RADAR CLUTTER CLASSIFICATION

by

WOLFGANG STEHWIEN, M.Eng.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

November 1989

RADAR CLUTTER CLASSIFICATION

DOCTOR OF PHILOSOPHY (1989)
(Electrical Engineering)

McMaster University
Hamilton, Ontario

TITLE: Radar Clutter Classification

AUTHOR: Wolfgang Stehwien
B.Sc.E.E. (University of Saskatchewan, 1976)
M.Eng. (McMaster University, 1983)

SUPERVISOR: Dr. Simon Haykin

NUMBER OF PAGES: *xx*, 350

ABSTRACT

The problem of classifying radar clutter as found on air traffic control radar systems is studied, and an algorithm is developed to carry out this classification automatically. The basis for the algorithm is Bayes decision theory and the parametric maximum a posteriori probability (MAP) classifier. This classifier employs a quadratic discriminant function and is optimum for feature vectors that are distributed according to the multivariate normal density. Separable clutter classes are most likely to arise from the analysis of the Doppler spectrum. Specifically, a feature set based on the complex reflection coefficients of the lattice prediction error filter (PEF) is proposed. These coefficients are also used in the maximum entropy method (MEM) of spectral estimation, and this link establishes many of their characteristics. A number of transformations are necessary, however, before they can be used as features.

The classifier is thoroughly tested using data recorded from two L-band air traffic control radars at different sites. The collected data base contains extensive bird, rain, and ground clutter, as well as thunderstorms, aircraft and ground-based moving vehicle echoes. Their Doppler spectra are examined; and the properties of the feature set, computed using these data, are studied in terms of both the marginal and multivariate statistics. Several strategies involving different numbers of features, class assignments, and data set pretesting according to Doppler frequency and signal-to-noise ratio, were evaluated before settling on a workable algorithm. Final results are presented in terms of experimental misclassification rates and simulated and classified PPI displays.

ACKNOWLEDGEMENTS

I extend my most sincere thanks to Dr. Simon Haykin for his continuing support and encouragement throughout this research. It was a privilege to have studied with his guidance.

Dr. Ed Jernigan of the University of Waterloo supplied much of the impetus toward the use of statistical pattern recognition theory for this work. His course and his personal comments contributed immeasurably to this research, and I am very grateful for these.

Dr. Max Wong and Dr. Tom Timusk served on my supervisory committee and provided many helpful directions, which I am very grateful for. I also acknowledge the contributions and discussions offered by other members of the Communications Research Laboratory, in particular Dr. D. Taylor, Dr. C. Carter, and Dr. J. Reilly. Thanks are extended to Mr. Terry Greenlay and Mr. Brian Currie, who offered many constructive suggestions throughout the research work and assisted with the computational facilities, and Mr. Peter Weber, who participated in many stimulating discussions.

The radar personnel at CFBs Trenton and Moose Jaw, and NDHQ in Ottawa, who gave of their time to support my data collection efforts and from whom I learned much about the TRACS radar, deserve special recognition.

To my wife, for her continuing patience and support: Thank You.

Finally, the financial support extended by the National Sciences and Engineering Research Council and the Department of National Defence is gratefully acknowledged.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

agl/asl	above ground level/above sea level
AR	auto-regressive
ASR	Area Surveillance Radar (also: Airport Surveillance Radar)
ARSR	Air-Route Surveillance Radar
ATC	Air Traffic Control
CFAR	constant false alarm rate
FFT	fast fourier transform
MAP	maximum a posteriori probability
MEM	maximum entropy method
MICD	minimum intra-class distance
MTI	Moving Target Indicator
MTD	Moving Target Detector
PEF	Prediction Error Filter
PPI	Plan Position Indicator
PRF	pulse repetition frequency
PRI	pulse repetition interval
RCS	radar cross-section
SNR	signal-to-noise ratio
SSR	Secondary Surveillance Radar
STC	Sensitivity Time Control
TRACS	Terminal Radar and Control System

List of Symbols

$E[x]$	expectation
$N(\mu,\sigma^2)$	normal density with mean μ and variance σ^2
$N(x:\mu,\sigma^2)$	normal density using variable x with mean μ and variance σ^2
χ^2_n	chi-squared (n degrees of freedom) distributed variate
$F_{n,m}$	central F (n,m degrees of freedom) distributed variate
$\Gamma(\ell)$	gamma function
P_D	probability of detection
P_{FA}	probability of a false alarm occurring
$P(c_i)$	a priori probability
$P(c_i \mid x)$	conditional probability
$p(x)$	probability density function
$p(x \mid c_i)$	conditional probability density function
$\lambda(\alpha_i \mid c_j)$	conditional loss function
R	Bayes risk
$R(\alpha_i \mid x)$	conditional risk
δ_{ij}	Kronecker delta
c_i	class i of L classes
x	d-dimensional feature vector
$\mathbb{R}$	set of feature vectors x
$\mathbb{R}_i$	region in the feature space of x where c_i is selected
μ_i	d-dimensional feature mean vector of class i

$\hat{\mu}_i$	d-dimensional feature mean vector estimate
Σ_i	d x d feature covariance matrix of class i
$\hat{\Sigma}_i$	d x d feature covariance matrix estimate
C_i	d x d sample covariance matrix
G_i	d x d transformation matrix
y_i	transformed d-dimensional feature vector
$g_i(x)$	discriminant function to be maximized over all classes
$g_i'(x)$	discriminant function to be minimized
$D_i^2(x)$	squared Mahalanobis distance from μ_i to x
$\hat{D}_i^2(x)$	squared Mahalanobis distance using estimates $\hat{\mu}_i$ and C
J_{ij}	Divergence between classes c_i and c_j
B_{ij}	Bhattacharyya distance
K_{ij}	Kolmogorov variational distance
T	class rejection threshold (Mahalanobis distance)
T'	significance test threshold (probability)
V	time-varying coherently received signal
A	amplitude of V
ω	frequency of V
ψ	arbitrary starting phase
P	relative signal power (squared voltage)
W	relative white noise power
X	coherently demodulated signal phasor
I	in-phase component of signal vector X
Q	quadrature component of signal vector X
S	amplitude of vector X

ϕ	relative frequency (normalized to f_s)
f_s	sampling frequency/pulse repetition frequency
k_e	in-phase to quadrature channel voltage ratio
e	in-phase channel gain error
δ	quadrature channel phase error
FT	false-target-level ratio
d_0	channel differential noise factor
x_n	complex data sample n of signal vector V
a_n	real (in-phase) component of x_n
b_n	imaginary (quadrature) component of x_n
w_n	complex noise sample n
w_{In}	real (in-phase) component of w_n
w_{Qn}	imaginary (quadrature) component of w_n
n_0	signal-to-noise ratio
$R(m)$	autocorrelation value for lag m
$h_i^{(m)}$	i^{th} coefficient of m^{th}-order PEF
$\gamma_i^{(m)}$	i^{th} parameter of m^{th}-order autoregressive model
$\bar{\gamma}$	vector of autoregressive model parameters
$\overline{W}$	prediction error power estimate
ρ_m	m^{th}-order Burg lattice PEF reflection coefficient
r	magnitude of ρ_1
$f_n^{(m)}$	m^{th}-order PEF forward prediction error sample n
$b_n^{(m)}$	m^{th}-order PEF backward prediction error sample n
P_0	signal-plus-noise power, equal to $R(0)$
P_m	prediction error power of m^{th}-order PEF

N	total number of data samples
M	order of Burg lattice
U_0	P_0-based feature
u_0	ρ_1-based feature
P_{var}	variance of sample power within measurement window
P_{dif}	difference between sample power in window center and edges

Chapter 1

INTRODUCTION

1.1 The Need for Classification

Modern air traffic control (ATC) radar systems rely heavily on digital technology to ease controller workload in identifying and tracking aircraft targets in busy airport terminal areas. Airborne transponder-based secondary surveillance radars (SSR) and fully processed synthetic displays have significantly altered the role of primary radar within the ATC system. It is now used mainly to provide a backup to the SSR and to "reinforce" their targets: primary radar measurements of range and azimuth are generally more accurate. Its ability to detect airborne hazards can, however, assist controllers immensely in vectoring aircraft around storms and, in some cases, bird flocks. Unfortunately, the advent of digital (synthetic) displays prevents radar operators from using their trained judgment to interpret the raw data as it would have appeared on the older style plan position indicator (PPI). There is thus a need to identify these hazards automatically and present the results on the processed display.

The two most common hazards to aircraft are weather and birds. The most dangerous forms of weather are thunderstorms and heavy precipitation (in which airframe icing frequently occurs). Winds within thunderstorms present a serious threat to airliners, especially during the landing phase, and can literally tear small aircraft apart. Aircraft collisions with birds can result in serious damage, crashes and fatalities. The small size of birds and the very high closing speed of jet aircraft makes collision avoidance by pilots extremely difficult. Low-flying jets are particularly vulnerable, as

are aircraft taking off and landing. Bird scaring techniques are partially successful at airports, but this is clearly no solution elsewhere. Periods of heavy night-time migration during the spring and fall 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 non-aircraft targets as undesirable "clutter", which must be suppressed to enhance the visibility of aircraft within the clutter areas. The moving target indicator (MTI) is essentially a Doppler filter designed to remove all non-moving echoes such as ground clutter. Modern digital MTI filters are quite effective and provide as much as 50 dB improvement for detecting targets flying over areas with strong 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 number of these false alarms is then controlled using constant false-alarm rate (CFAR) or programmable sensitivity time control (STC) circuits, or both. These, unfortunately, also reduce the detectability of legitimate radar targets.

The latest generation of primary radars will improve on the MTI filters using the so-called moving target detector (MTD) [Rabinowitz *et al.*, 1985; Taylor and Brunins, 1985]. This detector 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. Other algorithms include model-dependent detectors which adapt to the clutter rather than use fixed filter banks, but they too aim to completely remove clutter. While these improved detectors allow progressively weaker targets to become visible in the presence of clutter, they also further suppress the availability of information regarding airborne hazards. At the same time it is not clear how these systems respond

to bird echoes, which can take on radar cross-sections (RCS) and groundspeeds similar to those of small aircraft, and don't often appear as distributed clutter. Their detection will only confuse controllers unless they are specifically identified as birds.

In this thesis an attempt is made to classify the different forms of radar signals, rather than detect moving targets. It can be said that detection is simply classification with two classes, and thus this work is a generalization of the detector. There is, however, an important distinction to be made: in classification the characteristics of the classes, once determined, are generally fixed or updated only gradually. The modern detector, on the other hand, adapts itself to the background and as such continuously changes the characteristics of the *background* class in order to maximize the probability of detection for the *target* class.

1.2 The Classifier Model

A generic block diagram of the classifier is shown in Figure 1.1. 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 its internal motion. Since not all of this information is unique to any one clutter type, those features which can reliably identify the clutter need to be *extracted* and passed on to the classifier.

The classifier must then decide which class c_j the measurement sample vector x belongs to. This is frequently achieved using the notion of a distance metric. The distance of the sample from each possible class is computed, and that class to

which the sample is closest is then chosen. If it is possible to define the statistics of the class and the measurement samples, then potentially fewer misclassifications may result whenever x is assigned to c_j such that the a posteriori probability $P(c_j|x) > P(c_i|x)$ for all $i \neq j$. Using *Bayes Rule* [Duda and Hart, 1973],

$$P(c_j|x) = \frac{p(x|c_j)P(c_j)}{p(x)} \tag{1.1}$$

where

$$p(x) = \sum_{j=1}^{n} p(x|c_j)P(c_j) \tag{1.2}$$

This maximum a posteriori probability (MAP), or Bayes classifier is in fact optimal when the a priori probabilities $P(c_j)$ and the probability densities $p(x|c_j)$ are known. In most cases, however, only estimates for these quantities are available and the classifier

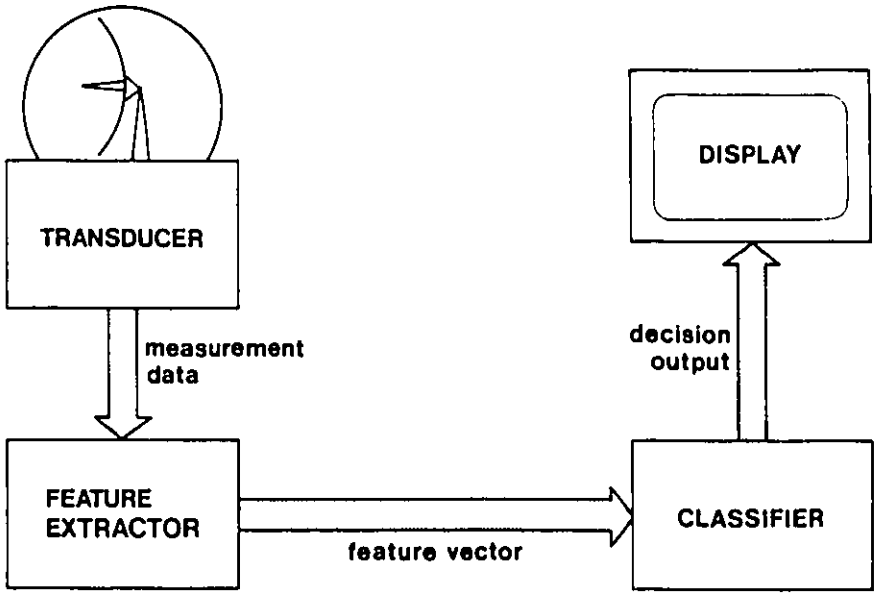


Figure 1.1 Block diagram of the classifier

will necessarily perform suboptimally.

The actual increase in the rate of misclassification through the use of a suboptimal classifier is not clear, and some classifiers have been devised to accommodate cases where parametric estimates (based on, for example, assumptions of normality) are clearly poor representations of the true statistics. These nonparametric techniques, such as the nearest neighbor rule [Duda and Hart, 1973], require that all labelled samples be stored and compared in some way to each new sample. This can lead to prohibitive memory and computational requirements, especially for radar data which becomes available and must be processed at rates of more than one million samples per second.

Parametric estimates which provide sufficient information for successful (even if sub-optimal) classification are thus essential for real-time systems. Acceptable performance must therefore be achieved by defining features and classes in such a way that the estimates will reflect the true statistics as closely as possible.

1.3 Spectral Parameters as Features

A central question for the classification of clutter concerns the *selection* of features, that is which features, if any, are present in coherent radar returns and able to uniquely 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 these. First is the modulation of the time series by the azimuth gain pattern of the main scanning beam, which is usually approximately Gaussian or $sin(x)/x$. Clearly, the fewer samples available from the target, the wider the spectral spread will be. 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.

1.4 Chronological Overview of the Research Activity

Previous work [Stehwien, 1983] established and analyzed the Burg reflection coefficients as a potential feature set for classification. The early stages of this research were concerned with the proof of concept of the application of the reflection coefficients to the parametric Bayes classifier. A comparison with other, distance metric based classifiers showed superior performance in virtually all cases [Stehwien, 1984]. Unfortunately, these analyses all suffered substantially from a lack of data and little could be said about the asymptotic statistics of the feature set or the classifier performance. In fact there is considerable suspicion that the apparently good results were simply due to local differences in the data, and that the necessity of testing the classifier on the training set provided excessively optimistic results. Hence, before any algorithm could be declared successful, a much greater volume of data was needed in order to allow exhaustive training and testing. (Exhaustive within practical limits, of course.)

This need could only be met by allocating a major portion of the research time to collect and analyze large quantities of radar data. Two radar sites were visited and nearly five hours of data in 18 minute segments were recorded [Stehwien and Haykin, January 1986]. Selected segments of this data were transferred to a research computer for analysis. Software was written to efficiently examine and display the data for verification of integrity and further selection of training sets. Daily weather and bird migration forecasts and reports along with time-exposure photographs of the PPI

and experience in observation served to establish the a priori knowledge about the nature of the recorded clutter which was necessary for the training of the classifier.

Several scenarios for classifier training were tested to maximize the separability of the clutter classes. It was not clear at this stage how many classes could be, or perhaps even needed to be defined for optimal performance given certain computational restrictions. Little theoretic analysis could be applied to such considerations as feature selection and extraction, dimensionality, data window size per feature set, number of classes per clutter type, and classifier performance. Hence several cycles of feature extraction from the data, class definition, and performance testing were required before the algorithm could be considered workable. The final test, of course, lies less in numbers but in the presentation of the classifier output as an air traffic controller might see it. Generation of such displays rounded out the research activity.

1.5 Thesis Outline

The thesis is structured to provide a fairly complete picture of the radar classification problem. It begins, in Chapter 2, with a review of the developments leading up to the current work. The requisite theory is presented in Chapters 3, 4, and 5. The presentations are necessarily aimed at achieving a practical solution and thus contain a strong component on the relationships between sample size and error performance. Chapter 6 details the experimental work and its results, and Chapter 7 examines the practicality of the final algorithm in terms of its implementation in real time, and its finite precision performance. Chapter 8 concludes by highlighting the research findings and contributions. A more detailed overview of the thesis is given in the following outlines.

Chapter 2 reviews the background to the research as reported in the literature, starting with the initial recognition of the problem (Section 2.1) and early attempts at classification using noncoherent radars in Section 2.2. These attempts were really studies in clutter characterization and feature selection, and were quickly expanded to include Doppler signatures as well (Section 2.3). Research into automatic classification itself, however, has progressed largely as an independent discipline and its application to radar clutter (Section 2.4) appears not to have received much attention.

A detailed analysis of the spectral content of radar clutter is given in Chapter 3. The radar is first introduced as a sensor of the environment followed by, in Section 3.2, a discussion of the physical phenomena which control the scattering of radar energy by clutter sources and targets of interest. This discussion is illustrated with sample spectra. Section 3.3 examines how the clutter phenomena determine the spread of these spectra, and Section 3.4 expands this to include the effects of the scanning beam modulation. A brief discussion on the consequences of the use of staggered PRFs, which is common in radar systems designed around conventional MTI filters, concludes the chapter.

Bayes decision theory as applied to classification is reviewed in Chapter 4. First, the relevant Bayes theory is reviewed in Section 4.1. Performance prediction is essential for the selection of effective features, but quite difficult for the parametric multiclass Bayes classifier. Some error bounds and their application to the multiclass problem are discussed in Section 4.2. Measures of confidence in the decision can be quite useful to the radar operator who has to interpret the classifier output. Two such measures are proposed in Section 4.3. Finally the problem of classifier training is addressed. The solution to this problem is critical to a satisfactory performance of the

classifier. *A priori* knowledge is required to allow the assignment of labels to training samples: bad labelling will inevitably lead to substandard, and perhaps even completely erroneous decisions. Because of the non-separable nature of the data, clustering algorithms used in unsupervised learning techniques will not likely perform well. In addition to a priori knowledge, a large sample size is also essential to develop confidence in the class prototype. This is particularly true for long feature vectors (high-dimensionality classifiers).

Chapter 5 is concerned solely with the feature vector. The question of optimal feature selection is examined first, along with the reasons for the continued ad-hoc nature of the feature selection process. In Section 5.2 the proposed feature set is derived from the Burg reflection coefficients. Several transformations are required before they can be used effectively. Their statistics, both asymptotic and experimental, are discussed in Section 5.3, along with the relationship to data availability. The problem of dimensionality is examined in more detail with respect to the chosen feature set in Section 5.4. Section 5.5 then presents an ad-hoc solution to the nonlinear dependence of all features on the signal-to-noise ratio (SNR) which, if not removed, is sufficiently strong to almost completely distort the classification process into one of simply assigning classes based on the SNR alone.

Experimental results are presented in Chapter 6, after first describing the radar system and the data derived from it. A performance analysis of the radar with respect to point targets in general and bird clutter in particular is also included. The training strategy and the selection of the labelled samples is discussed in Section 6.2. During this process considerable insight was gained into the behavior and statistics of the clutter data. Various experiments and their results are interpreted in Section 6.3. Finally, classified radar images are shown as the "proof of the pudding".

Chapter 7 deals with the implementation of the algorithm in real time. Computational requirements of the algorithm are summarized and compared to the real-time needs of the radar. Some thoughts are presented on the feasibility of using finite precision and fixed point arithmetic, and modern, highly parallel processors.

Several appendices are included to supplement the understanding of the radar system. The first three deal with its specifications, detection performance, and the elusive nature of the proof for the presence of bird clutter. The expected value of the reflection coefficients given a complex-sinusoid-in-noise case is derived in Appendix D. In Appendix E, an attempt is made to quantify the sensitivity of the first reflection coefficient to in-phase and quadrature (I and Q) channel misalignments, namely different gain and noise levels per channel, as well as phase misadjustments (Q not orthogonal to I).

Chapter 2

HISTORICAL PERSPECTIVE

This chapter 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; an excellent overview, particularly from the British perspective, may be found in Eastwood's book on Radar Ornithology [Eastwood, 1967]. More recently, Skolnik [1985] prepared a brief history and discussion of the state of radar today. References provided in this chapter have been chosen to be representative and significant to this research. Other and related work is amply referenced in the comprehensive review papers and books.

2.1 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 systems as a secondary

surveillance radar (SSR) or, more formally, the *Air Traffic Control Radar Beacon System* (ATCRBS) [Skolnik, 1970]. 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 [Eastwood, 1967].

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 the moving target indicator (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 [Eastwood, 1967; Flock and Green, 1974]. 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 [Vaughn, 1985]. 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 [Stables and New, 1968]. The use of radar to help avoid bird strikes seemed logical, especially since most busy airports use radar for air traffic control [Blokpoel, 1976].

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. Photographic records are more reliable, 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 ornithologist in

migration studies, and for migration forecasts used in the planning of flight operations [Gunn and Solman, 1968; Blokpoel, 1973].

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 [Blokpoel, 1976]. 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.

2.2 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 [Skolnik, 1970]. 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 useful 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 [Schaefer, 1968; Flock and Green, 1974; Vaughn, 1985]. 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 is, however, 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 [1974] 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 examined by Barry *et al.* [1973], and Dill and Major [1977] actually studied the distributions of bird flocks in three dimensions. Some references to the work on polarization properties were provided by Vaughn [1985].

Meteorological clutter has long been the subject of study and much is known about the characteristics of rain and storm clutter [see Nathanson and Reilly, 1968; Smith *et al.*, 1974]. Long-term pulse-to-pulse correlation is much greater than it is for bird clutter [Zawadzki, 1973], as is the spatial uniformity in amplitude distribution and area movement. 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 [Taylor and Brunins, 1985]. 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 [Long, 1975]. Even today, ground clutter characterization cannot be considered complete [Barton, 1985]. Air traffic control radars generally combat clutter through the use of moving target indication (MTI) and moving target detection (MTD) techniques and ground clutter maps [Muehe *et al.*, 1974; Rabinowitz *et al.*, 1985; Taylor and Brunins, 1985].

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 [1977] describes automatic equipment which 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 [1975] 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 U.S. agencies. Currie *et al.* [1977] 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 useful. It appears that this work has not been continued beyond this point, with the emphasis shifting to the study of Doppler spectra.

2.3 Evolution of the Use of Spectral Parameters

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, 1985]. Spectral analysis is in fact a form 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 [1975] proposed, and Currie *et al.* [1977] tested the viability of using the spread of the Doppler spectrum as a discriminant. Chan and Haykin [1976] 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 [1978] 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. The authors found, using their recorded data, that significant differences appear to exist in the spectral spreads of bird, weather, and ground clutter. These results may also be found in [Haykin *et al.*, February 1979; Haykin and Kesler, 1979]. Chan and Haykin [1979] further explored the properties of the MEM for use with radar signals, and Currie [Currie and Haykin, 1981; Haykin *et al.*, 1982] actually constructed a working "classifier" prototype which computed "turbulence indices" to warn of the dangers of the airborne hazards. These indices were then related to individual clutter types for identification.

The CRL work did not, however, relate the observed characteristics to any of the physical phenomena at work in the scatterers which 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 non-scanning radars which allowed continuous measurements for several seconds [Flock and Green, 1974; Vaughn, 1985]. From these, 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 [Nathanson and Reilly, 1968; Smith *et al.*, 1974]. 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, the Doppler or radial velocity, the Doppler velocity profiles and gradients with height and range [Wilson *et al.*, 1980], and Doppler spectral spread [Aviation Week, 1983] to detect turbulence.

2.4 The Application of Pattern Recognition Techniques

The work discussed up to this point essentially all dealt with finding characteristics of clutter on which basis 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 [Fukunaga, 1972; Duda and Hart, 1973; Young and Calvert; 1974]. Comprehensive summary papers on the state of the research at that time were written by Ho and Agrawala [1968] and Kanal [1974]. Most of this work is still valid; and more recent research results have been compiled by Fu [1980] and Jain [1987].

In regard to radar, Skolnik [1970] 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 (one 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, 1973]. Barry *et al.* [1973] 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 [Skolnik, 1970], but also much more effective. While this work is now considered relatively mature [Jain, 1987], it has not been applied broadly to radar signal analysis. Non-radar applications appear to have received considerably more attention. Electrocardiogram (ECG) waveforms have frequently been used as examples of time series classification [Young and Calvert, 1974], and "optimal" rules have been devised for speaker waveform recognition [Kashyap, 1978]. Chen [1982] 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 [1984, 1985] 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 intra-class 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 [1984; Stehwien and Haykin, 1986, 1989] appears to have been the first to apply the Bayes classifier to parameters arising out of the estimation of the Doppler spectrum. This work is continued in this thesis.

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 [1973] applied such techniques to weather radar data to trace echo contours and track weather systems. Blackmer *et al.* [1973] extended the work and used the echo descriptions to forecast weather system movements.

Chapter 3

SPECTRAL CONTENT OF RADAR CLUTTER

The basic premise for radar clutter classification is that the radar returns contain sufficient information to allow such classification to be made. In this chapter the information content of clutter is examined in some detail. No attempt has been made to describe and test for macro level behavior such as spatial coverage and overall movement of the clutter, both within a scan and from scan to scan. Such behavior could be described and analyzed using syntactic pattern recognition techniques [see, for example, Duda and Hart, 1973], rather than the statistical methods proposed in this thesis. The following discussions thus deal with relatively short term (measurable within 50 ms) changes in and distributions of RCS and position relative to the radar. Section 3.1 introduces the radar as a sensor of reflectivity and velocities. Section 3.2 describes the physical phenomena inherent in the various clutter and target types, followed by a discussion of the effects of internal motion on their spectra in Section 3.3. The scanning motion of the radar further modulates the target echoes and is dealt with in Section 3.4. Finally, the consequences of using staggered PRFs, a common technique used to resolve range and velocity ambiguities, are outlined in Section 3.5.

3.1 The Radar as a Sensor

A surveillance radar consists of a high power pulse transmitter, a rotating antenna, and a receiver [Skolnik, 1970, 1980]. 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 pulse repetition frequency (PRF), and the rate of antenna rotation, and is given by

$$N_\theta = f_s t_d = f_s \frac{\theta}{\omega_a} \tag{3.1}$$

where N_θ is the number of pulses, f_s is the PRF in Hz, t_d is the dwell time on the target, θ is the 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 change in distance or velocity. The phase φ, as measured at a distance x from the transmitting radar t seconds after transmission, is

$$\varphi(x,t) = \varphi(0,0) + \omega_c t - \omega_c \frac{x}{c} \tag{3.2}$$

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

$$\frac{d\varphi}{dt} = \omega_c - \omega_c \frac{1}{c}\frac{dx}{dt} = \omega_c + \omega_d \tag{3.3}$$

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

$$\omega_d = -\omega_c \frac{1}{c}\frac{dx}{dt} = -\omega_c \frac{2v}{c} \tag{3.4}$$

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 $2x$ due the velocity of the target v. Equation (3.4) can be written in terms of the wavelength λ and the Doppler frequency expressed in Hertz

$$f_d = -f_c \frac{2v}{c} = -\frac{2v}{\lambda} \tag{3.5}$$

The carrier frequency will have been removed in a coherent receiver, reducing Equation (3.3) to $d\varphi/dt = \omega_d$. A pulsed Doppler radar is inherently a sampled system and therefore

$$2\pi f_d = \omega_d = \frac{\Delta\varphi}{\Delta t} = f_s \Delta\varphi = f_s \phi \tag{3.6}$$

where $\phi = \Delta\varphi$ is the relative Doppler frequency. Combining Equations (3.5) and (3.6) then provides the relationship between target velocity and the pulse-to-pulse phase change

$$v = -f_s \frac{\lambda}{2} \frac{\Delta\varphi}{2\pi} \tag{3.7}$$

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 frequencies 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.

This coherent video data is sampled at fixed intervals, whose length depends on the range resolution. To ensure a low probability of missing targets the data is typically oversampled, that is the sampling interval is much 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 which can be regarded as having originated from the same target to N_θ (per Equation 3.1). Because of the oversampling, adjacent range samples will usually contain returns from the same target. The number of adjacent time series which may be used to extract information from any one target will depend on the length of the sampling interval relative to the pulse width.

A number of excellent books on radar and radar theory have been published, notably by Skolnik [1970, 1980] and Nathanson [1969]. In-depth studies of radar reflectivity and range performance were published by Long [1975] and Blake [1980], respectively. Barton (among others) compiled a rather extensive collection of papers about various aspects of radar, of which two volumes are of particular interest here [Barton, 1974, 1975]. More recently, Brookner [1977, 1988] comprehensively summarized the state of radar technology.

3.2 Physical Phenomena in Clutter and Targets

Radar echoes arise from any reflecting surface or object within the line of sight of the radar. These objects include, besides aircraft, a range 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 9^{th}-order spectra were computed using the multisegment

Burg formula and the procedure outlined in Section 6.1.3. To allow comparisons, similar signal strengths (30 dB SNR) were selected except where not available (rain clutter at 20 dB). Mean Doppler shifts were also removed.

3.2.1 Targets and Moving Vehicles

Targets are often modeled as a moving sphere with a given radar cross-section. Such an object, moving in a straight line, will present a radial velocity component to the radar and as such result in a line spectrum in the absence of any modulation. This model works well in some cases; however, slight adjustments are frequently made. In a scanning radar the finite width of the beam will modulate the amplitude of the pulse return in accordance with the beamshape. If this shape is Gaussian, then the spectrum will also be Gaussian and no longer a line. This is evident from the spectrum of an idealized non-fluctuating target, shown in Figure 3.1. The effects of beamshape modulation are explored in more detail in Section 3.4.

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, 1969; Skolnik, 1970]. Neither do aircraft move in pure straight lines. 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 [Skolnik, 1970]. 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.

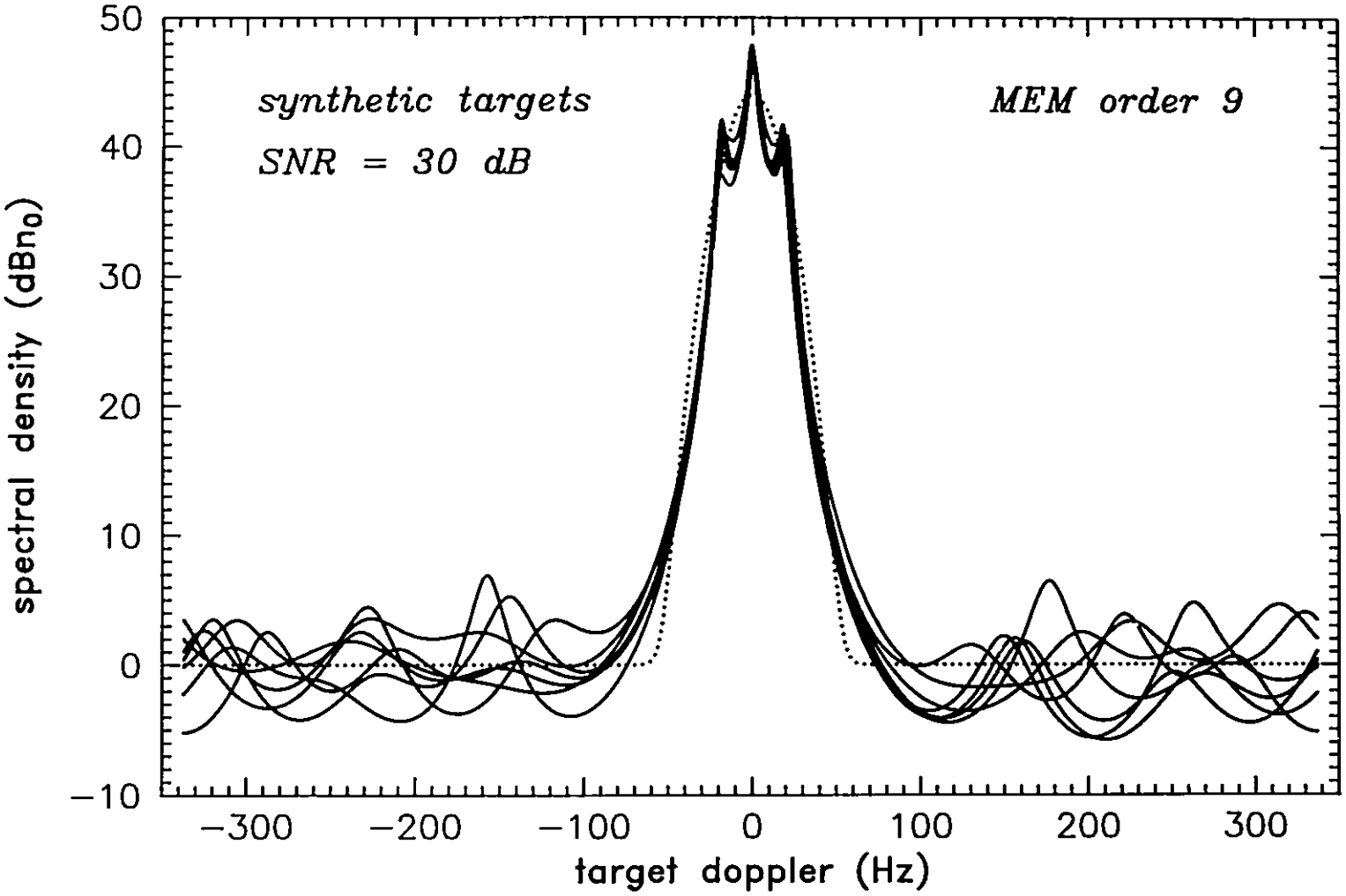


Figure 3.1 Maximum entropy spectra of synthetically generated targets

The dotted line represents the true spectrum of the beamshaped complex sinusoid in noise.

Figure 3.2 presents sample aircraft spectra which display some of these characteristics.

Ground-based moving vehicles are often mistaken by radar operators as light aircraft following a highway due to their similar speeds. There are a few significant differences, however. The RCS of cars is likely quite low compared to aircraft whose wings present a large reflecting surface. Trucks are therefore the most likely ground-based vehicles to be observed, 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 will also result in much lower RCS fluctuations. Neither do cars and trucks have propellers, and as such 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 Figure 3.3 have been normalized to move the primary peak to zero frequency, hence the ground clutter peak is elsewhere.

3.2.2 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 they tend to be the strongest of all clutter types and dominate 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 low or massive buildings. Trees, transmitter towers, and tall buildings all sway by varying amounts, depending on wind conditions. Higher

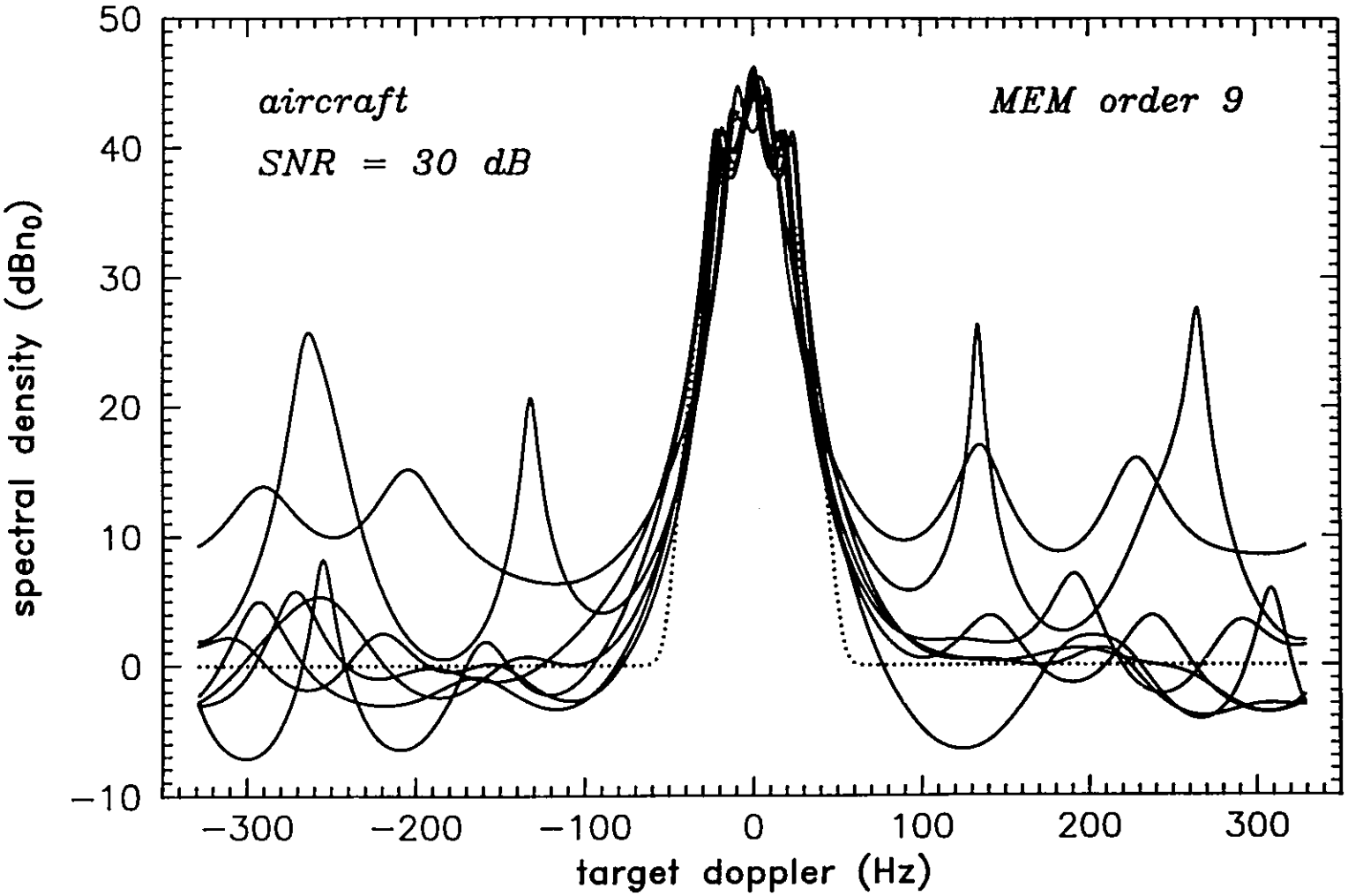


Figure 3.2 Maximum entropy spectra of aircraft targets

Note the additional sidebands and the greater variability in the shape of the peak.

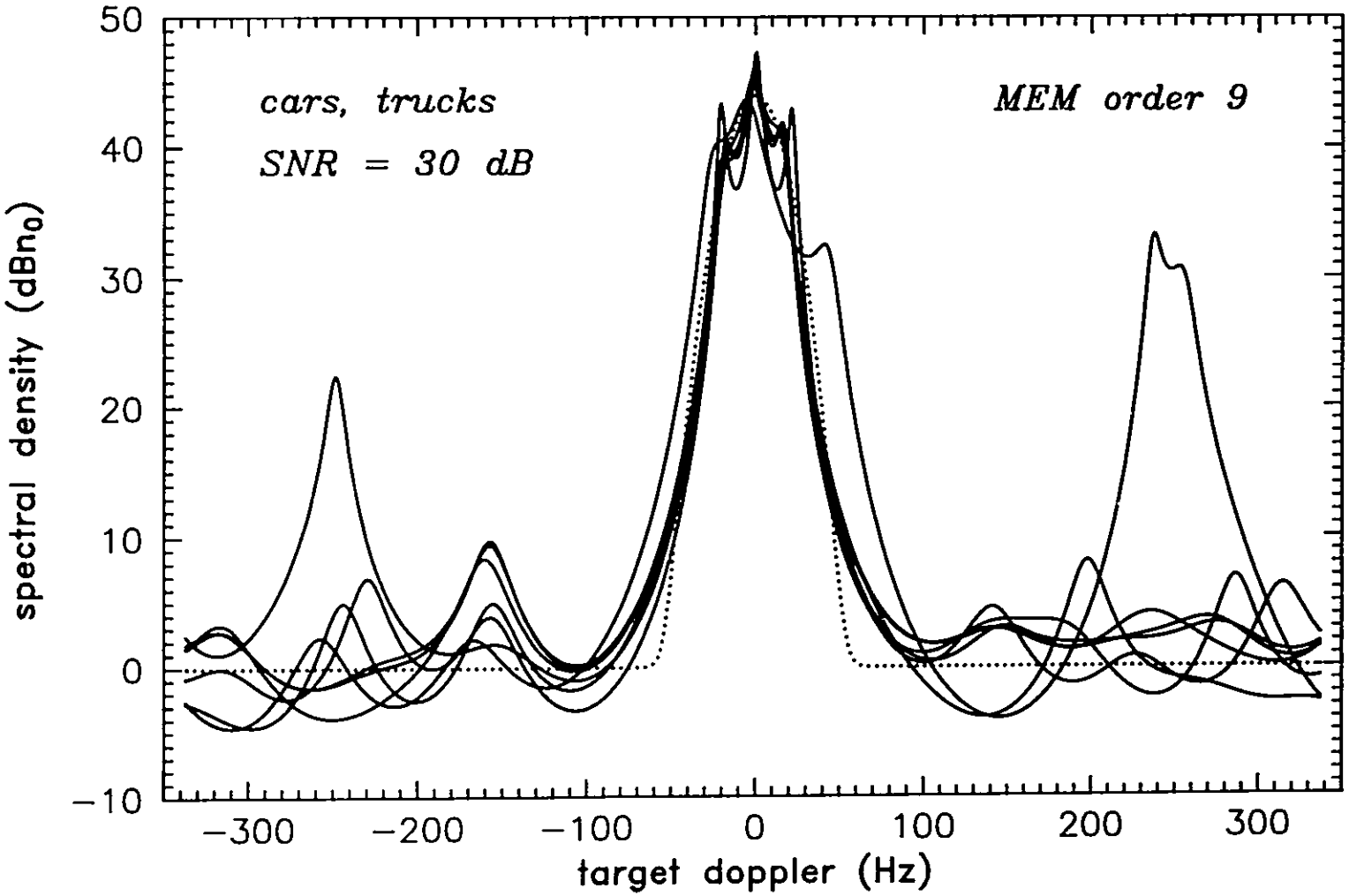


Figure 3.3 Maximum entropy spectra of car and truck targets

The sidebands are due to residual ground clutter (not at zero due to spectrum shifting).

frequency vibrations, depending on strength, may also be visible to the radar.

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 add 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 which 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 Figure 3.4. More details on ground clutter may be found in [Long, 1975; Skolnik, 1985].

3.2.3 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

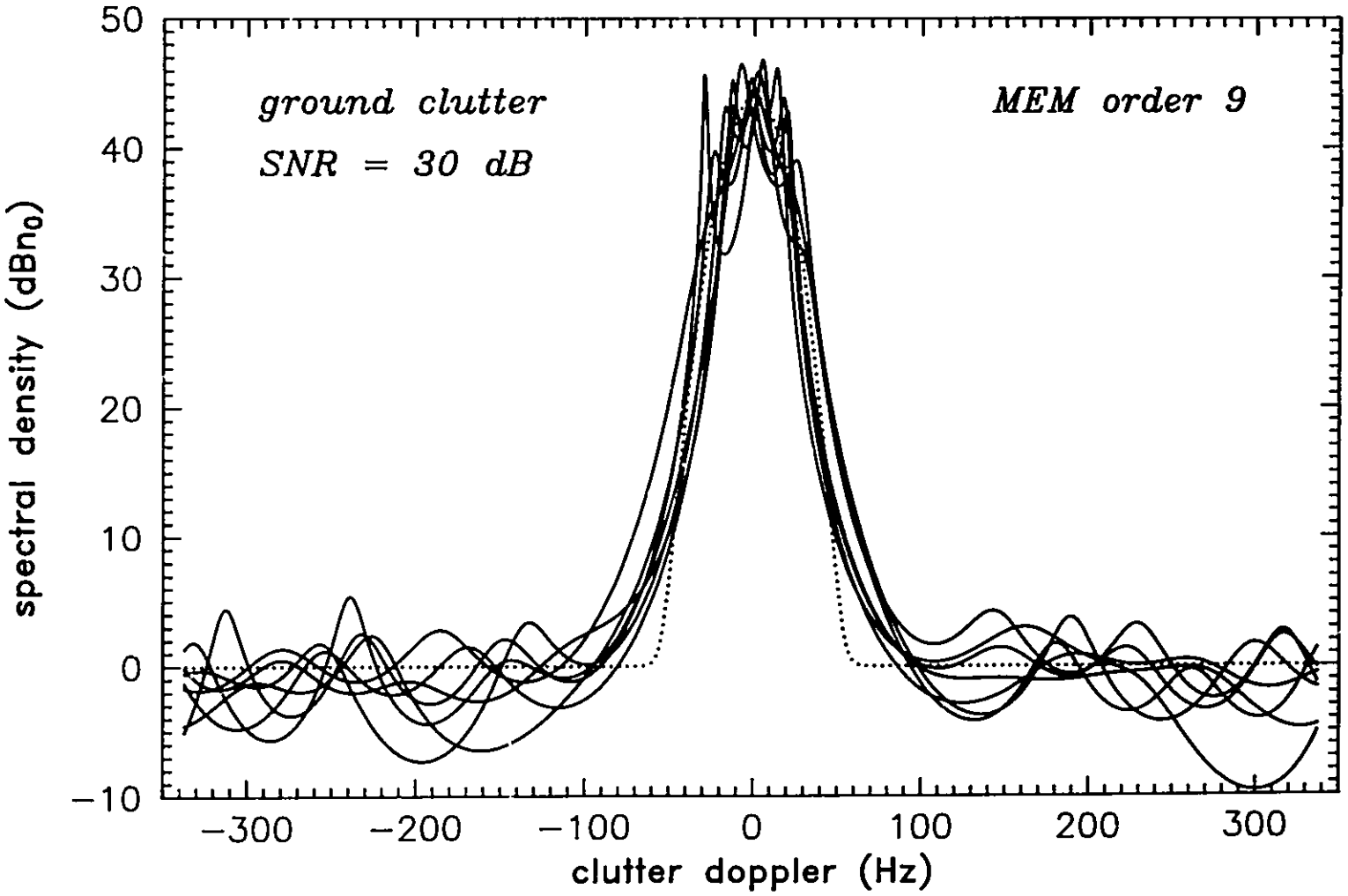


Figure 3.4 Maximum entropy spectra of ground clutter

Note the absence of sidebands and the considerable variability of the peak shape.

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 travelling 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 Figures 3.5 and 3.6. While the main spectral peaks are similar in both cases, the presence of a windshear component is visible in Figure 3.6. 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 [1968] and Smith *et al.* [1974].

3.2.4 Birds and Bird Flocks

Birds have long perplexed radar operators who refer to their echoes as "angels" [Eastwood, 1967]. It is very difficult to correlate bird echoes with visual observations (see Appendix C), 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 who fly in a particular formation. This causes large fluctuations in net echo strength as individual echoes sum constructively at one moment and destructively the next. Clearly the variation in net velocities, and consequently the spectrum, will be quite large for the same reason.

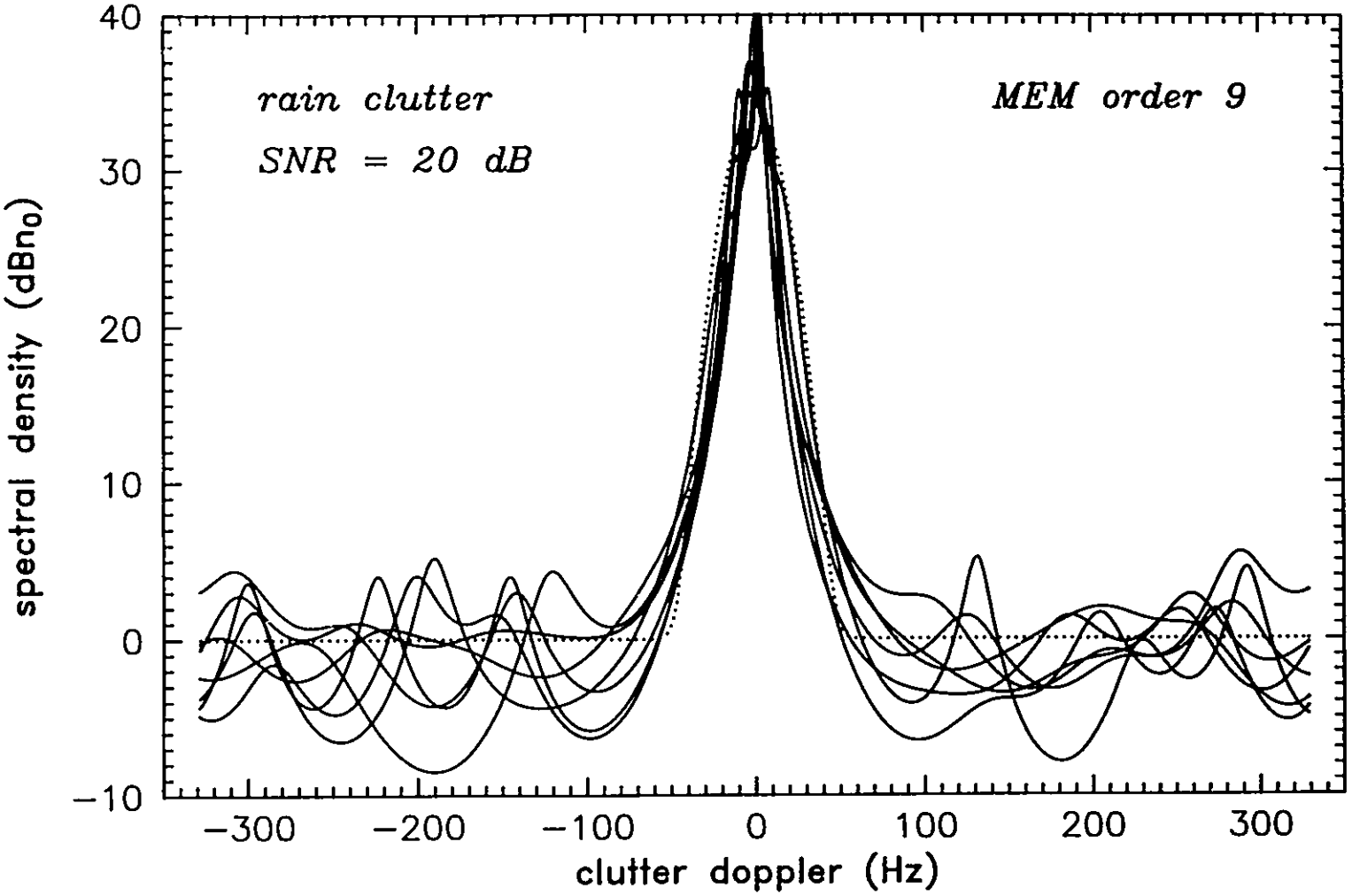


Figure 3.5 Maximum entropy spectra of rain clutter

Distributed clutter is affected less by beamshape mdulation, resulting in a narrow peak.

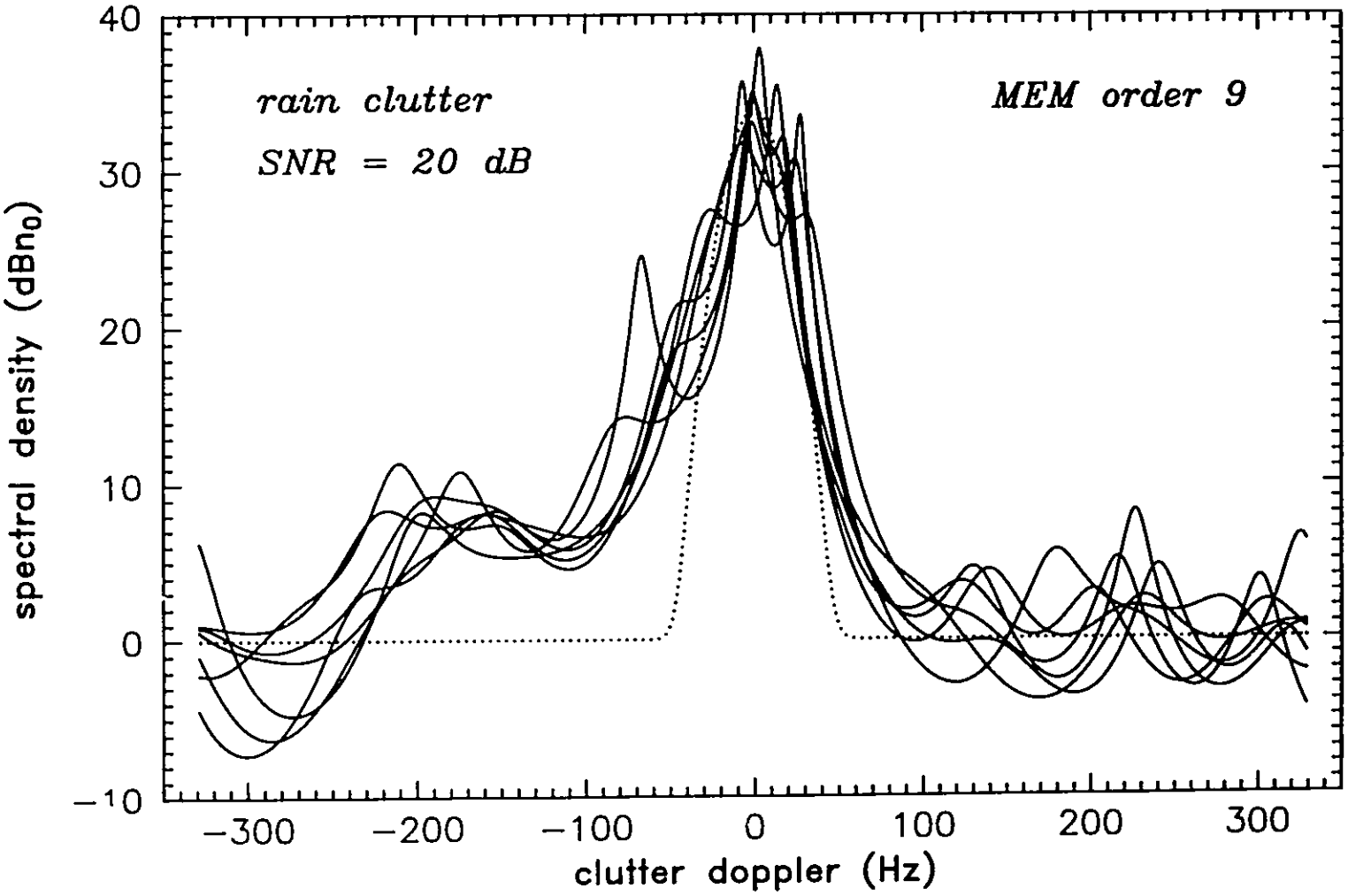


Figure 3.6 Maximum entropy spectra of rain clutter with a windshear component
Note the secondary peak and the greater variability of the primary peak.

Bird echoes are well known to exhibit wingbeat modulation, that is a variation in RCS as a function of the wing motion [Schaefer, 1968; Flock and Green, 1974]. This wing motion also causes instantaneous changes in the relative velocity as the bird 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 Figure 3.7. An excellent and more complete summary of the current knowledge about birds as radar targets has recently been published by Vaughn [1985].

3.2.5 Insects and Other Angels

Insect echoes have been observed with short-range high resolution radars. Their RCS is much too small, however, to be a factor in the long-range surveillance radar of the type used for this research [Riley, 1985]. Other unidentified clutter, by definition, cannot be characterized unless it is known to be all from the same source. The fact that it has been observed [see, for example: Hardy and Katz, 1969] should lead to the conclusion that not all is known about clutter and an 'unknown clutter' class should be included in the classifier. More concerns about the classifier training strategy are addressed in Chapter 6.

3.3 Spectral Spread and Shape

Equation (3.5) shows the linear relationship between target velocity and Doppler frequency. Figure 3.8 illustrates the geometry of the target velocity and its effects on the Doppler spectrum. Clearly only the radial component v_r contributes to

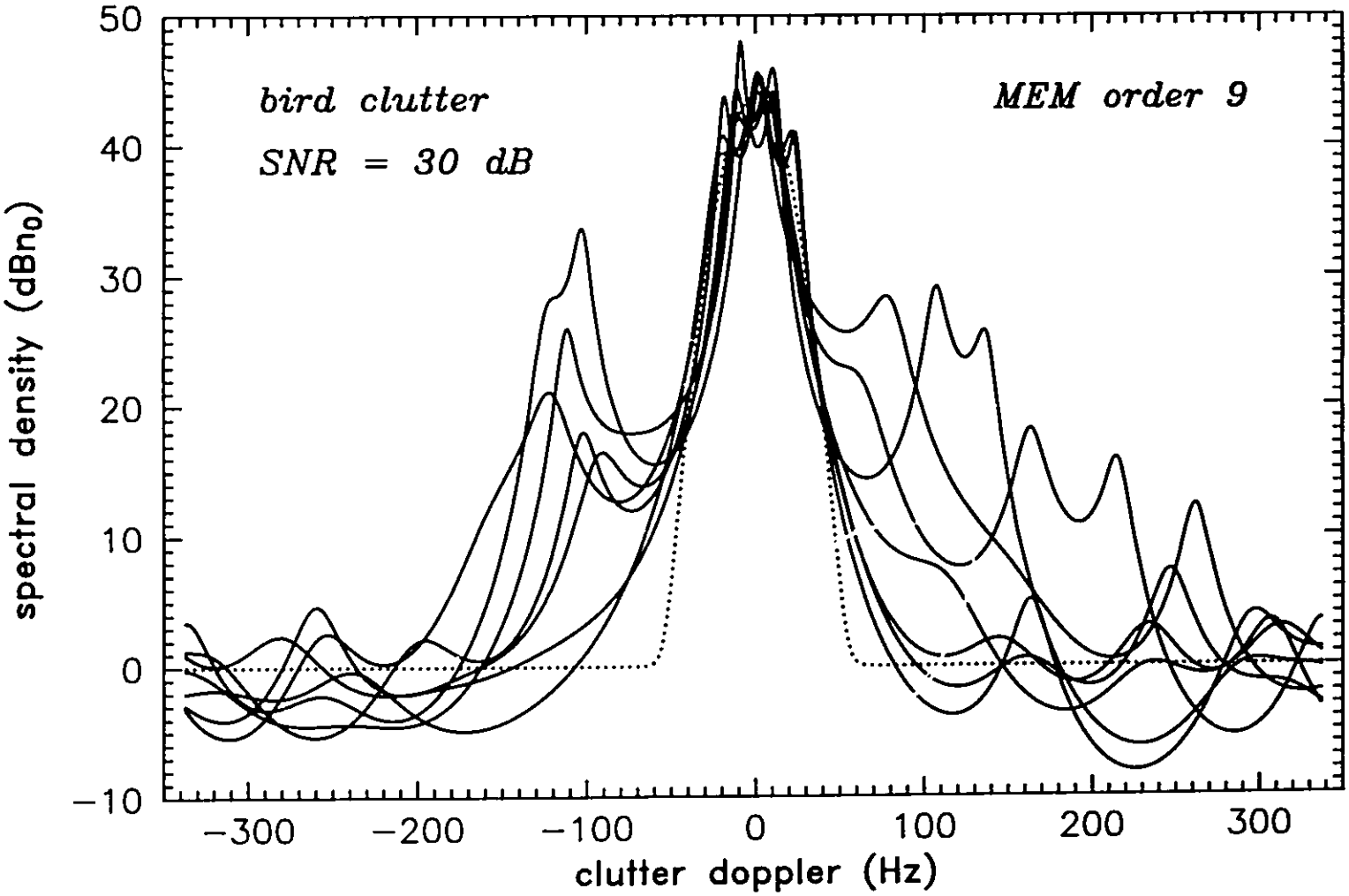


Figure 3.7 Maximum entropy spectra of bird clutter

The presence of many large independent scatterers results a wide and variable spectrum.

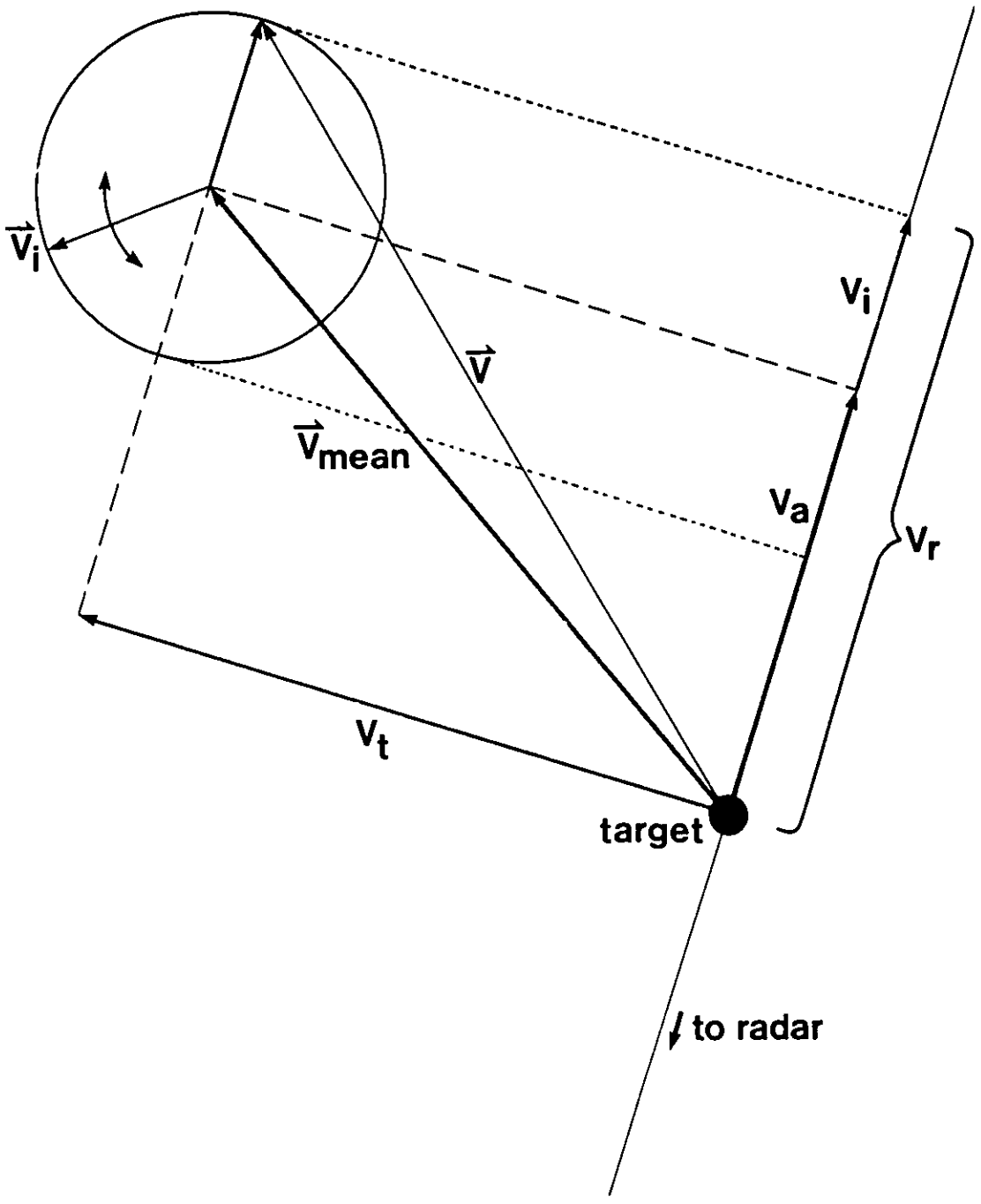


Figure 3.8 Moving target geometry

The target velocity *V* is resolved into its radial and tangential components *Vr* and *Vt*. The internal motion relative to *Vmean* is *Vi*, which modulates the radial velocity *Vr* about its average *Va*. These velocities give rise to frequency modulated doppler shifts. The diagram contains no information regarding target strength and amplitude modulation.

the Doppler shift. The width of the Doppler spectrum is determined by the bandwidth and amplitude of the modulation waveform which is proportional to the velocity v_t relative to the average radial velocity v_a. The shape itself will depend on the nature of the internal motion and the value of the modulation index.

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 3.9 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 labelled AM and FM. Both 5^{th}- and 9^{th}-order sets of spectra are shown for comparison. 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 knots and is not an unreasonable level of internal motion. A 20 Hz modulation is typical and representative of the wingbeat for bird echoes [Vaughn, 1985] and the Doppler spectrum deviations for rain clutter [Nathanson and Reilly, 1968].

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 peaks. Superposition of these spectra

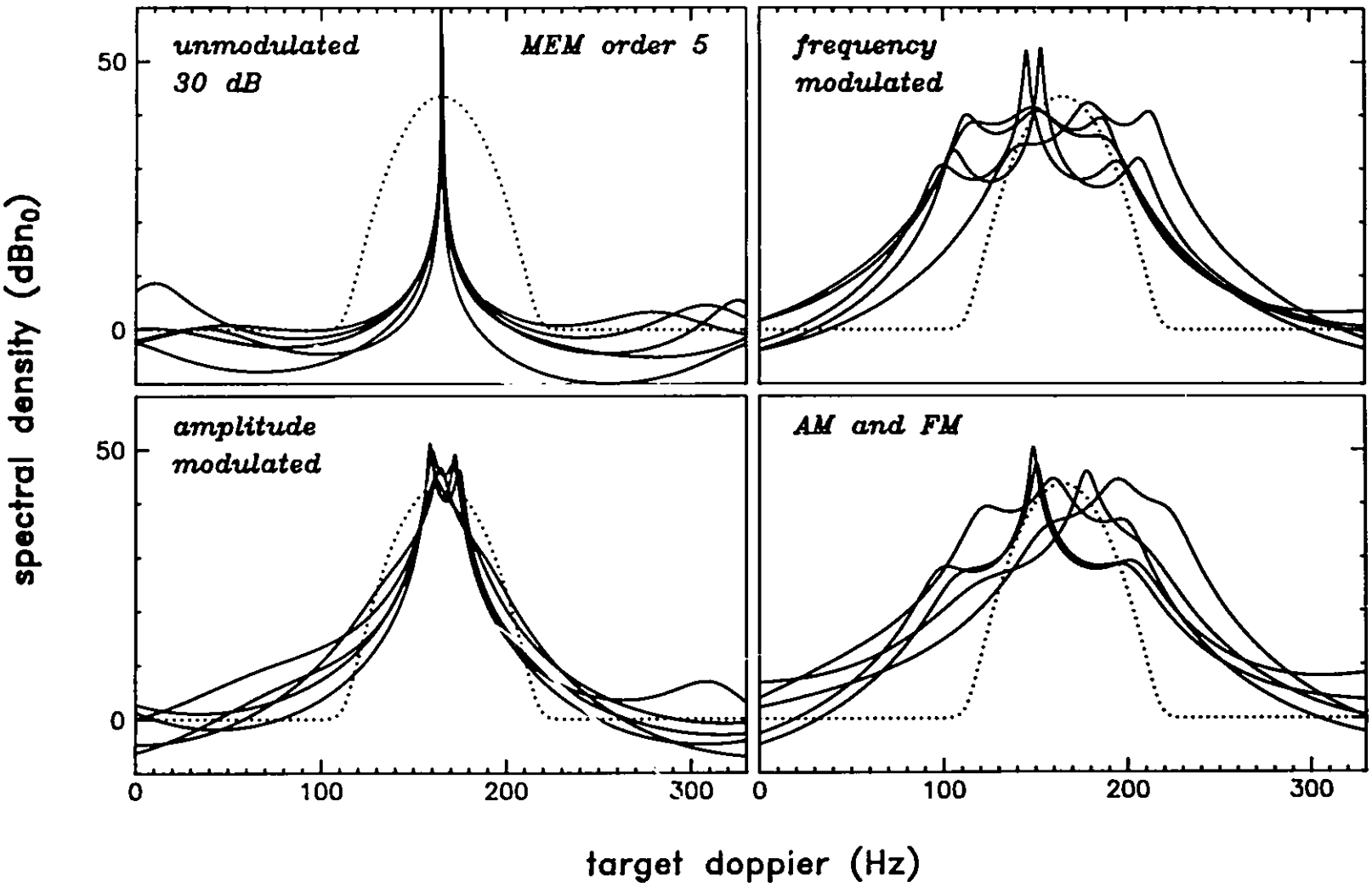


Figure 3.9a Maximum entropy spectra of amplitude and frequency modulated sinusoids in noise
The dotted lines represent the true spectrum of a sinusoid after beamshape modulation.

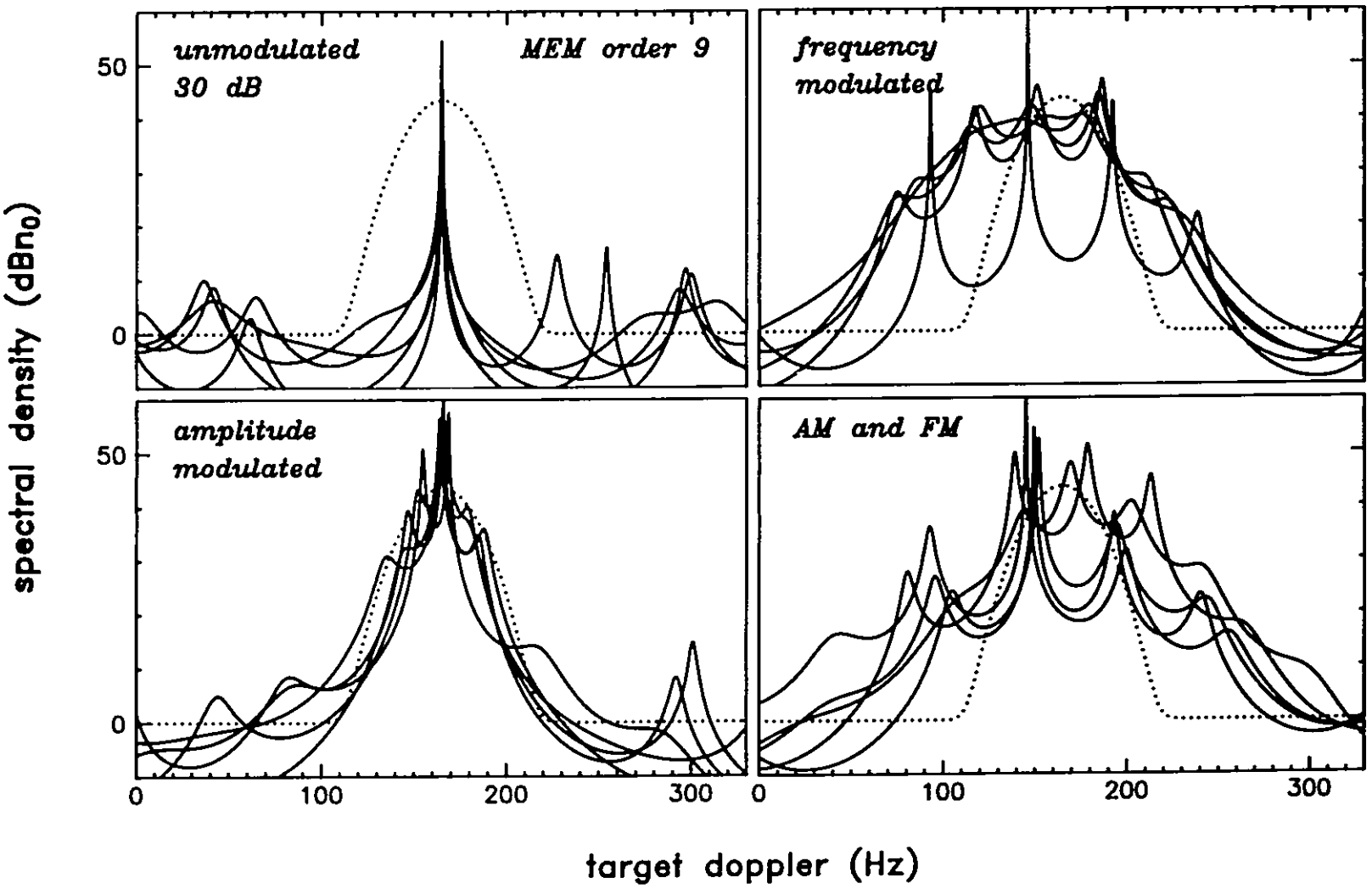


Figure 3.9b Maximum entropy spectra of amplitude and frequency modulated sinusoids in noise
9^{th}-order spectra of the same data used for Figure 3.9a.

leads to rather 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.

3.4 Effects of the Antenna Scanning Motion

The presence of unique characteristics within the Doppler spectrum of the echo 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 as the beam scans past it [from Equation (3.1): 13.5 to 14 hits/beamwidth for the TRACS–ASR, see Appendix A]. This will lead to a relatively large variance in the spectral estimate and, conversely, any feature set arising from it. The statistics of maximum likelihood correlation and covariance estimates, which will provide the basis for the feature extractor, are known to be [Burington and May, 1970]

$$E[(\hat{x}-\bar{x})^2] = \sigma^2/N \tag{3.8}$$

$$E[\hat{\sigma}^2] = \sigma^2(N-1)/N \tag{3.9}$$

$$E[(\hat{\sigma}^2-\sigma^2)^2] = 2\sigma^4/N \qquad (3.10)$$

where σ^2 is the true variance of the underlying, normally distributed (Gaussian) process, and N is the number of samples used for the estimate. 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 [Blake, 1980], will convolve with the basic spectrum, introduce additional spreading, and further reduce the maximum available resolution (see Figure 3.10). For distributed clutter this may not appear to be a problem since new scatterers are continuously entering the beam. These new scatterers are moving independently, however, and this can, in many cases, lead to significantly different spectra. As they superimpose the resulting spectrum will spread, and therefore the resolution of individual clutter spectra may not necessarily be any better.

The third limitation is the introduction of an apparent Doppler shift due to the size and orientation of reflecting objects. A building which presents a reflecting surface wider than the beamwidth, but with a non-tangential 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 non-moving point target with its Gaussian spectrum.

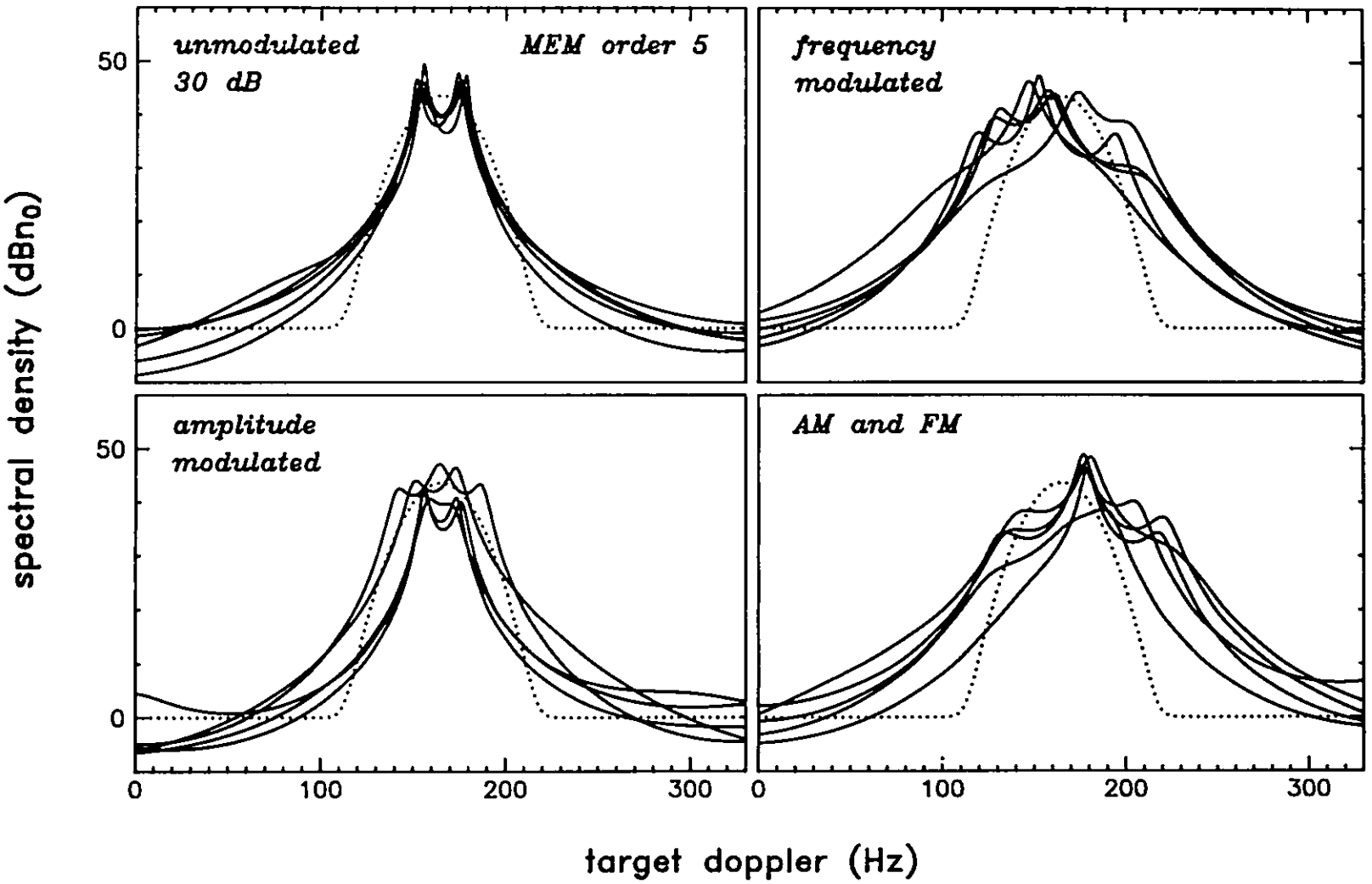


Figure 3.10a Maximum entropy spectra of modulated sinusoids in noise after beamshaping

Note the degree to which the sidebands have been obscured compared to Figure 3.9a.

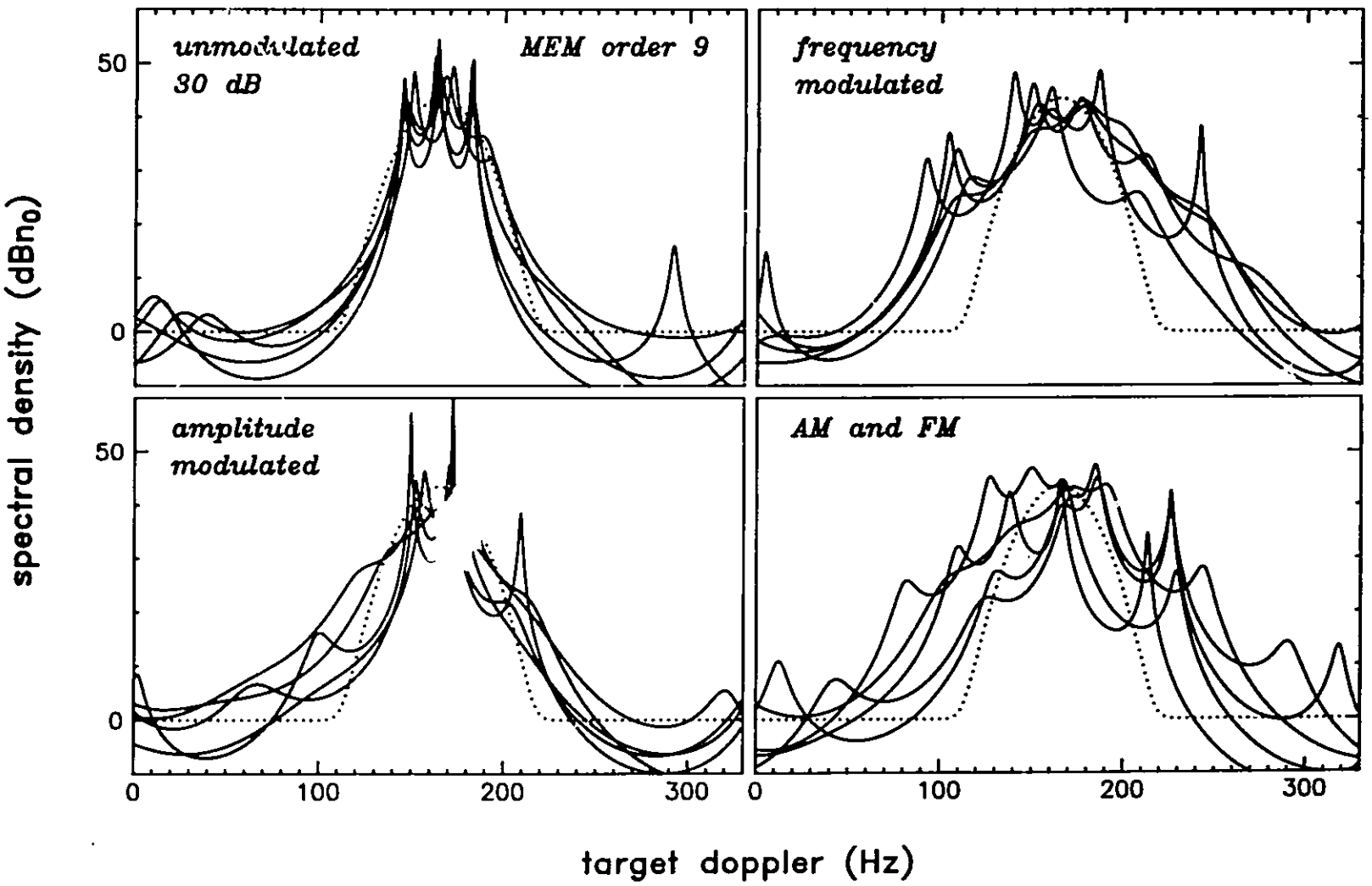


Figure 3.10b Maximum entropy spectra of modulated sinusoids in noise after beamshaping
9^{th}-order spectra of the same signals as in Figure 3.10a.

3.5 Effects of the Use of Staggered PRFs

The use of fixed pulse repetition frequencies in MTI radars has two problems associated with it. First, aliasing can occur about the PRF and if, as in the case of the TRACS radar, this frequency corresponds to a velocity much lower than is expected from aircraft, blind speeds can result. Second, spatial aliasing is possible if strong targets are found at distances greater than that allowed by the interpulse period. This will give rise to *second-time-around echoes* and cause problems with the operation of the radar.

Staggered PRFs solve both difficulties by spacing the pulses at variable intervals, thus forcing phase shifts to occur for all moving targets travelling at any speed and placing second-time-around echoes at different ranges to "tear" them apart. In practice, the number of different intervals is fixed and the sequence repeats itself continuously. This means that blind speeds still occur, although at a much higher frequency. For maximum benefit, the sequence contains at least as many different intervals as the number of hits on a target per beamwidth.

The forced phase shifts, when taken from successive pulses, constitute a signal with a considerably broader spectrum than before. As a result, some energy will appear at most or all frequencies and will pass the MTI filter and allow detection. Unfortunately, this same broadening obscures much, if not all of the discriminating information required for successful classification. Staggered PRFs cannot therefore be used for the classifier if conventional, fixed sampling interval based spectral estimation techniques are to be used.

The modeling of time series with unequally spaced observations has been studied by a number of researchers, notably Jones [1980, 1981] and Dunsmuir [1981]. The resulting models are generally discrete ARMA or continuous time AR, and are

found using maximum likelihood iterative techniques and the exact sample spacings. Such iterative techniques are not very suitable for real-time applications, and will not be discussed further. It is interesting, however, to consider the type of information extracted by these algorithms compared to that found when using equally spaced samples. An insight into the differences may be found in [Masry *et al.*, 1978], where the problem of finding the true and complete spectrum (without aliasing) is approached through the use of random sampling.

Suppose the process to be analyzed is stationary and has no well defined bandwidth. Clearly, the use of fixed intervals will always introduce aliasing and obscure the true spectrum. To minimize the foldover energy the sampling rate must be much higher than the highest frequency, and this may not be possible or practical. The use of random sampling will, at least in theory, provide information about the entire spectrum if carried on long enough. If, however, only a limited number of samples are available, then an information theoretic tradeoff must be made: random sampling provides information for as many different frequencies as possible, but no frequency will be looked at twice and the estimates remain noisy. Fixed interval sampling, on the other hand, repeatedly examines the same low frequencies and reduces the variance of the estimates, but provides fewer of them. The tradeoff is thus one of a low variance spectral estimate with a limited frequency range versus a high variance estimate with an extended frequency range. As will be seen in Chapter 4, a low variance is paramount for low error rate classification. The actual Doppler frequency is not very important, and the spectral width is well within the range of the fixed PRF (see Figures 3.1 to 3.7). Consequently, random or staggered sampling is undesirable even if it could be analyzed in real-time.

3.6 Summary

In this chapter the expected spectral content of radar clutter, as seen with a coherent Doppler surveillance radar system, was examined. The significant informational content of successive echo returns collected within a single beamwidth was determined to be the amplitude and Doppler signature. The spectral spread and shape, rather than the Doppler frequency itself, are most likely to allow discrimination between clutter types. It was also found that:

1. Ground clutter and aircraft targets have the narrowest intrinsic spectra.
2. Bird clutter has the widest and most variable spectra.
3. Weather clutter, due to its distributed nature, has narrower measured spectra than either ground clutter, aircraft, or birds. Its spectra still have finite width, however, and may exhibit several spectral peaks due to windshear.
4. An unknown clutter class is required to accommodate other, less common clutter types.
5. Both amplitude modulation (RCS fluctuation and scintillation) and frequency modulation (internal target or clutter motion) affect the clutter spectrum.
6. Antenna scanning motion contributes most of the spectral spread, but does not completely obscure the intrinsic spectral characteristics.
7. Spectra measured using staggered PRFs provide information different from that needed for reliable classification.

Chapter 4

BAYES CLASSIFICATION

Chapter 3 addressed the question: which characteristics of the clutter sources are visible to a surveillance radar? Having this information then leads naturally to the follow-up question: how can a decision regarding the identity of the clutter source be based on radar data alone, and how good is this decision? The answer may be found in Bayes decision theory, which provides optimum decision rules when the statistics of the problem are known. Section 4.1 presents a brief review of Bayes decision theory, and develops several discriminant functions which may be used to test the data. Section 4.2 attempts to answer the second part of the question, and offers some thoughts on performance predictors. Such predictors function essentially "off-line"; that is, they do not provide an immediate answer as the decision is being made. They are important, however, for the *selection* of features to use in the discriminant function, and to test the effectiveness of the *feature extraction* algorithm. "On-line", and working back from the discriminant function itself, it is possible to assign confidence measures to the decisions, as is shown in Section 4.3. These measures may be used under operational conditions to assess the quality of the decisions made.

Up to this point, knowledge of the underlying statistics has been assumed and required. This proviso does, of course, lead to another question: can this theory still be used if only an estimate of the statistics is available? The answer is clearly dependent on the quality of the estimates. Quantifying performance degradation is difficult at best when a priori knowledge about the statistics is not available, and may

not even be possible in the multiclass case. Hence the emphasis must be on ensuring that all estimates are as good as they can be. The discussion on classifier training in Section 4.4 addresses this need for sufficient data for each class of interest.

4.1 Review of Bayes Classification Theory

The basic classification system has been briefly described in Chapter 1. Figure 1.1 related the transducer to the radar itself. The feature extractor is the main signal processing element and is fully described in Chapter 5. This section deals with the classifier itself, which operates on the feature vector and assigns it to one of many possible and predefined classes. A thorough exposition of Bayes decision theory may be found in [Young and Calvert, 1974]. A more readable, but less rigorous treatment is presented by [Duda and Hart, 1973].

Designating these classes c_i, and the probability that the d-dimensional feature vector $\boldsymbol{x}$ belongs to any one of them as $P(c_i|\boldsymbol{x})$, then a simple rule for assignment of $\boldsymbol{x}$ to c_j is one which chooses the class with the largest *a posteriori* probability

$$P(c_j|\boldsymbol{x}) > P(c_i|\boldsymbol{x}), \qquad \text{all } i \neq j \tag{4.1}$$

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 $p(\boldsymbol{x}|c_i)$, which describe the statistics of the feature vector $\boldsymbol{x}$, given the underlying stochastic process belonging to class c_i. Then we may use *Bayes Rule*

$$P(c_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|c_i)P(c_i)}{p(\boldsymbol{x})} \tag{4.2}$$

where

$$p(x) = \sum_{i=1}^{L} p(x|c_i)P(c_i) \tag{4.3}$$

and L is the number of classes under consideration. Note that the *mixture* density $p(x)$ is independent of the index i, and therefore acts only as a scale factor which may be ignored when searching for the maximum per Equation (4.1).

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 x is

$$R(\alpha_i|x) = \sum_{j=1}^{L} \lambda(\alpha_i|c_j)P(c_j|x) \tag{4.4}$$

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, 1973] that this rule minimizes the overall Bayes risk R. Combining (4.2) and (4.4):

$$R(\alpha_i|x) = \sum_{j=1}^{L} \lambda(\alpha_i|c_j) \frac{p(x|c_j)P(c_j)}{p(x)} \tag{4.5}$$

Hence,

$$R(\alpha_i|x)p(x) = \sum_{j=1}^{L} \lambda(\alpha_i|c_j)P(c_j)\, p(x|c_j) \tag{4.6}$$

The quantity $R(\alpha_i|x)p(x)$ is a scaled risk and may be tested instead of the risk $R(\alpha_i|x)$

without changing the outcome of the decision. The quantities $\lambda(\alpha_i|c_j)P(c_j)$ present, in effect, a bias toward the most likely classes and 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

$$\lambda(\alpha_i|c_j) = 1 - \delta_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases} \quad \text{all } i, j \tag{4.7}$$

which, when substituted in (4.4) yields

$$R(\alpha_i|x) = \sum_{j \neq i}^{L} P(c_j|x) = 1 - P(c_i|x) \tag{4.8}$$

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 (4.1) (*maximum a posteriori probability* (MAP) *classifier*). Further simplifying by setting $P(c_i) = 1/L$ for all i in (4.2) gives

$$(R(\alpha_i|x) - 1)\, p(x)\, L = -p(x|c_i) \tag{4.9}$$

and provides a classification rule based solely on the class-conditional probability density function. Equation (4.9) 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 x to class c_j if

$$g_j(x) > g_i(x), \quad \text{all } i \neq j \tag{4.10}$$

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

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

$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} exp[-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)] \tag{4.11}$$

where

$$\mu = E[x] \tag{4.12}$$

$$\Sigma = E[(x-\mu)(x-\mu)^t] \tag{4.13}$$

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

$$g_i(x) = ln[p(x|c_i)] \tag{4.14}$$

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i) - \frac{d}{2}ln(2\pi) - \frac{1}{2}ln|\Sigma_i| \tag{4.15}$$

and removing constants yields a new discriminant function

$$g_i'(x) = (x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i) + ln|\Sigma_i| \tag{4.16}$$

The change of sign requires $g_i'(x)$ to be minimized, that is, x is assigned to class c_j whenever

$$g_j'(x) < g_i'(x), \quad \text{all } i \neq j \tag{4.17}$$

The quantity $D_i^2(x) = (x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i)$ is also known as the squared *Mahalanobis distance* from x to μ_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 x, or the "volume" inside the multivariate distribution's equiprobability contours. Adding the term $ln|\Sigma_i|$ to $D_i^2(x)$ thus biases the discriminant toward the denser, lower variance classes.

Omitting this bias leads to the so-called *minimum intra-class distance* (MICD) classifier. This is the most general of a class of *minimum-distance* classifiers which use the notion of a metric to assign feature vectors to a class. In this interpretation the inverse of the covariance matrix, Σ_i^{-1}, can be viewed as the product of a transformation matrix G_i with its transpose. The linear transformation

$$y_i = G_i(x-\mu_i) \tag{4.18}$$

is designed to transform all samples x belonging to class c_i into samples y_i such that the mean squared distance between them (intra-class distance) is minimum [Jernigan, 1984]. This turns out to be an orthonormal whitening transformation: the features in y_i exhibit complete independence and are zero mean with unit variance. If the original vector x is assumed to have independent features with unit variance, the matrix G_i will

become the identity matrix and Equation (4.16) reduces to the *Euclidean distance* classifier

$$g_i''(x) = (x-\mu_i)^t(x-\mu_i) \tag{4.19}$$

The computational advantage which may be gained through the use of such simplified structures may make them attractive alternatives to the full Bayes or MAP classifier. It has been shown, however, that the performance loss is quite significant for the analysis of radar clutter [Stehwien, 1984]. This may be attributed to the non-separable behavior of the clutter features. In other words, no well defined boundary exists between classes: clutter characteristics are a continuum with considerable overlap amongst clutter types. Hence classification must proceed on a statistical basis for which the Bayes classifier is best suited.

The loss of performance expected and experienced when using minimum distance classifiers cannot be justified with the minimal computational gain obtained (see also Chapter 7): the classifier uses few resources compared to the proposed feature extractor. In fact, the largest computational burden lies in the training process which need not take place in real time. For the interested reader, [Fukunaga, 1972] presents an excellent overview of suboptimal linear classifiers.

Nonparametric Bayes classifiers make no assumptions about the form of the probability density function in (4.9). In principle, therefore, they offer an improved performance potential over and above the parametric classifier using the normality assumption (4.11). This is especially true for nonlinear feature extractors such as the Burg algorithm used in this work. Two popular nonparametric classifiers are the Parzen window density estimator, and the nearest-neighbor (or k-nearest neighbor) rule. The Parzen window approach uses labelled samples to compute an estimate of the true

density $p(x|c_i)$, which is then used in (4.9) directly. In the limit, as the number of samples approaches infinity, the estimate converges on the true density. The nearest neighbor rule examines the labelled samples closest to the new feature vector and assigns it to that class to which most of these k nearest samples belong.

Both approaches require virtually all the data, with which the classifier is to be trained, to be stored to assure good performance. This can lead to severe memory and computational requirements, especially for high dimensionalities and large numbers of classes. The vast quantities of radar data available and necessary for successful classification make these approaches difficult to manage. Consequently they were not investigated further. Comprehensive discussions on nonparametric classifiers can be found in [Fukunaga, 1972; ' uda and Hart, 1973; Young and Calvert, 1974; Patrick, 1972].

4.2 Classifier Performance Predictions

Predictors of classifier performance are useful in two ways. First they provide an indication of what sort of performance to expect in an operational environment: if complete separability of the classes cannot be achieved, is the classifier still of value? The second, and to the designer the most important use of predictors is as an aid to the selection of features and the design of the feature extractor.

The probability of making an error is, for most probability densities, a function of class separation. Hence measures of separability may be used to estimate error rates. Two will be discussed in this section; they and several others are also discussed in [Fukunaga, 1972; Lissack and Fu, 1976].

The conditional error rate was already shown above to be equivalent to the conditional Bayes risk using the zero-one loss function. The overall error rate is

$$P(\varepsilon) = \int P(\varepsilon|x)p(x)\,dx \tag{4.20}$$

Clearly,

$$P(\varepsilon|x) = 1 - P(c_i|x) = \sum_{j\neq i}^{L} P(c_j|x) \tag{4.21}$$

where c_i is the correct class. Using Bayes rule

$$P(\varepsilon) = \sum_{i=1}^{L} \int_{\mathbb{R}i} \sum_{j\neq i}^{L} p(x|c_j)P(c_j)\,dx \tag{4.22}$$

where the integral is over all sub-regions $\mathbb{R}i$ of the feature space wherever x is classified into class c_i. For the general multiclass case these regions are difficult to evaluate and no closed form solution to (4.22) is known to exist, even for the generally well behaved normal density. For the general two-class case, the problem is solvable but difficult and involves numerical integration [Fukunaga, 1972].

The probability of error for the two-class, equal covariance matrix case may be found by using log-likelihood ratios. The *divergence* J_{ij} is defined as the difference between the expected value of these ratios

$$J_{ij} = E\left[ln\frac{p(x|c_i)}{p(x|c_j)}\,|c_i\right] - E\left[ln\frac{p(x|c_i)}{p(x|c_j)}\,|c_j\right] \tag{4.23}$$

For $p(x|c_i)$ normal

$$J_{ij} = \frac{1}{2}(\mu_i-\mu_j)^t(\Sigma_i^{-1}+\Sigma_j^{-1})(\mu_i-\mu_j) + \frac{1}{2}tr(\Sigma_i^{-1}\Sigma_j+\Sigma_j^{-1}\Sigma_i-2I) \tag{4.24}$$

which, in the case of equal covariance matrices $\Sigma = \Sigma_i = \Sigma_j$, reduces to

$$J_{ij} = (\mu_i - \mu_j)^t \Sigma^{-1} (\mu_i - \mu_j) \tag{4.25}$$

This is the squared Mahalanobis distance between the two class mean vectors. The probability of error is computed as follows

$$P(\varepsilon_{ij}) = Q(\tfrac{1}{2}\sqrt{J_{ij}}) \tag{4.26}$$

where

$$Q(\alpha) = \int_{z=\alpha}^{\infty} \mathcal{N}(z{:}0,1)\, dz = \frac{1}{\sqrt{2\pi}} \int_{z=\alpha}^{\infty} exp(-z^2/2)\, dz = \frac{1}{2}\, erfc(\alpha/\sqrt{2}) \tag{4.27}$$

and $\mathcal{N}(z{:}0,1)$ is the normal density with zero mean and unit variance. This unique relation to the probability of error is valid only for the two-class equal covariance matrix case. For the unequal covariance case only bounds are available. From (4.22), the two-class probability of error can be written as

$$P(\varepsilon_{ij}) = \int_{\mathbb{R}} min[\, p(x|c_i)P(c_i), p(x|c_j)P(c_j)\,]\, dx \tag{4.28}$$

which, according to the inequality $min[a,b] \le \sqrt{ab}$, may be written as

$$P(\varepsilon_{ij}) \le \sqrt{P(c_i)P(c_j)} \int_{\mathbb{R}} \sqrt{p(x|c_i)p(x|c_j)}\, dx \tag{4.29}$$

The quantity

$$B_{ij} = -\ln \int_{\mathbb{R}} \sqrt{p(x|c_i)p(x|c_j)}\, dx \tag{4.30}$$

is known as the *Bhattacharyya distance*. The upper bound on the probability of error may then be written as

$$P(\varepsilon_{ij}) \leq \sqrt{P(c_i)P(c_j)}\, exp(-B_{ij}) \tag{4.31}$$

or, if $P(c_i) = P(c_j) = \frac{1}{2}$

$$P(\varepsilon_{ij}) \leq \frac{1}{2} exp(-B_{ij}) \tag{4.32}$$

The Bhattacharyya distance can be evaluated relatively simply if the densities are multivariate normal

$$B_{ij} = \frac{1}{4}(\mu_i-\mu_j)^t(\Sigma_i+\Sigma_j)^{-1}(\mu_i-\mu_j) + \frac{1}{2} ln \frac{\frac{1}{2}|\Sigma_i+\Sigma_j|}{|\Sigma_i|^{1/2}|\Sigma_j|^{1/2}} \tag{4.33}$$

It is interesting to note that, in the case of equal covariance matrices $\Sigma = \Sigma_i = \Sigma_j$, this reduces to

$$B_{ij} = \frac{1}{8}(\mu_i-\mu_j)^t\Sigma^{-1}(\mu_i-\mu_j) = \frac{1}{8} J_{ij} \tag{4.34}$$

and thus, using (4.26)

$$P(\varepsilon_{ij}) = \frac{1}{2} erfc\,(\sqrt{B_{ij}}) \tag{4.35}$$

A lower bound on the probability of error is instructive and may be found using the *Kolmogorov variational distance*

$$K_{ij} = \int_{\mathbb{R}} |p(x|c_i)P(c_i) - p(x|c_j)P(c_j)|\, dx \tag{4.36}$$

which is uniquely related to the probability of error as

$$P(\varepsilon_{ij}) = \frac{1}{2}(1-K_{ij}) \tag{4.37}$$

Noting that

$$\int_{\mathbb{R}} [\sqrt{p(x|c_i)P(c_i)} - \sqrt{p(x|c_j)P(c_j)}]^2 \, dx =$$

$$= P(c_i) + P(c_j) - 2\sqrt{P(c_i)P(c_j)} \int_{\mathbb{R}} \sqrt{p(x|c_i)p(x|c_j)} \, dx$$

$$= 1 - 2\sqrt{P(c_i)P(c_j)} \exp(-B_{ij}) \tag{4.38}$$

and that $|a - b| = |\sqrt{a} - \sqrt{b}|\,|\sqrt{a} + \sqrt{b}|$, the Schwartz inequality

$$\left[\int |\sqrt{a} - \sqrt{b}|\,|\sqrt{a} + \sqrt{b}|\right]^2 \leq \int (\sqrt{a} - \sqrt{b})^2 \int (\sqrt{a} + \sqrt{b})^2 \tag{4.39}$$

leads to

$$K_{ij}^2 \leq 1 - 4\,P(c_i)P(c_j)\,[\exp(-B_{ij})]^2 \tag{4.40}$$

Whenever $P(c_i) = P(c_j) = \frac{1}{2}$

$$K_{ij} \leq [1 - \exp(-2B_{ij})]^{1/2} \leq 1 - \frac{1}{2}\exp(-2B_{ij}) \tag{4.41}$$

From (4.32) and (4.37) it can be seen that $P(\varepsilon_{ij})$ is bounded by

$$\frac{1}{4}\exp(-2B_{ij}) \leq P(\varepsilon_{ij}) \leq \frac{1}{2}\exp(-B_{ij}) \tag{4.42}$$

Note that although these bounds diverge as B_{ij} becomes large, the probability of error at that point will be so small that, in practice, only the upper bound will be of interest. These bounds are plotted in Figure 4.1. The probability of error for the equal covariance matrix case (Equation 4.35) is included for comparison.

Up to this point, only two-class separability has been discussed. While this is useful when selecting classes and class groupings in the multiclass problem, an

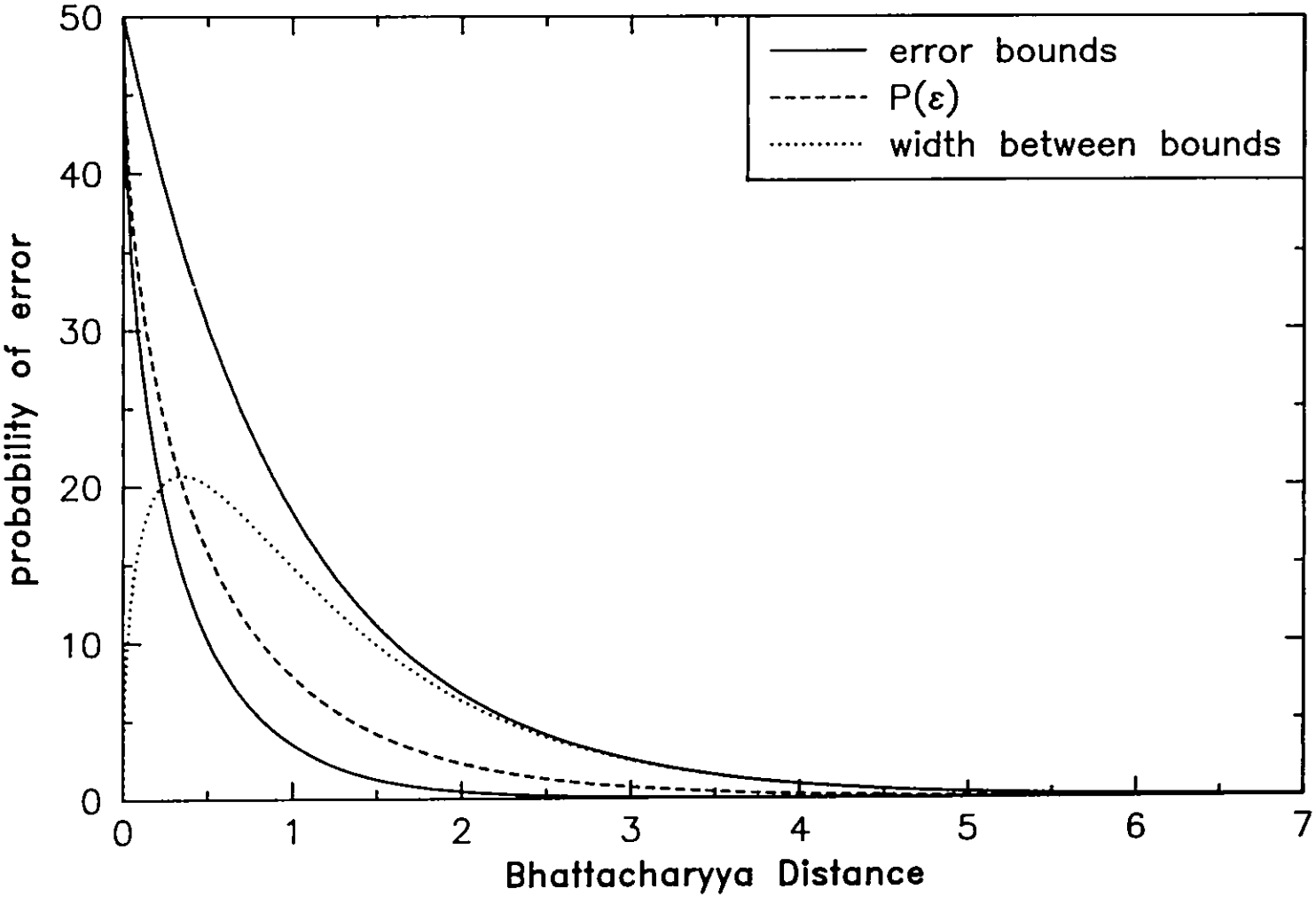


Figure 4.1 Bhattacharyya distances and error bounds

The upper bound is that given in equation (4.42); the lower bound is the tighter of the two arising from equation (4.41). The probability of error $P(\varepsilon)$ is based on equation (4.35).

overall error rate is desirable. Unfortunately, such a general measure has not been found to date. While an upper bound may be established

$$P(\varepsilon) \leq \sum_{i>j}^{L} \sum_{j=1}^{L} P(\varepsilon_{ij}) \tag{4.43}$$

this bound will invariably be much too conservative if the pairwise upper Bhattacharyya bounds are used. Even a class error rate bound, found by summing only those bounds belonging to a single class

$$P(\varepsilon_i) \leq \sum_{j \neq i}^{L} P(\varepsilon_{ij}) \tag{4.44}$$

will likely be so conservative as to be useful only for relative comparisons rather than true performance predictions.

It has been pointed earlier out that the divergence-based computation of a pairwise error rate using (4.26) is *not* valid for the unequal covariance case. It is not clear, however, what the deviation from the true value is if the calculation is nevertheless applied. It may be hypothesized here that the error is not great in practice, and that an approximate class error rate may in fact be computed using (4.24), (4.26), and (4.44) with approximate equality replacing the inequality. The validity of this hypothesis may be seen in the experimental results in Chapter 6. It is shown that the class error rates calculated in this manner are quite close to the experimentally observed rates.

The performance predictors discussed in this section provide not only the theoretical basis for the evaluation of experiments, but also the insight required to

improve performance. For example, Equations (4.24) and (4.33) show that increasing separability without changing the mean vectors μ_i requires the determinant of the covariance matrices to become small. The determinant is the product of the variances of the features after an orthogonality transformation. Variance reduction must therefore be an important consideration during the process of feature extraction. When adding a new feature, its variance must be carefully weighed against the distance between the class means in order to net a true increase in the divergence or the Bhattacharyya distance.

4.3 Assigning Confidence Measures

In addition to establishing the classifier performance, it is very desirable to know how good the decision of the classifier is at the time it is made. Obviously the a posteriori probabilities themselves are such a measure. Recalling Bayes rule

$$P(c_i|x) = \frac{p(x|c_i)P(c_i)}{\sum p(x|c_j)P(c_j)} \tag{4.45}$$

where the summation in the denominator is over all $j = 1$ to L. From (4.14) and (4.16)

$$p(x|c_i) = exp[g_i(x)] = exp[-\tfrac{1}{2} g_i'(x)]\ exp[-\tfrac{d}{2} ln(2\pi)] \tag{4.46}$$

Setting $P(c_i) = P(c_j)$, Equation (4.45) becomes

$$P(c_i|x) = \frac{exp[-g_i'(x)/2]}{\sum exp[-g_j'(x)/2]} \tag{4.47}$$

Only one exponential needs to be evaluated for each class from the discriminants computed for the classifier. It is then possible to report this probability, or to reject the sample x altogether if some arbitrary probability threshold is not exceeded. Chow [1970] discusses such a strategy and presents rejection rate versus error rate tradeoff curves. Clearly, any such rejection rule can only decrease the probability of correct classification obtained using the Bayes classifier. Hence the decision to reject samples simply to reduce the error rate must be made with care and with due consideration to the application itself. The strategy which will be used for the clutter classifier is to report low probabilities, but not reject any data on this basis.

There are, however, situations where rejection is in order. One concerns the reduction of the misclassification rate by rejecting data that lie too close to the decision boundary [Chow, 1970]. It is also entirely possible to define numerous class prototypes and still not account for the entire feature space. A new measurement may thus be an outlier with respect to all defined classes, but still lie much closer to one class than any of the others. Equation (4.47) clearly shows that a high a posteriori probability may still result, preventing the above strategy from rejecting the measurement. Instead, rejection should be based on the likelihood that the samples could come from the class itself. Such a test may be applied to the discriminant function $g_i'(x)$ directly. The squared Mahalanobis distance $D_i^2(x)$ can be shown to be χ_d^2 distributed with d degrees of freedom, if the probability density $p(\boldsymbol{x}|c_i)$ is normal, and both the mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$ are known [Young, 1978]. This knowledge immediately provides a significance test:

$$\textit{reject } \boldsymbol{x} \textit{ if}: \quad P[\chi_d^2 > D_i^2(x)] \; < \; T'' \tag{4.48}$$

where T'' is some application-specific threshold. A more practical, but equivalent rule is to test against a threshold T'

$$reject\ x\ if:\ \ D_i^2(x)\ >\ T' \tag{4.49}$$

where T' is such that $P[\chi_d^2 > T'] = T''$. This rule works well with the MICD classifier, but presents a difficulty for the multiclass Bayes or MAP classifier. In that case using (4.49) before evaluating $g_i'(x)$ may lead to rejection of the preferred class as determined by (4.17), and the selection of a less likely class. This is the result of adding the different biases $ln|\Sigma_i|$, and is illustrated in Figure 4.2. The sample x would be assigned to class c_1 when choosing the smallest discriminant $g_1'(x)$. In this case, however, the likelihood of the sample originating from c_1 is less than T''. In fact, according to the χ^2 test, the class which x appears most likely to have originated from is c_3, which exhibits the lowest Mahalanobis distance of all three. The answer to this apparent contradiction is that the χ^2 test simply gives the likelihood of x, given that it originated from c_i. It does not provide a comparison between classes in that such comparison is necessarily influenced by factors (such as the a priori probabilities and costs) which do not enter the χ^2 test. One must therefore apply caution before rejecting samples on the basis of (4.49) alone. One possible rejection rule could be applied to the discriminant itself

$$reject\ x\ if:\ \ g_j'(x)\ >\ T\ ,\qquad T\ =\ T'\ +\ max(ln|\Sigma_i|,\ i{=}1\ to\ L) \tag{4.50}$$

where $g_j'(x)$ belongs to c_j, which is the class selected according to (4.17). This ensures that rejection does not take place before condition (4.49) is simultaneously satisfied for

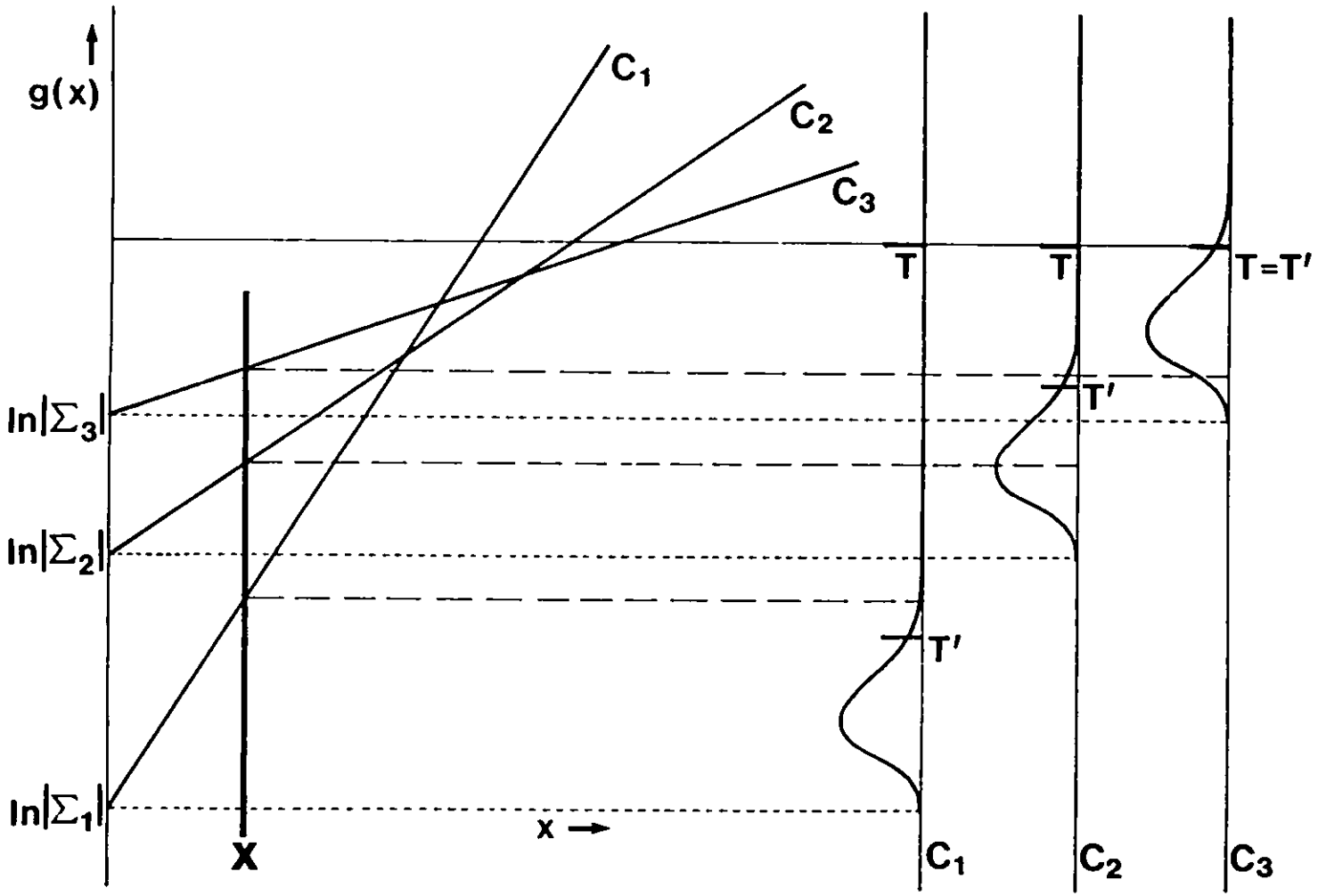


Figure 4.2 Illustration of the process of discriminant testing

The lines marked c_1 to c_3 represent the transformation of the feature vector x into the discriminant $g_i(x)$ according to equation (4.16). The χ^2 densities of $g_i(x)$ are shown at right.

all L classes.

4.4 Classifier Training

The solution to the problem of defining the parametric class prototypes is, perhaps, deceptively simple. All one needs to do, given a sufficient number of labelled samples, is to find the quantities μ_i and Σ_i in accordance with Equations (4.12) and (4.13), and then set up a classifier architecture to implement decision rule (4.17) and, if desired, rejection rule (4.50). It must be remembered, however, that this theory is based on *knowledge* of the underlying statistics. For most applications this knowledge is simply not available and one must be satisfied with the best available estimates. The questions then naturally arise: How good are these estimates? How well do they fit the assumptions used? What is the impact on the classifier? The answers are no longer simple, and this section will address them in the context of the clutter classification problem.

It has been mentioned previously that the observed clutter features are not likely to be separable, but rather present a continuum of characteristics with significant overlap amongst classes. Clustering techniques [e.g. Fukunaga, 1972], which attempt to collect all samples most likely to belong together, do not work well when non-separable behavior is encountered. Other unsupervised learning techniques [Duda and Hart, 1973], which attempt to divide the observed mixture density $p(x)$ into its most likely components are computationally complex and can never outperform supervised techniques. This is especially true for the radar clutter data whose mixture density is not conducive to simple separation. A starting set of labelled samples is often required and as such the trade-off appears to be one of computational complexity versus the task of assigning labels to samples. Clearly, if the samples can be labelled, then supervised

parameter estimation procedures are preferred.

Several such techniques are available, including maximum likelihood estimation and Bayesian learning. Maximum likelihood estimates of the mean vector and covariance matrix are well known [Anderson, 1984]

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x_n \quad \text{and} \quad \hat{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\hat{\mu})(x_n-\hat{\mu})^t \tag{4.51}$$

The estimate $\hat{\mu}$ is unbiased, whereas $\hat{\Sigma}$ is not. An unbiased estimate is the *sample covariance matrix C*

$$C = \frac{1}{N-1}\sum_{n=1}^{N}(x_n-\hat{\mu})(x_n-\hat{\mu})^t \tag{4.52}$$

Clearly, for large N the difference between these two estimates is small and may be ignored. Both estimates are consistent (they converge toward Σ in the limit as $N \longrightarrow \infty$) and either may be used successfully. The question of which one is better cannot be answered simply, and most certainly will depend on the application. The unbiased estimate leads to larger variances and thus projects less confidence about the "discovery" of dense clusters. The influence such "doubt" will have on the classification is complex and difficult to generalize, but it does reduce the rejection rate (4.50) and leads to more conservative error rate predications (4.26). For those reasons alone use of the unbiased instead of the maximum likelihood estimate appears justified.

Bayesian learning is another approach to parameter estimation. Here both $\hat{\mu}$ and $\hat{\Sigma}$ are considered random variables with probability densities of a known parametric form. The procedure is then to find the joint a posteriori probability density $p(\mu,\Sigma|\mathcal{X})$ given known or assumed forms for the a priori densities $p(x|\mu,\Sigma)$. $\mathcal{X}$ is the set

of available and labelled samples. Initial guesses μ_0 and Σ_0 are provided along with their measures of uncertainty. The parameters of the resulting densities turn out to be weighted averages between the guesses and the maximum likelihood estimates of (4.51); and the weighting is directly related to the uncertainties [Fukunaga, 1972]. The lack of any prior knowledge about the clutter distributions effectively rules out the Bayesian approach which, in any case, is equivalent to maximum likelihood estimation if the confidence in the initial guesses is deemed to be zero.

The need for labelled samples for classifier training is paramount. Clearly, good estimates can only be assured when the labels are substantially correct. An incorrectly labelled sample can contribute to a substantial distortion of the estimate away from its true value. It is possible to keep such distortions to a minimum by using large quantities of samples which contain, at most, only a very small percentage of incorrect labels. It is not possible, however, to definitively state how many such samples can be tolerated since they may exhibit similar statistics to those correctly labelled and thus have minimal impact. Yet, good performance can only be assured by keeping their numbers very small.

Assuming that all samples are correctly labelled, it is then natural to ask how many are required to ensure quality estimates, and more importantly, good classifier performance. The latter question is rather difficult to answer, especially for multidimensional, multiclass cases. No such measure has been found, and deriving empirical estimates using Monte Carlo techniques is extremely time consuming. The approach taken here is to verify classifier performance experimentally, and to use as many samples as is practical for the estimates. Nevertheless, some quantitative measure for the goodness of the estimates is helpful. Young [1978] describes such a measure. If the feature vectors x are distributed according to $N(x:\mu,\Sigma)$ and the

parameters $\hat{\mu}$ and C are computed per (4.51) and (4.52), the squared Mahalanobis distance estimate

$$\hat{D}^2(x) = (x-\hat{\mu})^t C^{-1}(x-\hat{\mu}) \tag{4.53}$$

may be transformed into the variate F

$$F_{d,N-d} = \hat{D}^2(x)\,\frac{1}{d}\left[\frac{N}{N+1}\right]\left[\frac{N-d}{N-1}\right] \tag{4.54}$$

which has a central F-distribution with $(d, N-d)$ degrees of freedom. This formulation also requires the random vector x to be independent of, but identically distributed as the set $\mathbb{Z}$ used to estimate $\hat{\mu}$ and C. From the properties of the F-distribution, when $N-d > 4$

$$E[\hat{D}^2(x)] = d\left[\frac{N+1}{N}\right]\left[\frac{N-1}{N-d-2}\right] \tag{4.55}$$

and

$$var[\hat{D}^2(x)] = 2d\left[\frac{N-2}{N-d-4}\right]\left[\frac{N+1}{N}\right]^2\left[\frac{N-1}{N-d-2}\right]^2 \tag{4.56}$$

It is immediately obvious that as N increases

$$\lim_{N \to \infty} E[\hat{D}^2(x)] = d = E[D^2(x)] \tag{4.57}$$

and

$$\lim_{N \to \infty} var[\hat{D}^2(x)] = 2d = var[D^2(x)] \tag{4.58}$$

which is reasonable since the consistent estimators $\hat{\mu}$ and C are expected to converge to the true values μ and Σ, and the mean and variance of a χ^2_d-distributed variate are known to be d and $2d$, respectively. This result then permits the computation of the

deviation of the mean of $\hat{D}^2(x)$ from the true mean d

$$E[\hat{D}^2(x)] = d\left[\frac{N+1}{N}\right]\left[\frac{N-1}{N-d-2}\right] = d(1+\beta) \tag{4.59}$$

hence

$$\beta = \left[\frac{N+1}{N}\right]\left[\frac{N-1}{N-d-2}\right] - 1 \tag{4.60}$$

The error βd represents $\beta\sqrt{d/2}$ standard deviations. This gives a measure of the goodness of the estimates $\hat{\mu}$ and C in the context of their use in the discriminant function $g_i'(x)$. Solving for N explicitly provides the minimum number of independent samples required to ensure the mean to be within 100 β percent of its true value

$$N \geq (1+\beta^{-1})(d+2) \tag{4.61}$$

For example, to maintain a 1% deviation from the mean, $\beta = 0.01$ and

$$N \geq 101d + 202 \tag{4.62}$$

This establishes a clear link between the dimensionality of the classifier and the number of independent samples required to maintain satisfactory performance. Consequently, increasing the number of features without the number of samples during classifier training does not necessarily bring about a performance improvement. In fact, if those new features do not add substantially to class separability, performance may actually decrease! This problem has been recognized previously [Duda and Hart, 1973].

4.5 Summary

This chapter presented several aspects of statistical pattern recognition theory, including a brief review of the parametric Bayes classifier. Pure distance based and linear classifiers were discussed and rejected for use with radar clutter data for reasons of inferior performance in return for a negligible reduction in computational requirements. Two error rate predictors were discussed as aids for the selection of separable features: bounds on the rates of misclassifications based on the Bhattacharyya distance, and an approximate multiclass error rate based on the concept of divergence. Even though the actual rates may deviate significantly from the predicted values, their use provides a quick means to assess trends in separability without the need for laborious classifier testing.

Knowledge of error rates does not, however, furnish any information about the quality of the individual decision itself. The actual a posteriori probability $P(c_i|x)$ can be computed relatively simply and can be used as an immediate confidence measure. It can also be used to reject the data sample for automatic classification if no class shows a clear dominance over all others. Sample rejection, or refusal to classify, may also be appropriate when the likelihood that the sample belongs to any of the classes is too small. Consequently, a rejection criterion based on testing the result of the discriminant function, which in turn is based on the χ^2_d-distributed Mahalanobis distance, was developed.

Finally, the important issue of classifier training was discussed. The importance of using as many labelled and statistically independent training samples as possible was stressed. To this end, an expression was given which relates the accuracy of the class prototype estimate, in terms of the expected value or mean of the Mahalanobis distance, to the number of independent samples used to arrive at this

estimate. This expression clearly shows that the number of samples required for a given accuracy is directly proportional to the dimensionality of the classifier.

Chapter 5

FEATURE SELECTION AND EXTRACTION

Before classification can proceed, the measurements must be transformed into a set of features in order to enhance class differences which 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. Linear transformations may not be successful in removing such information, and "optimal" feature extractors may perform quite poorly under these conditions. The first section of this chapter addresses this difficulty and the importance of matching the choice of features to the problem itself. The features proposed for use in clutter classification are described in Section 5.2, and their statistics are discussed in Section 5.3. The chapter concludes with a discussion on the problems of dimensionality and signal-to-noise ratio, both of which have a significant effect on classifier performance.

5.1 Features for Optimal Classification

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 which apply only to a specific group of problems [Kashyap, 1978]. These procedures may be optimal if those assumptions are correct, but when they are suspect one necessarily returns to ad hoc rules. In this thesis no claim regarding optimality is made. Heuristic reasoning is used extensively to arrive at a workable classifier. Due consideration is also given to implementation complexities. Ultimately only

experimental verification will supply the "proof of the pudding".

Radar data arrives in the form of a time series of finite length due to the finite dwell time on any one point in space (see also Section 6.1). Considerable work has been done on the classification of time series in fields such as speech recognition and the analysis of electrocardiograms (ECGs), among others. An "optimal" feature selection procedure for the use on time series is developed in [Kashyap, 1978]. The sampled data is assumed to originate from a stationary autoregressive (AR) process of order M

$$x_n = \sum_{i=1}^{M} \gamma_i x_{n-i} + w_n \qquad n > M \tag{5.1}$$

where w_n is normally distributed according to $N(w{:}0,W')$. Stationarity of the autoregressive process is ensured when the zeros of the polynomial

$$\gamma(z) = 1 - \sum_{i=1}^{M} \gamma_i z^i \tag{5.2}$$

lie outside the unit circle [Box and Jenkins, 1976]. The decision rule of (4.10) is used with

$$g_i(\bar{\gamma},\overline{W}') = -(\bar{\gamma}-\bar{\gamma}_i)^t S_i^{-1}(\bar{\gamma}-\bar{\gamma}_i)/2\overline{W}' + \beta_i ln(\alpha_i/\overline{W}') - \alpha_i/\overline{W}' + k_i \tag{5.3}$$

The features $\bar{\gamma}$ and $\overline{W}'$ are computed from the time series x_n as follows

$$\bar{\gamma} = [\Sigma x_{n-1} x_{n-1}^t]^{-1}[\Sigma x_{n-1} x_n] \tag{5.4}$$

$$\overline{W}' = \frac{1}{N}\Sigma[x_n - x_{n-1}^t \bar{\gamma}]^2 \tag{5.5}$$

where

$$x_{n-1} = [x_{n-1}, x_{n-2}, \cdots, x_{n-M}]^t \tag{5.6}$$

The summation range in expressions (5.4) and (5.5) is from $n = M+1$ to N. The quantities $\bar{\gamma}$ and $\bar{W}'$ can be recognized as the least squares estimates of the AR coefficients [Box and Jenkins, 1976] and the scaled prediction error power [Haykin and Kesler, 1979], respectively. The quantities γ_i, S_i, α_i, β_i, and k_i together form the class prototype for class c_i, and their estimates are computed from labelled sample time series. Details may be found in [Kashyap, 1978].

This classifier is computationally not as attractive as the simple quadratic form in (4.16). A major difficulty is the requirement for matrix inversion in real time, as well as the computation of a logarithm. Expression (5.3) is based on the assumption that the number of data samples N is large, in which case $\bar{W}'$ converges towards W', and $\bar{\gamma}$ towards γ. If N is small, however, a much more complicated decision rule must be used if performance is not to be degraded. Use of Equation (5.4) may in this case lead to nonstationary model estimates, that is $\bar{\gamma}$ may lie outside the bounded feature space specified by the assumption of stationarity in (5.2) [Markel and Gray, 1976]. Clearly, such estimates are subject to rejection as they will not fall within the class regions with high probability if those regions are accurately determined using large sample sets. Improved estimates for $\bar{W}'$ and $\bar{\gamma}$, such as those derived from forward-backward linear prediction (FBLP) and ones which ensure stationary estimates, may help in this regard. Computationally efficient algorithms are also available [Kay and Marple, 1981].

Another problem is the fact that the feature vector contains all information contained in x_n, some of which may hinder successful classification. This may sound

surprising, but if the identifying and non-identifying information are not linearly separable, nonlinear mapping techniques become necessary to allow parametric classification to proceed. To date, there exists no unified approach to the analysis of nonlinear feature extractors, and the question of how to separate or remove this information then leads once again to ad hoc rules [Young and Calvert, 1974; Fukunaga, 1972]. As will become clear later, the parameter vector γ does not lend itself readily to the required transformations.

There are, however, important similarities between the two classifiers. These similarities add some strength to the heuristics applied to the proposed feature extractor. In particular, Kashyap's derivation provides some justification for the use of the autoregressive parameters as features for time series classification. The equivalence between autoregressive and spectral analysis has been discussed by several researchers [van den Bos, 1971; Ulrych and Bishop, 1975; Jaynes, 1982]. The implication here is that the autoregressive model of (5.1) contains all measurable spectral information which is also present in the lattice filter reflection coefficients, the importance of which will be seen in the next section. In fact, there is a one-to-one correspondence between both sets of parameters [Haykin and Kesler, 1979].

What, then, are the most likely features of radar clutter which may lead to successful classification? As pointed out in Chapter 3, 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 pulse repetition frequency (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.

5.2 Selection of Separable Features

Previous work [Stehwien, 1983] established the utility of Burg's *reflection coefficients*, arising from the maximum entropy method (MEM) of spectral analysis [Burg, 1967], for the extraction of spectral features. The coefficients arise out of the lattice implementation of the prediction error filter (PEF), which attempts to minimize the prediction error power at each stage [Burg, 1968; see also Markel and Gray, 1976; Haykin and Kesler, 1979]. 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. Burg dealt only with real valued coefficients derived from real valued data, the complex form of the algorithm is given in [Haykin and Kesler, 1976]:

$$\rho_m = \frac{-2\sum_{n=m+1}^{N} f_n^{(m-1)} b_{n-1}^{*(m-1)}}{\sum_{n=m+1}^{N} [\,|f_n^{(m-1)}|^2 + |b_{n-1}^{(m-1)}|^2]} \tag{5.7}$$

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*

$$f_n^{(m)} = f_n^{(m-1)} + \rho_m b_{n-1}^{(m-1)} \tag{5.8}$$

$$b_n^{(m)} = b_{n-1}^{(m-1)} + \rho_m^* f_n^{(m-1)} \tag{5.9}$$

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

$$f_n^{(0)} = b_n^{(0)} = x_n \tag{5.10}$$

This, then, forms the basis of the feature extractor. Note that the prediction error filter coefficients are related to the reflection coefficients via the recursive relation

$$h_i^{(m)} = h_i^{(m-1)} + \rho_m h_{m-i}^{*(m-1)} \qquad i = 1, 2, \ldots, m \tag{5.11}$$

which is also known as *Levinson's recursion*. Note also that

$$h_0^{(m)} = 1 \qquad \text{and} \qquad h_m^{(m)} = \rho_m \tag{5.12}$$

The equivalence between MEM and autoregressive modeling has been established by

van den Bos [1971] and leads to the following relationship between the PEF coefficients and the M^{th}-order AR model parameters of (5.1)

$$\gamma_i^{(M)} = -h_i^{(M)} \tag{5.13}$$

The last coefficient of an autoregressive model is also known as the partial autocorrelation coefficient (PARCOR) [Box and Jenkins, 1976]. Hence, by relations (5.12) and (5.13), the reflection coefficients are equal to the negative of the partial autocorrelation coefficients

$$\rho_m = -\gamma_m^{(m)} \tag{5.14}$$

This useful equivalence establishes many of the properties of the reflection coefficients. Stability of the prediction error filter requires, for instance, that the magnitude of ρ_m must be less than unity. This is also the condition for stationarity of the AR model, which is ensured by Burg's estimate of (5.7). Another property of interest is the true value of ρ_m when the time series x_n represents samples of a complex sinusoid of amplitude A in white noise with power W

$$x_n = A\ exp(jn\phi) + w_n \tag{5.15}$$

where ϕ is the frequency of the sinusoid normalized to the sampling frequency f_s. As shown in Appendix D, this value is

$$\rho_m = -\frac{n_0}{1 + mn_0} exp(jm\phi) \tag{5.16}$$

where n_0 is the signal-to-noise ratio A^2/W. Result (5.16) has two important consequences. The first is that frequency translation of a time series simply rotates the resulting reflection coefficients without affecting their magnitudes and relative phase angles. If the assumption, that the measured signal is closely related to the sinusoid-in-noise model, is correct, the center frequency may be taken from the phase angle of ρ_1. Even if the model is not valid, the phase angle reflects a weighted average of all frequency components present (see also Appendix D)

$$\rho_1 = \sum_{k=1}^{K} q_k \rho_{1_k} \tag{5.17}$$

where

$$\rho_{1_k} = -\frac{n_{0_k}}{1 + n_{0_k}} exp(j\phi_k) \quad \text{and} \quad q_k = \frac{A_k^2 + W}{R(0)} \tag{5.18}$$

A_k and ϕ_k represent the magnitude and frequency of the complex sinusoids present, and $n_{0_k} = A_k^2/W$. This "average" frequency can then be removed from the coefficients by rotation, allowing the comparison of various spectra to be based on their shape alone

$$\phi = arg(-\rho_1) \tag{5.19}$$

$$\rho_m' = \rho_m exp(-jm\phi) \tag{5.20}$$

The second important consequence of (5.16) is the relationship of ρ_m with the signal-to-noise ratio n_0. It shows that the reflection coefficients are *not* independent of the signal strength unless the clutter process itself can be considered pure autoregressive and thus not affected by the scale of measurement. If this were true, the spectral equivalence of AR modeling requires that the spectral shape and the dynamic range remain unaffected

by signal strength changes. The residuals of the PEF must then be interpreted as a driving function of the process in (5.1) with power W', and not as the measurement noise of (5.15) with power W. The implication of the AR interpretation is that, with negligible measurement noise, power must be present at *all* frequencies in the spectrum for the model to be stable and positive definite. In reality, it is unlikely that the clutter echoes exhibit such a broad spectrum, unless aliasing is severe. Hence any power that does show up at the minima in the spectrum must be due to the white measurement noise, and the interpretation of the residuals as a pure driving function is indeed not valid. Further, if the absolute amount of measurement noise is fixed, the shape of the spectrum and its dynamic range *must* change as the signal strength changes. This is also evident from Equation (5.21), which relates the prediction error power W' to the mean signal power P_0 [Burg, 1968]

$$W' = P_m = P_{m-1}(1-|\rho_m|^2) = P_0 \prod_{i=1}^{m} (1-|\rho_i|^2) \tag{5.21}$$

This relationship shows that in the pure AR interpretation, where the model is independent of the scale of measurement, the prediction error power W' changes with the signal strength, and thus cannot possibly represent the measurement noise power W. Instead, if spectral energy is not present at all frequencies then W' must equal W for sufficiently large m, regardless of the value of P_0. Hence the dynamic range and the best fitting AR model will also change with signal strength, complicating the use of γ_i or ρ_m as features for the classifier.

The dependence of ρ_m on signal strength suggests that the signal power itself should be determined, if for no other reason than to bring the comparisons to an equal basis. The relative power is simply

$$P_0 = R(0) = E[x_n x_n^*] \approx \frac{1}{N} \sum_{n=1}^{N} x_n x_n^* \tag{5.22}$$

If the measurement noise W is known, then the signal-to-noise ratio can be simply determined. Since $P_0 = P + W$, where $P = A^2$, the relative power of the sinusoid

$$n_0 = \frac{P}{W} = \frac{P_0}{W} - 1 \tag{5.23}$$

Recall that Kashyap's feature extractor uses W' directly (Equation 5.5). Also recall that the value calculated using Equation (5.21) will be equal to the measurement noise power only if the autoregressive model order is sufficiently large. The difference between W' and W is thus a measure of the "goodness of fit" of the model to the data. Since W is known relative to P_0, it is clear that this "fit" information is completely contained in the right hand side of (5.21). Hence the use of P_0 in the feature extractor is equivalent to the use of W'.

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 pulse repetition frequency (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 $s = |x|$ is such a measure

$$P_{var} = \frac{E(s-\bar{s})^2}{E(s)^2} = \frac{P_0 - \bar{s}^2}{P_0} = 1 - \frac{\bar{s}^2}{P_0} \tag{5.24}$$

A feature to aid separation of distributed clutter and 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_O) will yield a feature which can distinguish between convex and concave signal shapes

$$P'_{dif} = \frac{P_I - P_O}{P_0} = \frac{N}{N_O}\left[\frac{P_I}{P_0} - 1\right] \tag{5.25}$$

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

$$P_{dif} = \frac{P_I}{P_0} - 1 \tag{5.26}$$

This completes the separable feature set which may be extracted from the radar data. To summarize, these features are

ϕ mean Doppler frequency

P_0 mean signal power relative to the known measurement noise W

ρ'_m heterodyned reflection coefficients: zero mean Doppler frequency

P_{var} normalized variance of the amplitude distribution in measurement window

P_{dif} normalized power difference for window center and edges

At this point it would be possible to train the classifier and to attempt to classify. This was in fact done with a subset of these features [Stehwien, 1984]. While a first glance the results look good, there is no assurance that they can be successfully extrapolated to more complete, and varied, data sets. In fact, the suspicion is that the success was more a result of the limited size of the data sets, that is local phenomena were found and tested against themselves. This problem of testing on the training data has lead many researchers to report overly optimistic results [Foley, 1972; Nagy, 1983]. Some considerable thought must therefore be given to avoid this difficulty. It is instructive, therefore, to examine the statistics of these features before designing a workable classification (class assignment) strategy.

5.3 Feature Statistics

Fundamentally, there are two causes for variations in the feature vector, given that it has been taken from the same class: measurement error and estimation errors, and the intrinsic characteristics of the features themselves. Once a measurement has been taken and the features computed, it is difficult to separate these causes and any examination of the statistics will necessarily have to consider their combined effect. Some general comments regarding the nature and dominance of each effect may, however, be made. The variance of low order features, which includes all those directly computed such as signal power and Doppler frequency, will be dominated by

intrinsic characteristics, and measurement and estimation errors are expected to be relatively small. High-order information contained in indirectly derived features such as the high-order reflection coefficients is, on the other hand, increasingly obscured by measurement and estimation errors. The reason is that, in their derivation, these features are subjected to all errors already contained in those from which they are derived.

In this section the combined statistics of the proposed feature set are thus examined. For the most part, if independence or at most linear dependence between features is assumed, looking at each feature separately will suffice. There is a potential danger in this: the parametric classifier of Chapter 4 is based on independent, normally distributed features. Residual dependence, or deviation from multivariate normality, can significantly affect performance if the classes of interest are not linearly separable, or exhibit significant overlap. Linear dependence between features is effectively removed in the orthonormal whitening transformation inherent in the computation of the Mahalanobis distance. Nonlinear dependence, however, cannot be removed this way. Hence, looking at the marginal statistics alone may only hide nonlinear dependencies which must be removed before classification can succeed. Multivariate statistics are, however, much more difficult to examine, especially if the number of dimensions is large. Many tests for multivariate normality have been proposed, but most are somewhat unsatisfying in that only a few aspects of normality are tested. Again, ad hoc rules must necessarily be used.

5.3.1 Marginal Normality

The first feature of interest is the Doppler frequency ϕ as computed from (5.19). If the clutter or target travel vectors are bivariate normal with zero mean

(which is reasonable if the direction relative to the radar is random), then the absolute Doppler frequency or range rate will be a marginal normal variate. Unfortunately, aliasing destroys this normality by folding over frequencies which exceed one half the sampling rate f_s. In fact, it can be shown that, if the sampling frequency is less than or equal to the standard deviation of ϕ, then the resulting distribution will be uniform between the limits $\pm f_s/2$. Approximate uniformity will still be observed if the sampling frequency is outside one but within two standard deviations. With the exception of ground clutter, it may therefore be prudent to assume a uniform distribution for all moving clutter unless aliasing is not a factor.

This, of course, prevents the Doppler frequency from acting as a useful feature except for the identification of ground clutter. Should it then be included in the classifier at all? The answer is not clear at this stage. It may well be that ground clutter may identifiable using Doppler frequency alone, in which case preselection before classification (or multistage classification) may well be in order. In other words ground clutter is eliminated immediately, as is done in moving target indicator (MTI) and moving target detector (MTD) algorithms. Only clutter which "breaks through" these filters is then classified. There is a penalty, of course, in the rejection of zero Doppler moving targets. This leads to blind speeds and the loss of tangentially moving targets in MTI systems.

Inclusion of the Doppler frequency in the classifier instead of preselection will prevent this problem, but presents another difficulty. Traditional radar target detection is a two-class problem where targets of interest must be separated from unwanted clutter. To maintain control over "false alarms" or detection errors, a *Neyman-Pearson* test is usually employed. The detection threshold is set such that the probability of false alarm P_{FA} is a fixed value, and the probability of *not* detecting the

target ($1-P_D$) is accordingly minimized [Fukunaga, 1972; Young and Calvert, 1974]. This technique is also known as *constant false alarm rate* (CFAR) processing. The Bayes rule differs from the Neyman-Pearson rule in that neither error probability is fixed and both are jointly minimized. It is common to set constant false alarm rates at 10^{-3} to 10^{-6}, whereas error probabilities for the Bayes classifier are usually not better than 10^{-2} (one percent). Large amounts of ground clutter misclassified as moving targets can thus cause a much larger false alarm rate than radar operators are used to. Depending on which class is selected by the classifier, this may be a highly undesirable situation and should be avoided. This problem is explored further in Chapter 6.

The signal power for most clutter classes has often been found to be approximately log-normally distributed [Long, 1975]. Since the measurement in (5.21) is linear in power, a conversion to decibels, or some scaled version thereof, is appropriate. If the measurement noise is known, P_0 may be transformed to decibels relative to W

$$U_0 = 10\, log(P_0/W) \tag{5.27}$$

U_0 will then replace P_0 as a feature.

The asymptotic statistics of the reflection coefficients, which are due to measurement and estimation errors, have been examined by a number of authors. Kay and Makhoul [1983] derive asymptotic probability densities for real valued coefficients as the number of data samples N, used in the estimation of $\hat{\rho}_m$, becomes infinite. They show that

$$var(\hat{\rho}_m) = \frac{1}{N}(1-\rho_m^2) \quad \text{for } m \geq M \tag{5.28}$$

where M is the order of the underlying autoregressive process. The variances and

covariances of $\hat{\rho}_m$ for $m < M$ are also given but are highly dependent on the model order and coefficient values themselves. The trend remains, however, and low order coefficients whose values are close to 1 have a very low variance. Note, of course, that ρ_m is zero for $m > M$ and the variance of their estimates thus remains steady at $1/N$. Another interesting result from Kay and Makhoul's work is that they find $\hat{\rho}_M$ to be uncorrelated with all $\hat{\rho}_m$ of orders $m < M$, which suggests that an AR order selection procedure might test for zero covariances. Of course, a test for model order using confidence limits based on the $1/N$ variance around the zero mean has long been used in time series analysis [Box and Jenkins, 1976]. Unfortunately, Box and Jenkins do not say how large N must be relative to M for this test to be valid, and Kay and Makhoul use sample sizes between 100 and 1000 in the experimental verification of their results. Experimental observations with smaller sample sizes suggest, however, that the variances are more closely related to $1/(N-m)$, especially as m approaches N. Durrani and Arslanian [1983] derive exact probability densities for noise-alone and sinusoid-in-noise cases without the requirement for $N \longrightarrow \infty$. Their results show that the behavior of the reflection coefficients begins to deviate from normal as they move away from zero. The experimentally observed dependence on the quantity $\ell = N-m$ is also shown in the following derivation. Durrani and Arslanian give the probability density function of the estimated real valued reflection coefficient, whose expected value is zero, as

$$p(\hat{\rho}_m) = \frac{\Gamma[(\ell+1)/2]}{\sqrt{\pi}\,\Gamma[\ell/2]}(1-\hat{\rho}_m^2)^{(\ell-2)/2} \tag{5.29}$$

To show that this density approximates the normal density when ℓ becomes large, compare the two following infinite series expansions

$$(1-x^2)^q = 1 - qx^2 + q(q-1)x^4/2! - q(q-1)(q-2)x^6/3! + \ldots \quad (5.30)$$

$$exp(-qx^2) = 1 - qx^2 + q^2x^4/2! - q^3x^6/3! + \ldots \quad (5.31)$$

Also note the identity of the following limit

$$\lim_{z \to \infty} \frac{\Gamma(z+a)}{\Gamma(z)} = z^a \quad (5.32)$$

Then, setting the following equivalencies

$$x = \hat{\rho}_m , \qquad q = (\ell-2)/2 \qquad \text{and} \qquad z = \ell/2 , \qquad a = 1/2$$

Equation (5.29) can be rewritten as, for large ℓ

$$p(\hat{\rho}_m) = \frac{1}{\sqrt{2\pi/\ell}} exp[-\frac{1}{2} \frac{\hat{\rho}_m^2}{1/(\ell-2)}] \quad (5.33)$$

Except for a scale factor, the functional form of this probability density is normal with a variance of $\sigma^2 = 1/(\ell-2) = 1/(N-m-2)$. The exact expression for the variance of (5.29), without the requirement of large ℓ, can be found as follows

$$var(\hat{\rho}_m) = \int_{-1}^{1} \hat{\rho}_m^2 \, p(\hat{\rho}_m) \, d\hat{\rho}_m \quad (5.34)$$

$$var(\hat{\rho}_m) = \frac{\Gamma[(\ell+1)/2]}{\sqrt{\pi}\,\Gamma[\ell/2]} \int_{-1}^{1} \hat{\rho}_m^2 (1-\hat{\rho}_m^2)^{(\ell-2)/2} \, d\hat{\rho}_m \quad (5.35)$$

The bounds on the integral arise from the requirement that $|\hat{\rho}_m| < 1$ for a positive definite spectrum and stable AR model. The integral can be recognized as the Beta function

$$\int_{-1}^{1} \hat{\rho}_m^2 (1-\hat{\rho}_m^2)^{(\ell-2)/2}\, d\hat{\rho}_m = B(3/2,\ell/2) = \frac{\Gamma[3/2]\Gamma[\ell/2]}{\Gamma[(\ell+3)/2]} \tag{5.36}$$

Therefore, using the functional relation $\Gamma(x+1) = x\Gamma(x)$, Equation (5.35) becomes

$$var(\hat{\rho}_m) = \frac{\Gamma[(\ell+1)/2]\Gamma[3/2]}{\sqrt{\pi}\,\Gamma[(\ell+3)/2]} = \frac{2\ \Gamma(3/2)}{\sqrt{\pi}\ (\ell+1)} = 1/(\ell+1) \tag{5.37}$$

The variance $\sigma^2 = 1/(N-m+1)$ can thus be seen to be a function of both N and m as was experimentally observed. The above analysis deals, strictly speaking, only with real valued reflection coefficients, although the statistics of complex valued coefficients are not expected to differ greatly. To some extent, it justifies the use the reflection coefficients directly in the parametric MAP classifier. But most important, it links the classifier performance with the number of samples used to estimate the feature vector. The importance of variance reduction was pointed out in Section 4.2, and this link establishes the way it must be done, at least for those features which are affected by large estimation errors.

As the expected values of the coefficients become increasingly non-zero and approach their bound at unity, the normality requirement becomes more difficult to satisfy, especially since the variance due to estimation errors decreases (Equation 5.28) and that due to intrinsic characteristics begins to dominate. The log-normal characteristic of the clutter amplitude and its relationship with the signal-to-noise ratio allows ρ_1 to be transformed into an approximately normal distributed feature. Recall

that the feature set uses the rotated coefficients ρ_m' from (5.20). The effect of the rotation on ρ_1 is to set its phase angle to π, leaving only the magnitude

$$\rho_1' = -\frac{n_0}{1 + n_0} \tag{5.38}$$

The feature based on ρ_1 is then the derived signal-to-noise ratio, expressed in decibels

$$u_0 = 10\, log(n_0) = 10\, log\left[\frac{|\rho_1'|}{1 - |\rho_1'|}\right] \tag{5.39}$$

It is very unlikely that the magnitude of the complex ρ_1 will ever be zero, even in the noise only case when it follows a Rayleigh distribution, and the above transformation should therefore not present any computational difficulties. It is interesting to compare this transformation to those proposed by Markel and Gray [1976], and one used by Agnel [1985]

$$g_i = log\left[\frac{1 + k_i}{1 - k_i}\right] \tag{5.40}$$

where the k_i are the real valued reflection coefficients arising from real valued data. The addition of 1 in the numerator assures a positive quantity inside the logarithm. This addition is not required in (5.39) since only positive magnitudes are involved. Such a transformation cannot, however, be applied to complex coefficients.

From Kay and Makhoul's work [Kay and Makhoul, 1983] it is also clear that a linear transformation will not likely be able to remove the correlations between the coefficients of the different lattice stages, even if these are due only to known

measurement and estimation errors. Nonlinear or heuristic techniques (or both) may therefore be required for successful classification.

5.3.2 Multivariate Normality

Assuring marginal normality does not, by any means, guarantee multivariate normality. In fact, nonlinear dependencies can easily be completely hidden to univariate tests for normality. Various graphical and significance tests for assessing deviations from multivariate normality have been proposed. General probability plotting methods for univariate statistics are described in [Wilk and Gnanadesikan, 1968], and include quantile and percentage probability plots. Plotting the squared Mahalanobis distance D_i^2 against the expected order statistics from the χ_d^2 distribution has been suggested in [Cox and Small, 1978]. Andrews [1972] plots the multivariate data directly using a function $f_x(\omega)$ which is effectively the real part of the complex polynomial $H_x(z)$ evaluated along the unit circle $z = exp(-j\omega)$, setting the complex coefficients h_i equal to the data (that is, $h_i = x_{2i} + jx_{2i+1}$). The similarity of this approach with the plotting of the power spectrum is evident, although each plot will necessarily emphasize different characteristics of the data. While it may be interesting to examine the complete and transformed feature set this way, this technique is not likely to provide new insights into the classification problem and the multivariate nature of the data, beyond what may be obtained from the spectra themselves.

Cox and Small suggest that scatter diagrams will always be required for interpretation of the data, and this may be a useful starting point. Such diagrams will reveal bivariate dependencies, and any observed curvature is an indication of non-normal behavior. Cox and Small also discuss several significance tests for detection of such nonlinearities. Young's test based on the χ^2 distribution [Young, 1978] can also

be used to assess multivariate normality, although the use of only one global statistic calls for caution: while rejection of the null hypothesis (that the probability density is normal) means that non-normal behavior has been detected, acceptance does *not* necessarily mean that the density is normal. References to a variety of other significance tests can be found in [Fukunaga and Flick, 1986].

Any test for multivariate normality must, unless its purpose is solely an exercise in mathematics, consider the reason for its application. Often complete normality is not required, and some deviation is quite tolerable. A test for local normality using a window function around the region of interest has been proposed by [Fukunaga and Flick, 1986]. This test does require a priori knowledge of the mean vector and covariance matrix and therefore has limited applicability. If good estimates for these parameters cannot be obtained, the results may be invalid.

Once non-normal behavior has been "discovered", there remains the question of what to do next. If nonparametric, or *distribution free* classifiers can be used then perhaps this is the road one should take. Unfortunately, the volumes of data encountered in radar make this approach difficult to implement and the data must somehow be molded to become acceptable for use with the parametric classifier. Several authors have suggested nonlinear power transformations on each of the marginal variates and have claimed reasonable success with various (limited) data sets [Draper and Cox, 1969; Andrews, 1971]. In fact, except for the logarithm, Equation (5.39) is a special case of such a transformation. Application of these transformations to the other features may be useful, although significant correlations between variates are not likely removable with this technique.

The degree to which deviations from normality can be tolerated in the classification problem is not easy to define. Clearly, the most desirable behavior is

close clustering around a well defined mean for each class of interest. If the clusters are completely separable, the precise nature of the feature distribution is not important since no misclassifications will result. In the case of overlapping classes, the degree of overlap is to be minimized; hence, short tails in the distribution are desirable, again regardless of the nature of the distribution. If the overlap is severe, then the positioning of the mean becomes important to allow overall minimization of the error rate. In this case the emphasis moves to the center of the distribution where the majority of feature samples are found. The goodness of the mean and covariance estimates are a function of the functional form of the density and therefore sensitive to any deviation from normality. The tails are now less important, although they still play a role in achieved error rate and the estimation of the means and covariances. In the case of particularly long tails it is possible to censor outliers before the estimation, although the results then become a function of the censoring procedure itself. As will be seen later, the clutter classes are not well defined and far from separable. In fact, the features turn out to be non-normal to such a degree that rather ad hoc solutions are required to achieve tolerable error rate. Part of the solution lies in the definition of the classes themselves, rather than on simply assuring normality.

5.4 Dimensionality

One question remains to be answered regarding the feature set proposed in Section 5.2: 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 adding additional independent features will increase separability, there is a very real problem with adding arbitrarily many features,

even if there are no computational constraints. This problem has been recognized in the literature [Hughes, 1968; Duda and Hart, 1973], and can generally be attributed to a lack of sufficient data for the estimation of the class prototypes. Young's result [Young, 1978], as reflected in Equations (4.59) to (4.61), yields a clue to the amount of data required or, conversely, how many dimensions can be tolerated given the number of available training samples N and a maximum allowable error β. This does not, however, solve the problem entirely.

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 who shows that, unless the statistical models are comparable, predicting a decrease in the Bayes risk with increasing complexity is not justified [van Campenhout, 1978]. 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 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 towards measurement errors and interfering noise. As pointed out in Section 5.3, high-order coefficients contain the generally weaker, higher-order information and are estimated using fewer samples. Hence their variance is greater and their individual separability is 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 (see Appendix E for an examination of the effects of quadrature

errors on the spectrum and the reflection coefficients). 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 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. It has already been stated that the clutter processes represent a continuum in the feature space, and that 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. The classifier will then proceed in two steps, first to identify the region within which the sample feature falls using the Bayes discriminant (4.16, 4.17), and second to identify the class to which the region belongs. Even in such a scenario, the intrinsic dimensionality of the irregular cluster must be considered when choosing the number of features to use.

5.5 Signal-to-Noise Ratio

The dependence of the spectrum and hence the reflection coefficients on the signal strength has already been pointed out. In light of the above discussions on

multivariate normality and dimensionality, it is worthwhile to reexamine this dependence and possible solutions to it. There is sufficient evidence that the early classification results [Stehwien, 1984] were in fact dominated by this dependence. The reason for the apparently successful identification was that most of the observed measurements stayed within a fairly narrow range of SNRs, which was different for every class and determined by the way the measurements were taken. Application to new data, which was measured differently, caused massive failure of the classifier. The fact that SNR information was included in the features was clear, however, that it should have such a significant impact on classifier performance was not appreciated.

Signal strength is, in fact, only a function of the measurement itself and contains little useful information about the clutter source. Only absolute radar cross-section (RCS) measurements may be viewed as a feature, but these are generally not derivable from surveillance radar data. Except for perhaps an upper limit on received signal strength for certain clutter types (for example, rain: see Appendix B), the SNR information in the feature set has proved to be detrimental, rather than helpful. The following discussion should serve to illustrate the problem.

As long as any clutter type can take on any signal strength, every class cluster in feature space will wind its way from the common low SNR *noise* cluster to its own high SNR *signal* cluster. Consequently, separability is also expected to be a function of SNR, which is clearly a reasonable conclusion. If these clusters are nothing more than elongated hyperellipsoids, then the parametric Bayes classifier should provide the optimum decision rule and all that would be required is to train it with sufficient data from all SNR ranges. This would prevent the class prototypes from representing only "local" phenomena. Recall the earlier discussion (Section 5.2) on the validity of the AR representation for the clutter data. A similar argument may be made

in reference to the modeling results on the signal itself, rather than only the dynamic range. Specifically, the shape of the spectrum of a non-fluctuating point target, which is scanned by a Gaussian shaped beam, is also Gaussian. Such a shape cannot easily be represented by an all pole spectrum, and techniques such as the MEM try to do the best they can by appropriately placing the spectral poles close to the true spectrum. As the dynamic range of the spectrum increases, spectral peaks become narrower and more are required for adequate representation. Hence the order of the best fitting AR model increases as the SNR increases. The fact that the first reflection coefficient is a linear combination of all observed frequency components (see Appendix D), implies that not only does the AR order increase with SNR, but the low-order coefficients also change. Hence the feature cluster cannot be considered a simple hyperellipsoid. This recognition suggests that the two step classification approach proposed in Section 5.4 may be appropriate.

A possible approach to the SNR problem has now emerged. It is clear that unless the nonlinear dependence of the reflection coefficients on SNR can be removed (it appears unlikely that this can be done without significant loss of other information), classification must occur with due consideration to the SNR. Several classifiers are constructed, one per SNR range. Both training samples and new measurements are directed to the appropriate classifier and the decision outputs recombined. Clearly, hardware constraints and the availability of training data will determine exactly how many classes and classifiers can be defined.

5.6 Summary

In this chapter so-called optimal features for time series were examined and found to be unsuitable, since much of the information contained in the data does not

contribute to separability. Linear methods may not be able to remove such information, and heuristic arguments and experimental experience were applied to the selection of practical features for the radar clutter classifier. These features were found to be the mean Doppler frequency, the mean signal-to-noise ratio, the reflection coefficients derived from the heterodyned time series, and two measures of the distribution of signal power inside the measurement window.

Both marginal and multivariate statistics of these features were examined next. The probability densities of the low order features were determined to be dominated by the intrinsic characteristics of the clutter itself: uniform for the Doppler frequency and log-normal for the signal-to-noise ratio. The asymptotic densities of zero-mean reflection coefficients were shown to be approximately normal, but nonzero coefficients, who are influenced considerably by the intrinsic clutter characteristics, could not be characterized in simple terms. Multivariate statistics are not expected to be normal, and nonzero covariances have been noted amongst the various reflection coefficients. The problem of testing for multivariate normality and how to achieve it in order to allow the use of the parametric classifier was also discussed.

The problem of dimensionality is closely related to classifier performance, and the desire to increase it simply to improve performance, and without examining the additional information (or lack thereof), was cautioned against. Finally, the role of the signal-to-noise ratio in the feature extractor was explored. The expected strong and highly nonlinear correlation of most features with the SNR, and the difficulties involved in removing these correlations, will most likely require ad hoc solutions such as the construction of one classifier per SNR range.

Chapter 6

EXPERIMENTAL RESULTS

This chapter presents the results of experimental application of the theory discussed in the previous three chapters. The description follows a logical progression starting with the process of collecting and preprocessing the data, transforming it into the required feature set, training the classifier, and finally, testing the classifier with the data. Any meaningful analysis and experimental verification requires large volumes of data, and considerable effort was spent obtaining them. For the most part, this effort was successful, and considerable confidence can be placed in the results presented in the following sections. On the other hand, dealing with such enormous data volumes places some restrictions on what can be tried within the scope of a limited time research project. As a result, some possibly worthwhile algorithms were not tried and the emphasis was placed on achieving the best results with the greatest simplicity. An unfortunate side effect of this approach is that there may well be other techniques available (or become available as research progresses) which would achieve comparable or perhaps even better results. As stated previously, in this thesis no claim is made regarding optimality, but rather that the concept of statistical classification of radar clutter is possible and practical. As with any new algorithm, only a trial implementation will allow a thorough assessment of its utility. In this research an attempt was made to emulate such an implementation.

The radar systems, and the data derived from them, are described first. Both radars were operational systems, and the restrictions this placed on the type of

data collected are also discussed. Section 6.1 then concludes with a summary of the preprocessing and feature extraction algorithms. Section 6.2 deals with the classifier training strategy, which is crucial to successful classification. The work in this thesis is based on the assumption that is it possible to apply supervised learning techniques, which require *a priori* knowledge about the nature of the clutter to allow sample labelling. With this knowledge, the statistics of the labelled samples can be examined and the feature space divided into classes using some very practical considerations. At this point classifier performance can be predicted. Experimental error rates are compared with these predictions in Section 6.3. Qualitative, but very illustrative results are presented as well in the form of classified PPI images, and a solution to the problem of ground clutter, which can lead to intolerable false alarm rates if not handled correctly, is also proposed. The final section then draws some conclusions from these results and cautions against their misinterpretation.

6.1 Data Preparation and Feature Extraction

Approximately 260 minutes (20 gigasamples) of radar data were collected from two L-band air traffic control radars. Some understanding of the radar is 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 can take place.

6.1.1 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 [Kesler, 1978].

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 Department of National Defence (DND), the Moose Jaw 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 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 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 has been reported to be as much as 4 dB greater than for vertical polarization [Vaughn, 1985]. (Information on polarization modulation of target echoes may also be found in [Skolnik, 1970], and some data about the effects of polarization on weather clutter has been compiled by Nathanson [1969] and Hendry and McCormick [1971]. Macikunas and Haykin [1989] have examined the use of polarization as a radar discriminant for several applications.)

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 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 PRFs 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 Appendices A and B.

In Chapter 3 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 PRFs, 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

$$I = S\cos\phi\,, \quad Q = S\sin\phi \tag{6.1}$$

The correction which must be applied to the measured values I' and Q' are thus

$$I = k_1 I' \quad \text{and} \quad Q = k_2 (Q' + k_3 I) \tag{6.2}$$

where

$$k_1 = \frac{1}{k_e}; \quad k_2 = \frac{1}{\cos\delta}; \quad k_3 = \sin\delta \tag{6.3}$$

The quantities k_e 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 Appendix E.

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 data, which has the lowest variance of any data available from the radar and is white relative to the PRF, can be found in abundance in the region above 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 (see Appendix E) and the computation of the SNR based feature U_0.

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/inch, 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 with the assumption that their effects are negligible when compared to the large volumes of correct data.

6.1.2 Data Library

The complete library consists of 15 indexed tapes, each containing 15 to 18 minutes (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 includes 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 nautical miles north-west of Moose Jaw (at a true bearing of 326.5°), 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 unusual clutter was observed. Later analysis showed that most of the suspected clutter was indeed caused by birds (see Appendix C), 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 travelling 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 Saskatchewan can increase the maximum range of the extent of ground clutter to over 80 nautical miles. 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 [Stehwien, January 1986].

Subsets of these tapes were transferred, after data selection and

preprocessing, to 9-track magnetic tape. Selection was made by range/azimuth gating while retrieving data from the tapes, and after examining whole scans for areas of interest using simulated PPI displays. These subsets form the basis of the experimental data used to train and test the classifier.

6.1.3 Feature Extraction Algorithm

The basis for the feature vector are, as discussed in Chapter 5, the complex reflection coefficients arising from the lattice implementation of the prediction error filter (PEF). They are computed using Burg's formula [Burg, 1968; Haykin and Kesler, 1979], which is restated here for convenience

$$\rho_m = \frac{-2 \sum_{n=m+1}^{N} f_n^{(m-1)} b_{n-1}^{*(m-1)}}{\sum_{n=m+1}^{N} [\,|f_n^{(m-1)}|^2 + |b_{n-1}^{(m-1)}|^2]} \tag{6.4}$$

where $f_n^{(m)}$ and $b_n^{(m)}$ are, respectively, the forward and backward prediction errors from the m^{th} stage of the lattice PEF

$$f_n^{(m)} = f_n^{(m-1)} + \rho_m b_{n-1}^{(m-1)} \tag{6.5}$$

$$b_n^{(m)} = b_{n-1}^{(m-1)} + \rho_m^* f_n^{(m-1)} \tag{6.6}$$

These lattice equations are initialized with the data values themselves

$$f_n^{(0)} = b_n^{(0)} = x_n \tag{6.7}$$

Recall that the data values x_n must first be error corrected and quadrature compensated as discussed in Section 6.1.1. The feature extractor need go no further than Equation (6.4), but should one wish to compute the spectrum, the PEF coefficients themselves must be computed. An efficient, although not necessarily the most numerically stable procedure to compute these is Levinson's recursion [Haykin and Kesler, 1979]

$$h_i^{(m)} = h_i^{(m-1)} + \rho_m h_{m-i}^{*(m-1)} \qquad i = 1, 2, \ldots, m \tag{6.8}$$

Which is initialized with

$$h_0^{(m)} = 1 \qquad \text{and} \qquad h_m^{(m)} = \rho_m \tag{6.9}$$

Finally, the m^{th}-order maximum entropy power spectral density function is given by

$$\hat{S}(f) = \frac{P_m}{B \left| \sum_{i=0}^{m} h_i^{(m)} exp(-j2\pi i f \Delta t) \right|^2} \qquad \text{for } |f| \leq \frac{1}{2\Delta t} \tag{6.10}$$

where B is the bandwidth of the complex process and the sampling interval Δt is set to its Nyquist rate $1/B$. For the radar case Δt is fixed at the pulse repetition interval (PRI), and the bandwidth B is therefore limited by the PRF ($B = f_s$). Any frequency present outside B will be folded into the spectrum. The reference power level P_m is the residual error power from the m^{th}-order PEF, computed from

$$P_m = P_{m-1}(1-|\rho_m|^2) = P_0 \prod_{i=1}^{m}(1-|\rho_i|^2) \tag{6.11}$$

which is initialized from

$$P_0 = R(0) = \frac{1}{N}\sum_{n=1}^{N} x_n x_n^* \tag{6.12}$$

The units of P_0 are then those of the data, squared. It is usually more convenient to express P_0 as a power ratio relative to the background noise level W, which may be estimated from a sector containing only thermal noise. Thus, by implication, the relative error power P_m/W becomes a measure of the goodness of autoregressive fit (see also Section 5.2), and should always remain greater than 1. Since the error powers computed per (6.11) can only become smaller as the number of stages increases, a value below 1 should raise suspicion regarding the validity of the model, and be cause to immediately terminate the growth of the lattice.

The importance of a low variance in the feature samples has been pointed out in Chapter 4. In Chapter 5 the variance of the zero mean reflection coefficients was shown to be proportional to the quantity $1/(N-m+1)$, where N is the number of independent samples used for the estimate. A rule of thumb for the largest filter order to be used can be constructed from the Bhattacharyya bound (Equation 4.32). The covariance matrix for a classifier using only the highest-order reflection coefficient - the real and imaginary parts are each considered to be one independent feature with one half of the total variance - is diagonal with elements $\sigma^2/2$. Equation (4.34) thus becomes

$$B_{ij} = |\mu_i-\mu_j|^2/4\sigma^2 \tag{6.13}$$

In Appendix D the upper limit on the magnitude of the m^{th} reflection coefficient is derived to be $1/m$, if the underlying signal is a complex sinusoid in white noise. If the variance σ^2 is to be taken as $(1-1/m^2)/(N-m+1)$, which combines the result of Equation (5.37) with the dependence on the magnitude shown in Equation (5.28), and the mean distance between two classes as $1/m$, then Equation (6.13) can be evaluated to be

$$B_{ij} = \frac{N-m+1}{4(m^2-1)} \tag{6.14}$$

A convenient upper limit on the error rate is 18%, which corresponds to a Bhattacharyya distance of 1. Placing this as a lower limit on B_{ij} then puts the following upper limit on m

$$m \leq \frac{1 + \sqrt{1+16(N+5)}}{8} \approx \frac{1}{2}\sqrt{N+5} \tag{6.15}$$

While this rule of thumb *must* be used with caution because of the assumptions and simplifications used in its derivation, it does illustrate the problem of a low number of samples rather nicely. If 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 filter may be used if a low error rate is to be maintained. Similarly, 63 samples are needed for a fourth-order filter, and 100 for a fifth-order filter. For even longer filters the number of samples rises rather dramatically. Removing the restriction for the maximum magnitude of ρ_m, which would be the case when high-order information is present in the spectrum, obviously improves the situation but the trend remains. It is therefore imperative to include as many samples as possible in the estimation procedure.

The danger in increasing N arbitrarily, however, 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. 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.* [1982] propose the use of the multisegment Burg formula - first described by Moorcroft [1978], 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

$$\rho_m = \frac{-2 \sum_{k=1}^{K} \sum_{n=m+1}^{N} f_{n,k}^{(m-1)} b_{n-1,k}^{*(m-1)}}{\sum_{k=1}^{K} \sum_{n=m+1}^{N} [\,|f_{n,k}^{(m-1)}|^2 + |b_{n-1,k}^{(m-1)}|^2]} \tag{6.16}$$

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

$$f_{n,k}^{(m)} = f_{n,k}^{(m-1)} + \rho_m \, b_{n-1,k}^{(m-1)} \tag{6.17}$$

$$b_{n,k}^{(m)} = b_{n-1,k}^{(m-1)} + \rho_m^* f_{n,k}^{(m-1)} \tag{6.18}$$

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

$$f_{n,k}^{(0)} = b_{n,k}^{(0)} = x_{n,k} \tag{6.19}$$

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 determine the form of the weighting and the number of samples accessed outside the resolution cell. 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 (6.16) if it is clear that they belong to a different cell. This test is more thoroughly explained in Section 6.2.

Similar data weighting is applied to the computation of P_0

$$P_0 = \frac{1}{NK} \sum_{k=1}^{K} \sum_{n=1}^{N} x_{n,k} x_{n,k}^* \tag{6.20}$$

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 was used throughout the research and 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 which should arise from use of only those samples inside the resolution cell and 87, which would result from weighting all 3×29 samples equally. A fourth-order filter may thus be used with a clear conscience. Slightly higher order filters may still be used; however, care must be taken to ensure a real increase in class separability can still be obtained.

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

$$\rho_m' = \rho_m \exp(-jm\phi) \tag{6.21}$$

where

$$\phi = arg(-\rho_1) \tag{6.22}$$

Rotating ρ_1 by its own argument simply leaves the magnitude $\rho_1' = -|\rho_1|$. The relationship of the magnitude r with the SNR n_0 (as shown in Appendix D),

$$r = |\rho_1| = \frac{n_0}{1 + n_0} \tag{6.23}$$

leads to the normality transformation

$$u_0 = 10\,log(n_0) = 10\,log\left[\frac{r}{1 - r}\right]. \tag{6.24}$$

The need for such a transformation has been discussed in Chapter 5, and its precise form arises from the knowledge that the probability densities of most clutter amplitudes of interest are approximately log-normal [Barton, 1975; Long, 1975]. Of course, the same applies to the signal strength feature P_0, conveniently expressed in decibels relative to the thermal noise floor W

$$U_0 = 10\,log(P_0/W) \tag{6.25}$$

A similar transformation may be appropriate for the second- or third-order coefficients as well; however, these remain complex valued and have less of a clear relationship with a normally distributed variate. For now, at least, they will be used directly.

Two additional features remain, that is 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 Chapter 5 (Equations 5.24 and 5.26). For the multisegment architecture of the feature extractor they are computed as

$$P_{var} = 1 - \frac{\bar{s}^2}{P_I} \quad \text{and} \quad P_{dif} = \frac{P_I}{P_0} - 1 \tag{6.26}$$

where

$$\bar{s} = \frac{1}{K}\sum_{k=1}^{K} s_k \quad \text{and} \quad P_I = \frac{1}{K}\sum_{k=1}^{K} s_k^2 \tag{6.27}$$

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

$$s_k = |x_{(N+1)/2,k}| \tag{6.28}$$

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.

This completes the feature set. It now remains to define a set of classes, to train the classifier, and to evaluate its performance.

6.2 Training Strategy

Proper training is imperative for acceptable classifier performance. In fact a poor choice of classes, or inadequate knowledge about the underlying structure of the class statistics will invariably lead to incorrect or even misleading results. The strategy

thus employed is one of supervised learning, and to refine the class structure based on the knowledge gained during this process. This knowledge includes the identity of the clutter source, some expected characteristics of the Doppler spectra, and the statistics of the feature set. To some degree, consideration must be given to the consequences of misclassification, and the confidence with which the decisions have been made.

6.2.1 A Priori Knowledge

Supervised learning requires that the identity of the clutter data be known. It also includes inspection, selection, and labelling of the desired sectors before processing. In most cases, performing these tasks was not difficult. Specific weather clutter types were easy to identify and simply required inspection to ensure that no unrelated clutter or targets were present. Much the same was true for aircraft, although isolating them was quite tedious due to their high speed (scan-to-scan displacement) and low density. Ground clutter was available in abundance but was also more susceptible to contamination by all kinds of unrelated clutter. It was thus important to select recordings made during environmental conditions which were not conducive to either precipitation or bird activity. It had to be restricted to areas with little air traffic, and where vehicular traffic was not generally visible. On the other hand, radar echoes caused by ground based vehicles are, by definition, contaminated by ground clutter and present a special problem. Most interesting was the identification of bird clutter which, once the data were analyzed, turned out to be plentiful. Their low cross-sections also presented a training challenge.

To maximize the confidence of the identifications, the radar data had to be faithfully reconstructed and inspected. It was simply not acceptable to select data at random from the massive data volumes on the tapes. To this end a completely

interactive data management and display program was written. Notes made during the recording sessions guided the choice of tape to use and the approximate location on the tape as identified by the scan number. Several full scans, or a sequence of smaller windows from many consecutive scans could be retrieved and viewed using a simulated PPI display. Quick access to Doppler information was also available by using a bipolar, two-color display for either of the in-phase or quadrature channels. Bit error correction occurred at this stage, as did the computation of quadrature error compensation factors. Homogeneous clutter areas were then selected and separated into labelled data sets.

Ground truth of the data was also required, especially for aircraft, vehicles, and bird clutter. True north alignment and exact zero mile offsets were found using precisely surveyed permanent echoes. Digitized video maps were overlaid to assist in the identification of ground clutter features such as powerlines and buildings, and help prevent tower echoes from being misidentified as aircraft. Maps of airways, airspace boundaries and airports served to find en-route aircraft, while maps of highways located and provided the necessary proof for car and truck echoes. Somewhat less successful was the use of maps of lakes to find concentrations of water fowl activity. They did, however, provide further identification of ground features such as sloped terrain, valleys, and shorelines. The north alignment allowed determination of the direction of travel of the clutter and thus was a significant contributor to the evidence for bird clutter. Noncoherently integrated maps were useful for the location of bird clutter and aircraft. Tracks left by the clutter allowed precise positioning of the data windows and minimized the introduction of unrelated clutter.

Much of the a priori knowledge regarding the identity of the clutter came from simply being present when the recordings were made. The ability to cleanly

separate clutter sectors can thus take full advantage of this knowledge in the labelling process. There are cases, however, where labelling is possible but clean separation is not. It may also be the case that some targets had been missed inside the sectors of interest, and these would contaminate the class prototype if not removed. Knowledge about the expected Doppler spectrum can help reject many of the contaminating feature sets. For most clutter types only one main peak is expected at the mean Doppler frequency. Two peaks will really only occur when two clutter sources moving at different speeds are present within the same resolution cell. Such cases are relatively rare and hence no classes were defined for them. They must therefore be rejected except, perhaps, for the case of two peaks being legitimately caused by windshear in areas of rain. Fortunately, as discussed in Chapter 5 and shown in Appendix D, the magnitude of ρ_1 moves away from unity as the second peak strengthens and thus can be used in a validity test for the time series of interest. The phase of ρ_1 is also affected by the second peak and a limit can be placed on its variation within the set of K time series to be combined. Such a test also prevents clutter edges and transitions from unduly influencing the clutter prototypes. Algorithmically,

$$\rho_m = \frac{-2 \sum_{k=1}^{K} \beta_k \sum_{n=m+1}^{N} f_{n,k}^{(m-1)} b^{*(m-1)}_{n-1,k}}{\sum_{k=1}^{K} \beta_k \sum_{n=m+1}^{N} [|f_{n,k}^{(m-1)}|^2 + |b_{n-1,k}^{(m-1)}|^2]} \tag{6.29}$$

and

$$P_0 N \sum_{k=1}^{K} \beta_k = \sum_{k=1}^{K} \beta_k \sum_{n=1}^{N} x_{n,k} x^*_{n,k} \tag{6.30}$$

The selection coefficient β_k is defined as

$$\beta_k = \begin{cases} 1, & \text{if } |\phi_c - \phi_k| \le \Delta\phi/2 \quad \text{and} \quad r_c - r_k \le \Delta r \\ 0, & \text{otherwise} \end{cases} \tag{6.31}$$

The phase and magnitude of the ρ_{1_k}, computed using Equation (6.4) and belonging to each of the k time series, are defined as

$$\phi_k = arg(-\rho_{1_k}), \qquad r_k = |\rho_{1_k}| \tag{6.32}$$

and ϕ_c is the phase of the reflection coefficient with the largest magnitude $r_c = max(r_k)$. The limits Δr and $\Delta\phi$ can be made as tight as required. To maintain a reasonable limit on the variances of the multisegment coefficients to be included in the class prototypes, the number of rejected time series must be kept low before the entire feature set is discarded. Thus

$$\sum_{k=1}^{K} \beta_k \ge K - N_{rej} \tag{6.33}$$

where $K - N_{rej}$ is the minimum allowable number of time series. In fact, the choice of N_{rej} was guided by another very practical consideration. During the computation of the feature sets used for training and testing, the window of K time series was moved by only one sample for each set. This was done to maximize the effective number of independent sets available for the prototype estimates. While a certain amount of dependence clearly remained and was acceptable, duplication was to be avoided. This could only be assured by restricting N_{rej} to be less than the smallest window dimension, that is 2 for the 3 × 15 size.

The test on ρ_1 reduces the probability of inclusion of unrelated clutter into the class prototype by outright rejection of data. There are cases, however, where it

may be more appropriate to remove such clutter from the data before applying feature extraction. To avoid distortion of the Doppler characteristics of the remaining clutter this must be done coherently. It also requires a priori knowledge about the amplitude and phase of the clutter to be removed. Such knowledge can usually only be obtained for completely stationary clutter which maintains its phase and amplitude exactly from scan to scan. A long term coherent clutter map can then be constructed. For every range r and azimuth φ of scan n, this may be done recursively

$$G_{r,\varphi}(n) = \lambda G_{r,\varphi}(n-1) + (1-\lambda)x_{r,\varphi}(n) \tag{6.34}$$

where λ is an exponential "forgetting" factor. The goodness of the mean estimate $G_{r,\varphi}(n)$ can be assessed by the amount of measurement noise remaining (the variance of the mean). Asymptotically, Equation (6.34) reduces this noise by a factor of $(1+\lambda)/(1-\lambda)$. The same reduction can be obtained using fewer scans with a uniformly weighted average

$$G_{r,\varphi}(n) = \frac{1}{N} \sum_{n=1}^{N} x_{r,\varphi}(n) \tag{6.35}$$

where the reduction factor is given by $1/N$. The mean estimates are then removed from the data

$$x'_{r,\varphi}(n) = x_{r,\varphi}(n) - G_{r,\varphi}(n) \tag{6.36}$$

Ideally, use of these maps will only remove ground clutter and leave all other clutter intact. Neither should they add any noise which could distort the clutter spectra. Therefore, due to a limitation in the number of available scans, Equation (6.35) was used almost exclusively and, with 64 scans integrated, no more than -18 dB of noise

would be added. A strong target present in only one scan would remain in the map 36 dB below its prime strength and would then be added into the data to be processed. This would really only cause problems with prime target strengths greater than 38 dB since no data with less than 2 dB SNR was processed. Such target strengths are unlikely for moving clutter, and for aircraft this level represents a peak amplitude of more than 43 dB, that is close to the saturation level of 46 to 49 dB.

In practice, these maps achieved only a 10 to 20 dB reduction in the average ground clutter levels. The reason is simply that this process is really a Doppler filter with a deep but very narrow null at a frequency of 0 Hz. Thus any clutter which moved only slightly over the period of N scans would pass through. At L-band, a 180° phase shift represents motion of about 5 to 6 cm which may be found in all but the most rigid objects. Scintillation is also induced by shifts in target position and aspect and widens the spectrum sufficiently to allow energy to pass through the filter, even if no phase shifts are present. Therefore any wind induced motion in towers, tall buildings, trees and other vegetation will prevent such clutter from being removed completely from the data. A tradeoff thus exists in the number of scans to use: the integration of more scans induces less distortion in (and due to) moving clutter but also achieves less of a reduction in ground clutter levels because of the narrowing Doppler null. Nevertheless, the maps were found to be useful for removing a significant portion of ground clutter during the processing of car and truck echoes. But bird and weather clutter, which were generally much weaker and thus more susceptible to interference from residual map clutter, were more successfully characterized when choosing only areas which were completely clear of ground clutter.

6.2.2 Feature Statistics

The statistics of the labelled sets of feature vectors are the basis of the class assignments and classifier performance predictions. Various tools are available and have been discussed at some length in Section 5.3. The classes examined in this section are aircraft, moving ground-based vehicles consisting mostly of cars and trucks, non-moving ground clutter, bird clutter, and weather clutter including rain, rain with an obvious windshear component, and thunderstorms. In some cases thermal noise data is included for comparison. Synthetic aircraft targets were used as a reference class. These targets were generated to emulate the radar characteristics as closely as possible, including the differential quadrature errors and noise characteristics. The underlying signal is a complex sinusoid with random frequency and starting phase, amplitude modulated in both range and azimuth with a Gaussian weighting function of the same width as seen in actual observed point targets. Peak amplitudes ranged from 0 dB to 46 dB, just below the saturation level of the 10 bit data word. The peak position inside the resolution cell was also randomized. The result is a set of idealized non-fluctuating point targets with no additional amplitude or frequency modulation. Such a target would be most representative of large jet aircraft in level flight, which have no part of their engine's moving parts exposed to the radar energy.

Scatter plots of 100 randomly selected feature sets of each of the various classes are shown in Figure 6.1 (6.1a to 6.1h). The frequency heterodyned complex reflection coefficients arising from a 9^{th}-order lattice PEF are plotted inside the unit circle to illustrate their bounded nature and to provide a sense of scale. The real part is taken as the abscissa. The SNR related features U_0 and u_0 are scaled such that 40 dB equals unity, and plotted with the P_0 derived feature U_0 as the abscissa and the ρ_1 derived feature u_0 as the ordinate. The shape parameters P_{dif} and P_{var} are also plotted

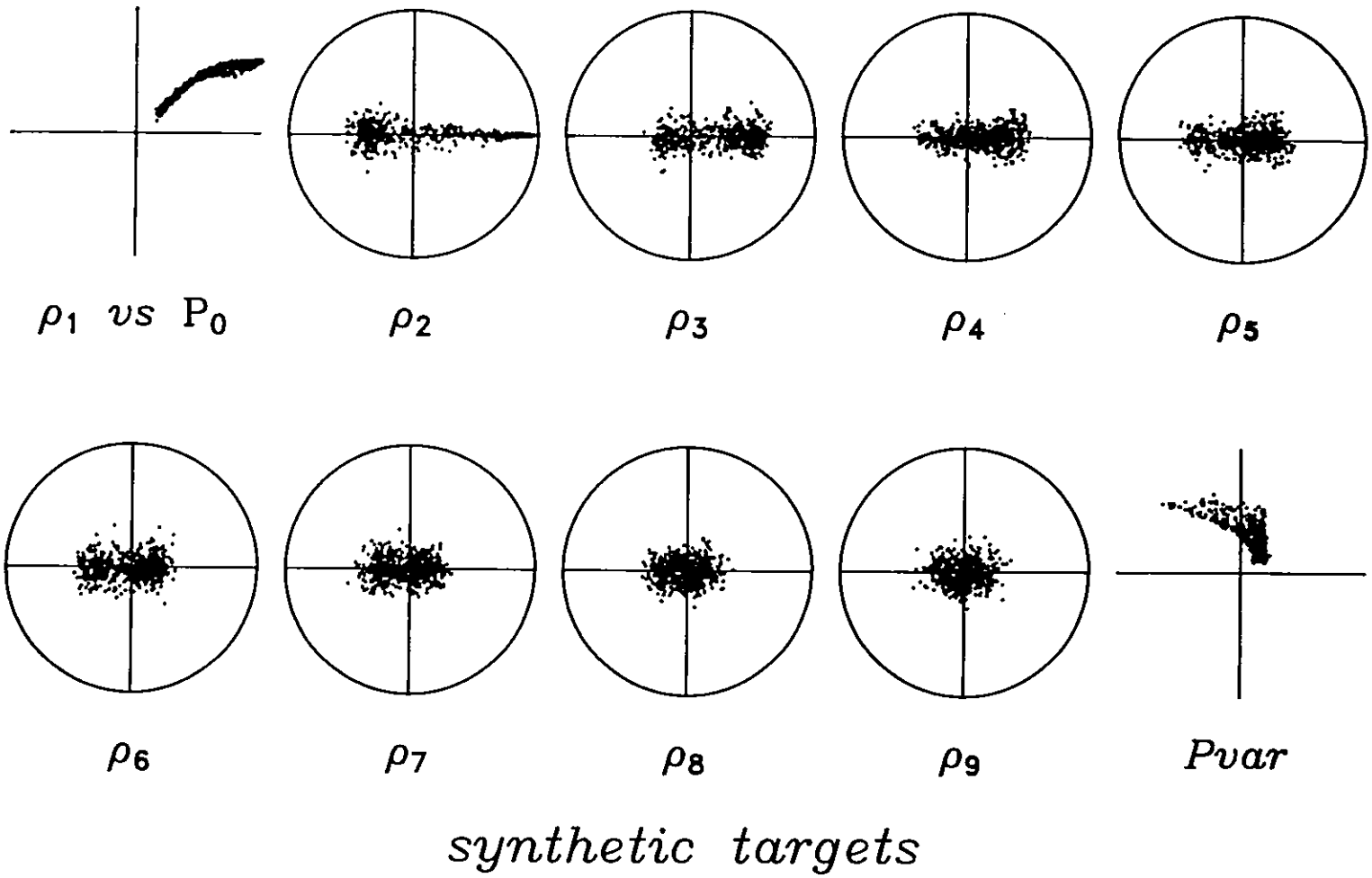


Figure 6.1a Scatter plot of the feature set arising from synthetically generated targets
see text for explanation

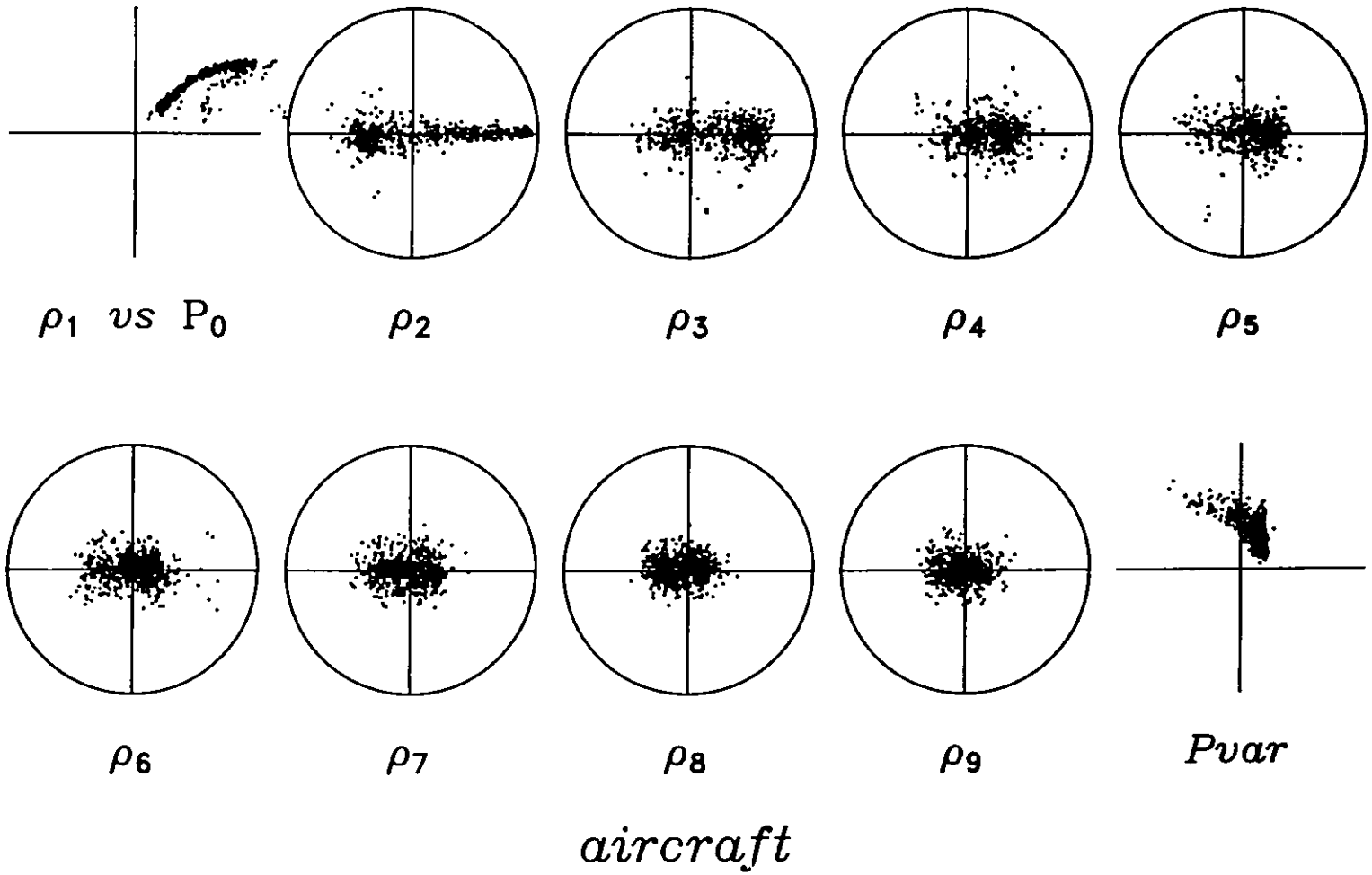


Figure 6.1b Scatter plot of the feature set arising from aircraft targets

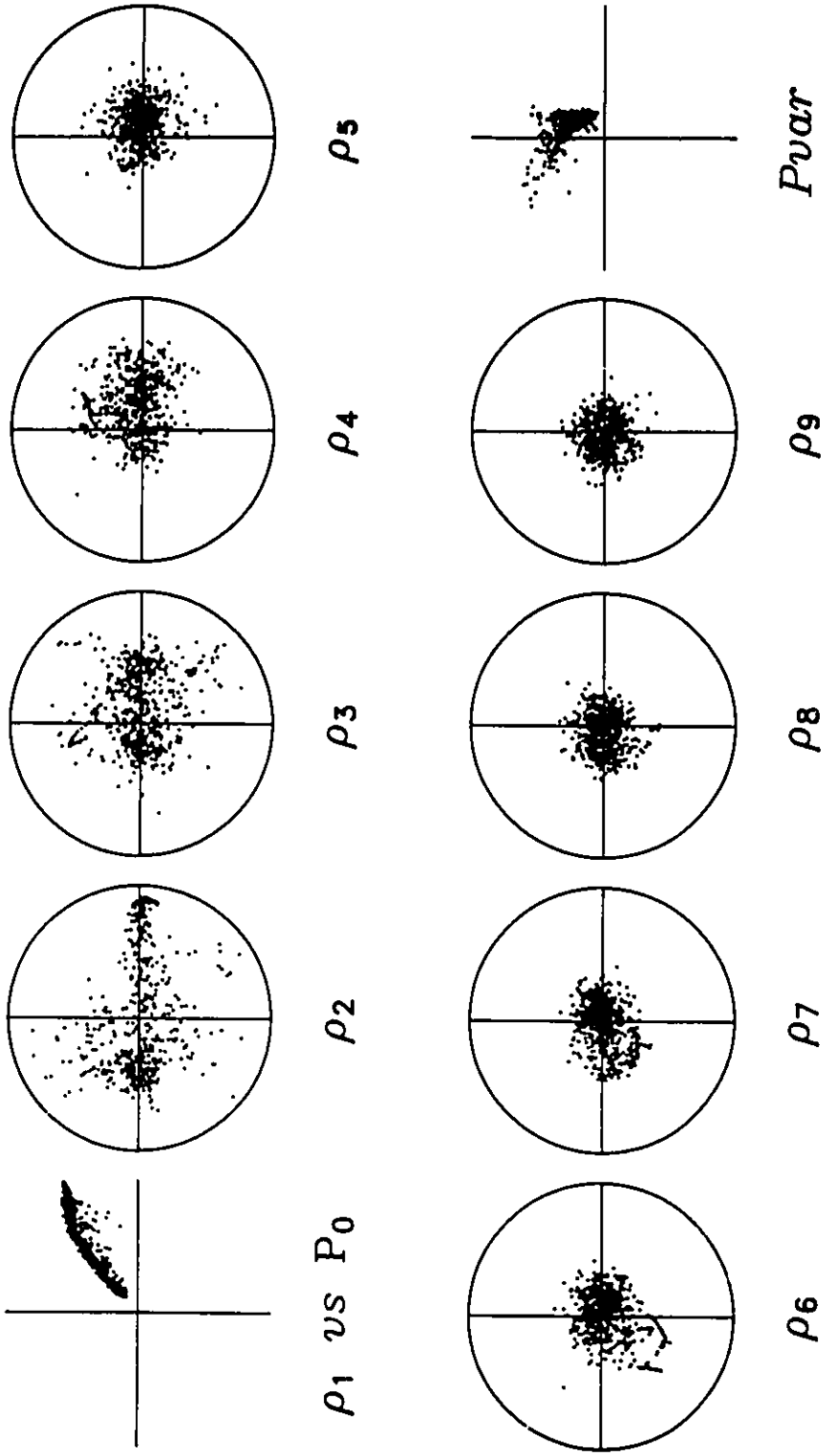


Figure 6.1c Scatter plot of the feature set arising from cars and trucks

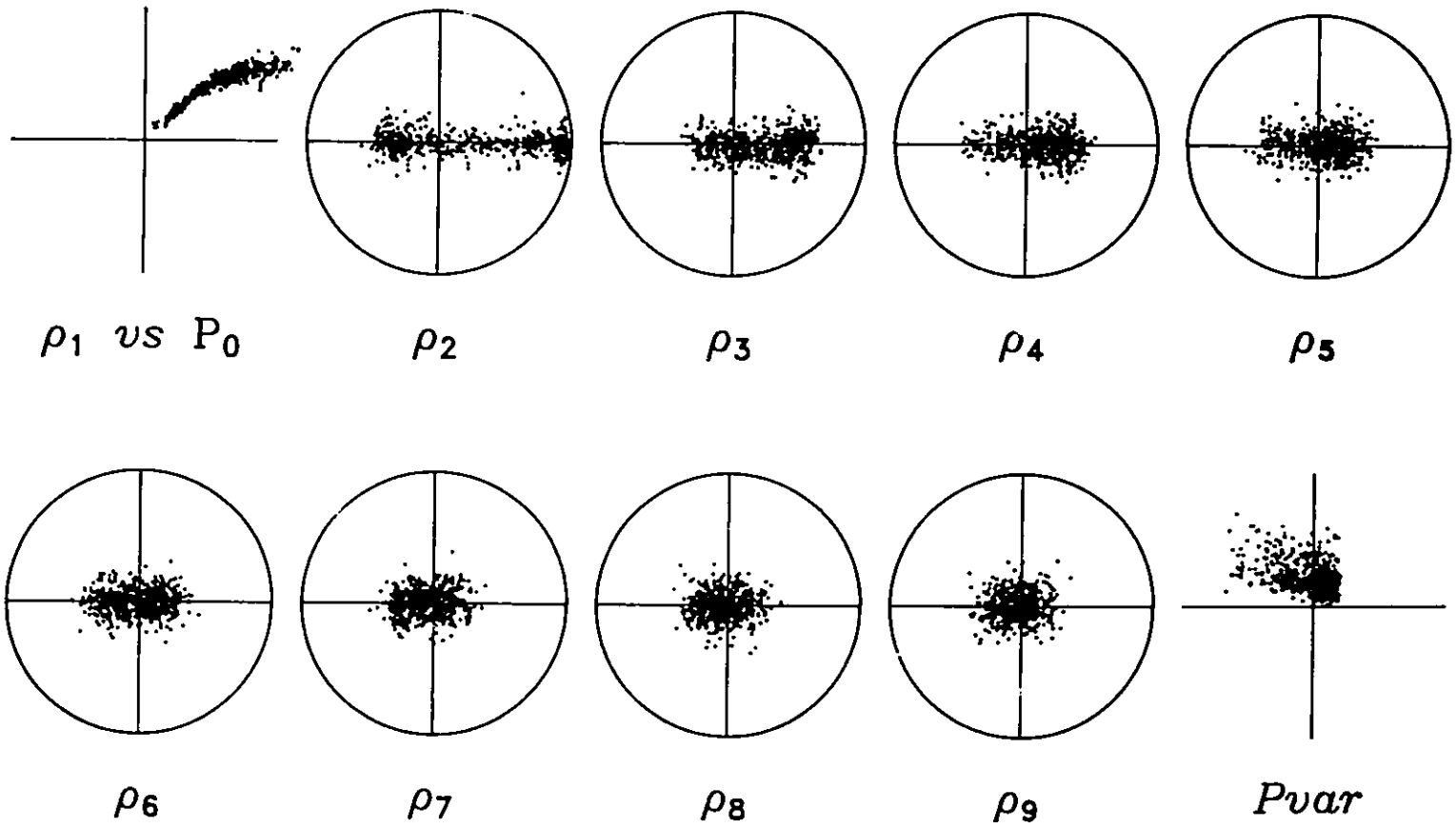


Figure 6.1d Scatter plot of the feature set arising from ground clutter

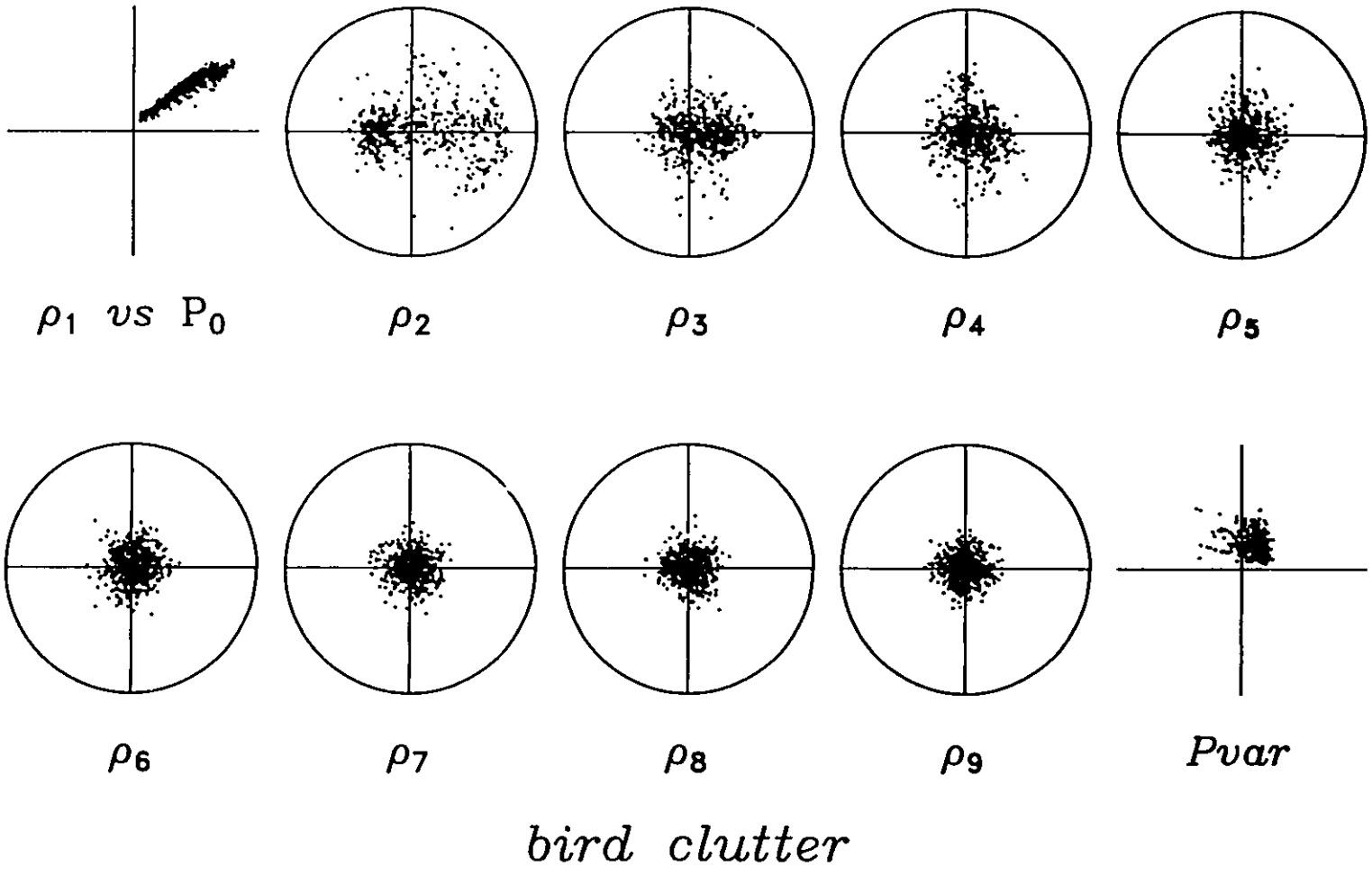


Figure 6.1e Scatter plot of the feature set arising from bird clutter

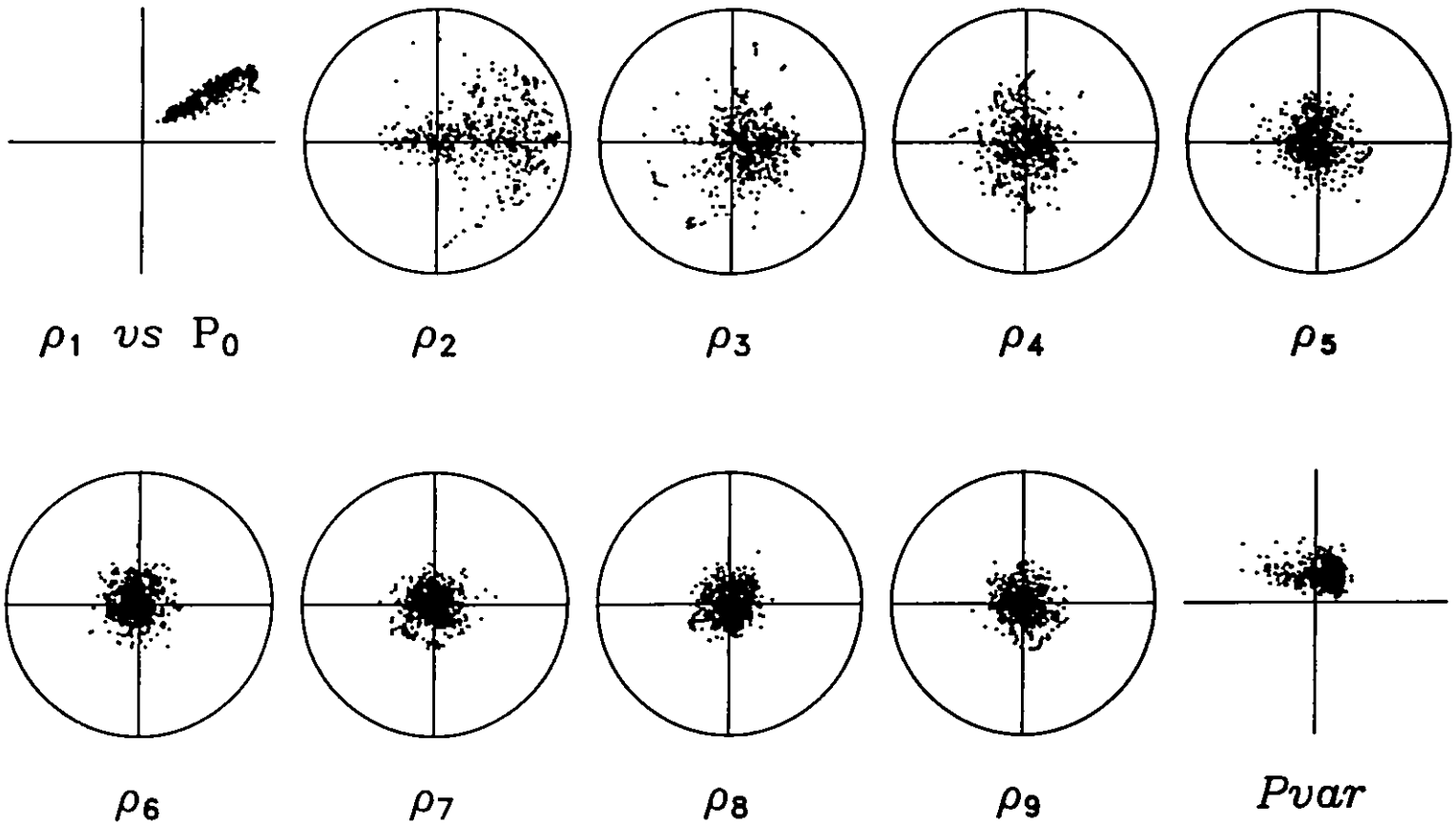


Figure 6.1f Scatter plot of the feature set arising from bird clutter
Note differences and similarities to the example in Figure 6.1e.

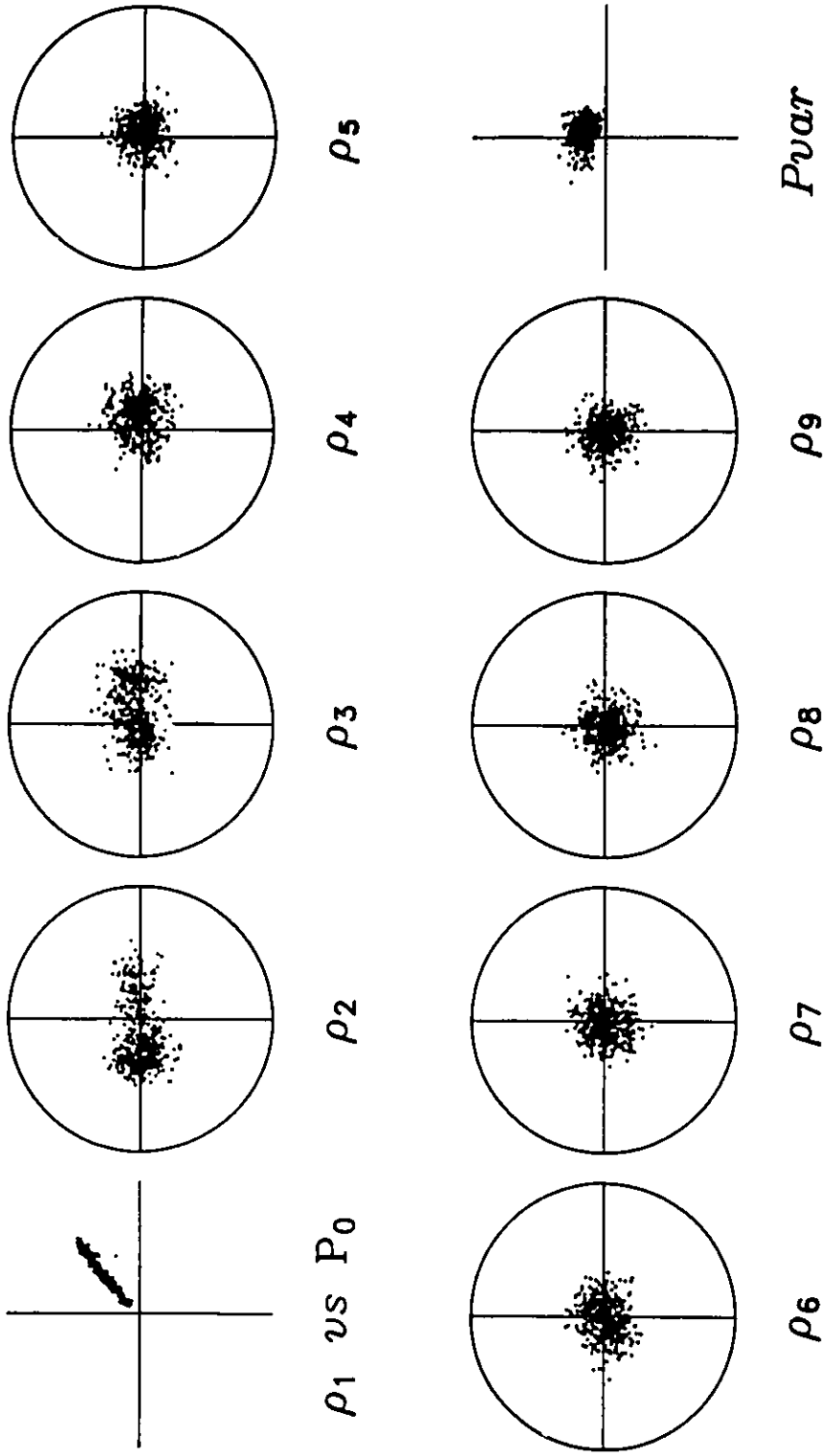


Figure 6.1g Scatter plot of the feature set arising from rain clutter

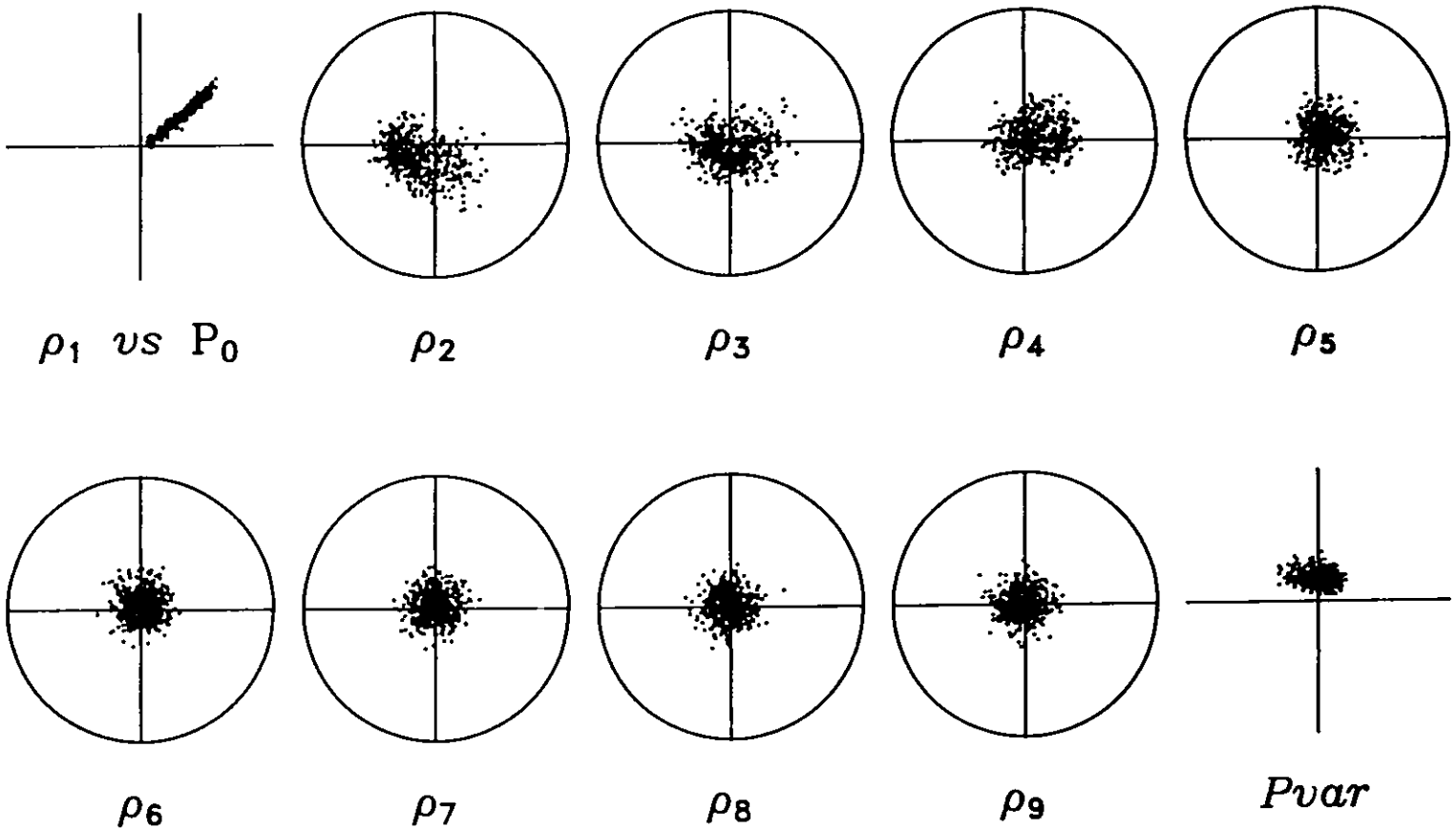


Figure 6.1h Scatter plot of the feature set arising from rain clutter with a windshear component

unscaled as the abscissa and ordinate respectively.

From these plots the similarity between the synthetic targets and aircraft is immediately apparent, thus validating the model used for the targets. These plots also show that most of the differentiating information appears to be contained in ρ_2' to ρ_4', with somewhat less in ρ_5' and P_{dif} and P_{var}. The last reflection coefficient ρ_9' clearly appears to contain little information. The fact that its variance is not noticeably greater than that of the lower order coefficients does show, however, that sufficient samples appear to have been used in its computation.

Another characteristic which can be clearly seen is the fact that most of the information in a reflection coefficient is contained in its real part. This is due to the symmetry of the spectrum. *Complete* symmetry will result in zero imaginary components, as is the case for real valued data. Symmetry arising from real valued signals in complex valued noise will generally result in nonzero imaginary components, but with a mean of zero. Any significant asymmetry in the spectrum will therefore manifest itself as a mean shift off the real axis. This may be found in clutter from cars, trucks, and birds, whose spectra can contain more than one peak. Windshear in rain systems can also contribute to a secondary peak and a nonzero mean for the imaginary component of at least ρ_2', as may be seen in Figure 6.1h.

To put these scatter plots into perspective, corresponding spectra are plotted in Figures 6.2 and 6.3. Seven 9^{th}-order spectra were randomly selected from those sets in each class with a signal strength of $U_0 = 20 \pm 0.5$ dB. For Figure 6.2, no frequency heterodyning was applied. The characteristics of each type of target and clutter and the distribution of their Doppler frequency (converted to knots) are apparent in Figure 6.2a. Secondary peaks can be seen in the aircraft spectra; these are likely caused by engines or propellers. The secondary peaks at zero Doppler in the car and truck spectra are due

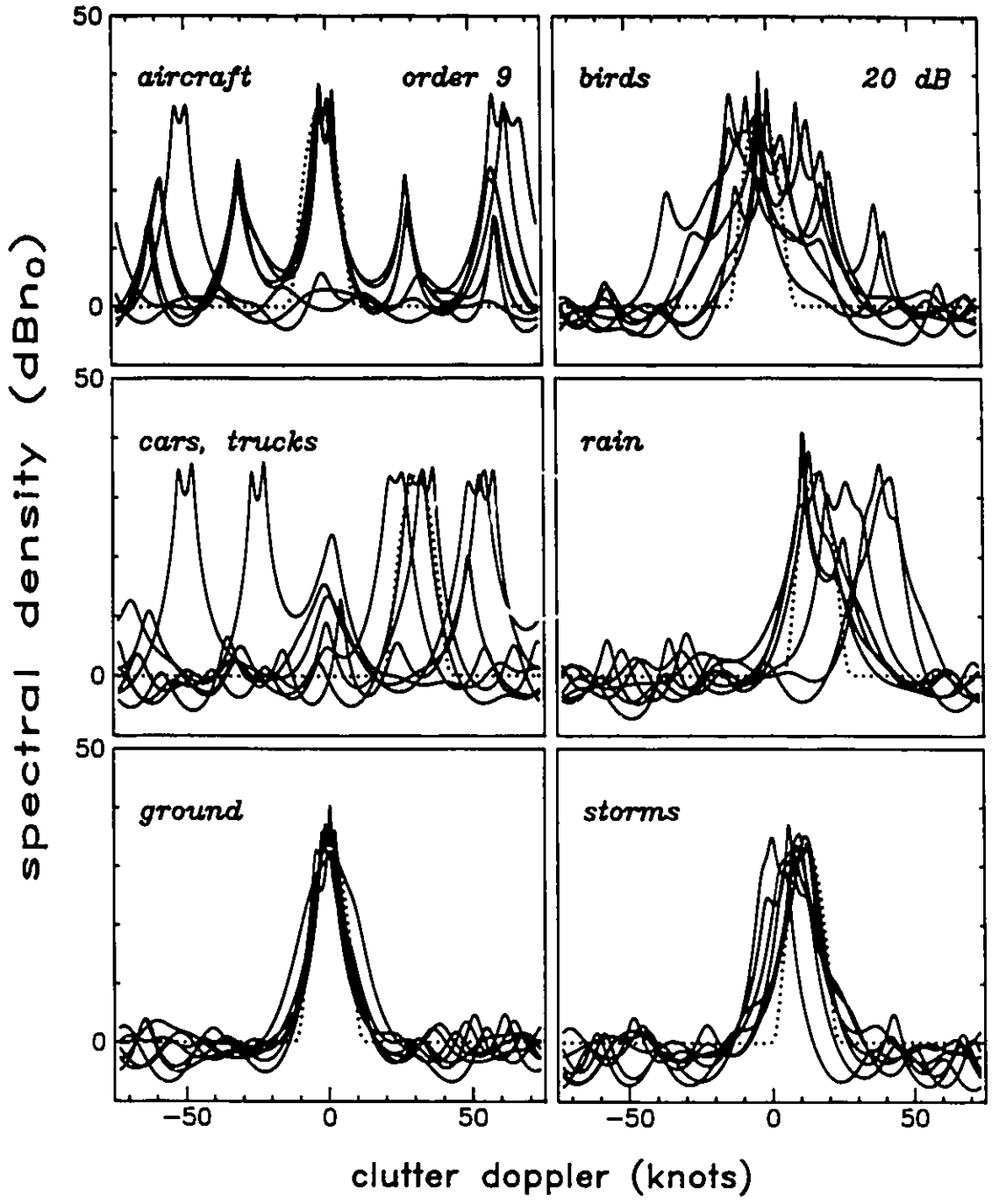


Figure 6.2a Sample 9th-order MEM clutter spectra

No frequency heterodyning was applied to these spectra.

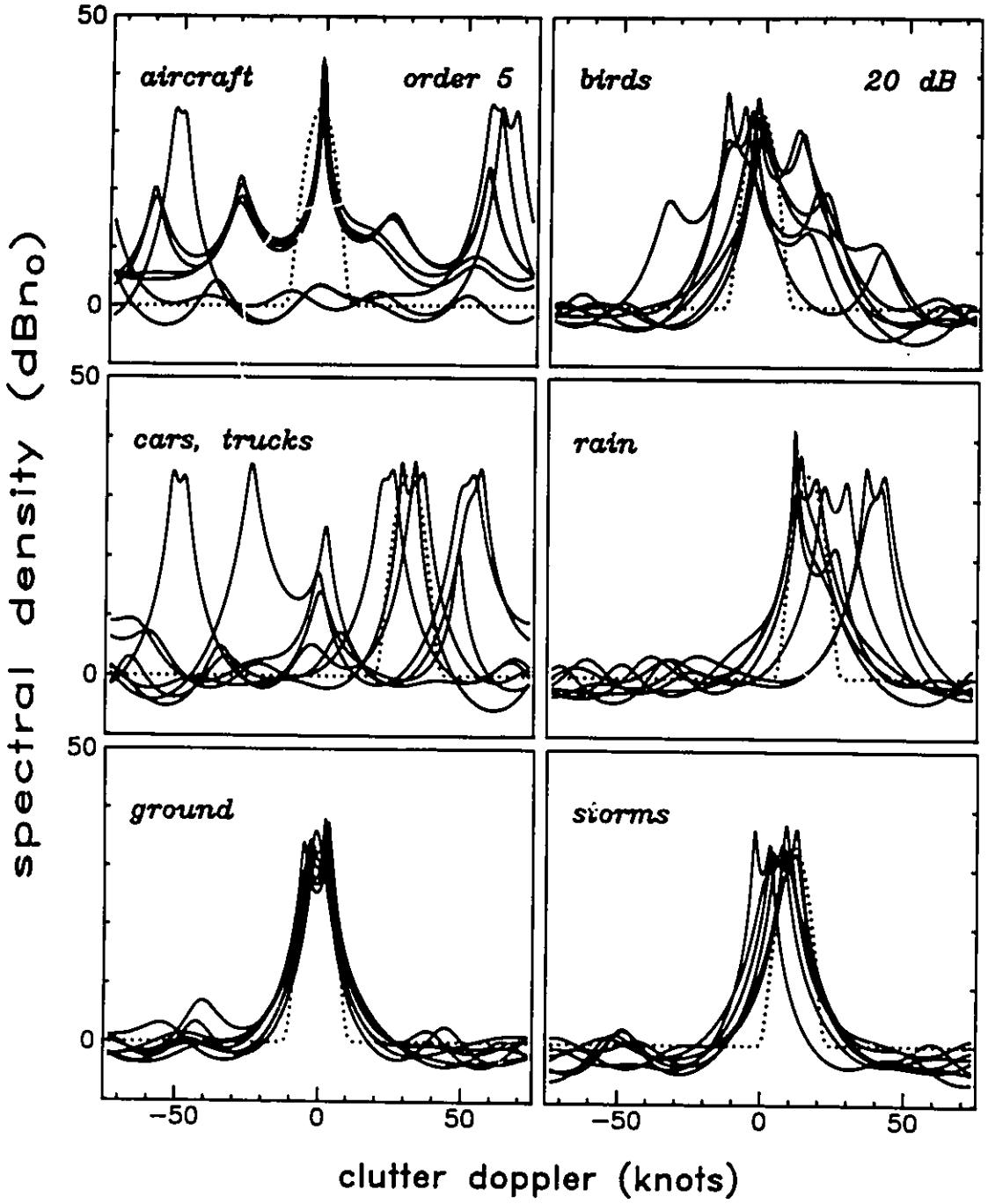


Figure 6.2b Sample 5th-order MEM clutter spectra

5th-order spectra of the same data as used in Figure 6.2a.

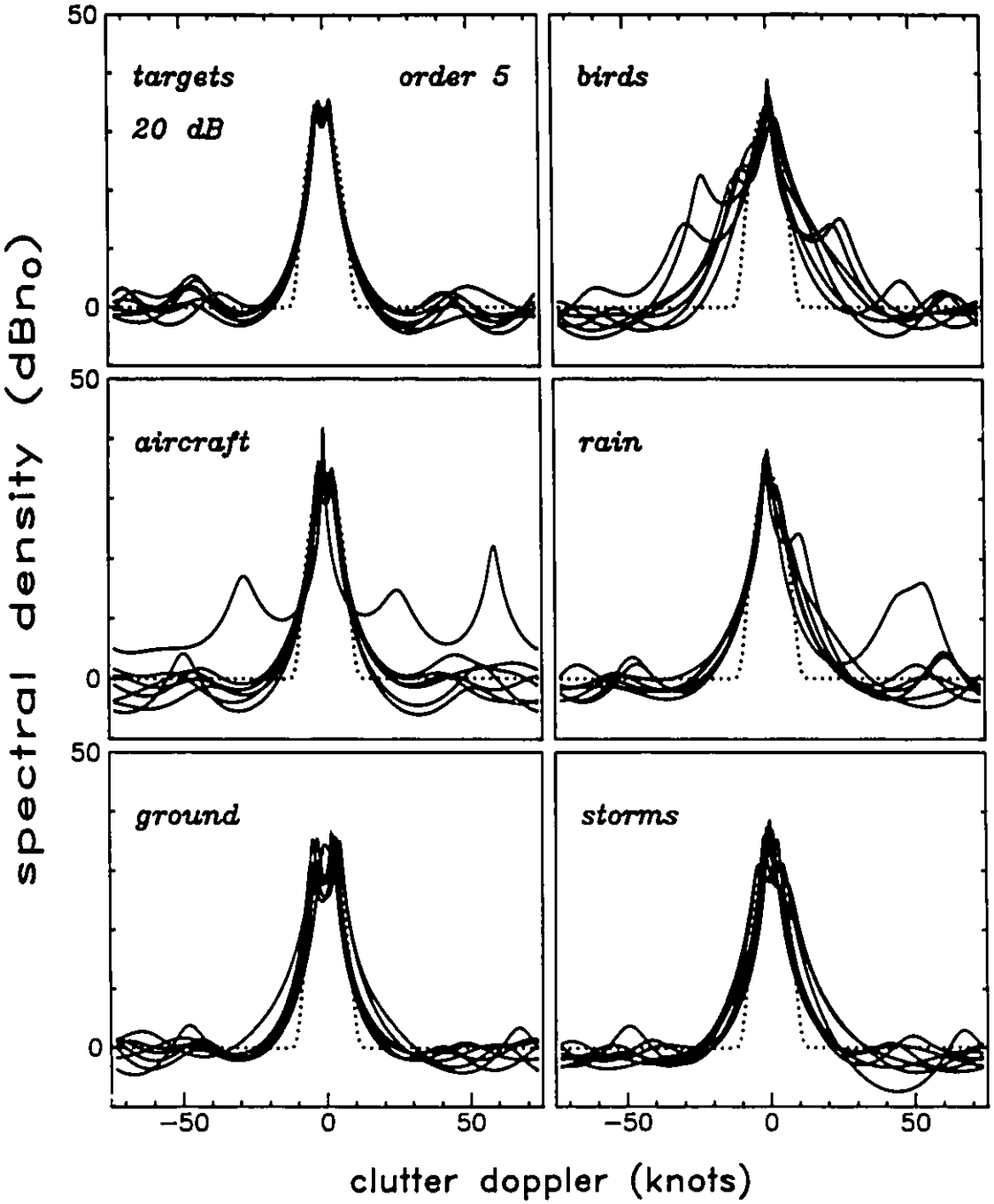


Figure 6.3a Comparative 5th-order MEM clutter spectra

Comparison between clutter types requires frequency heterodyning.

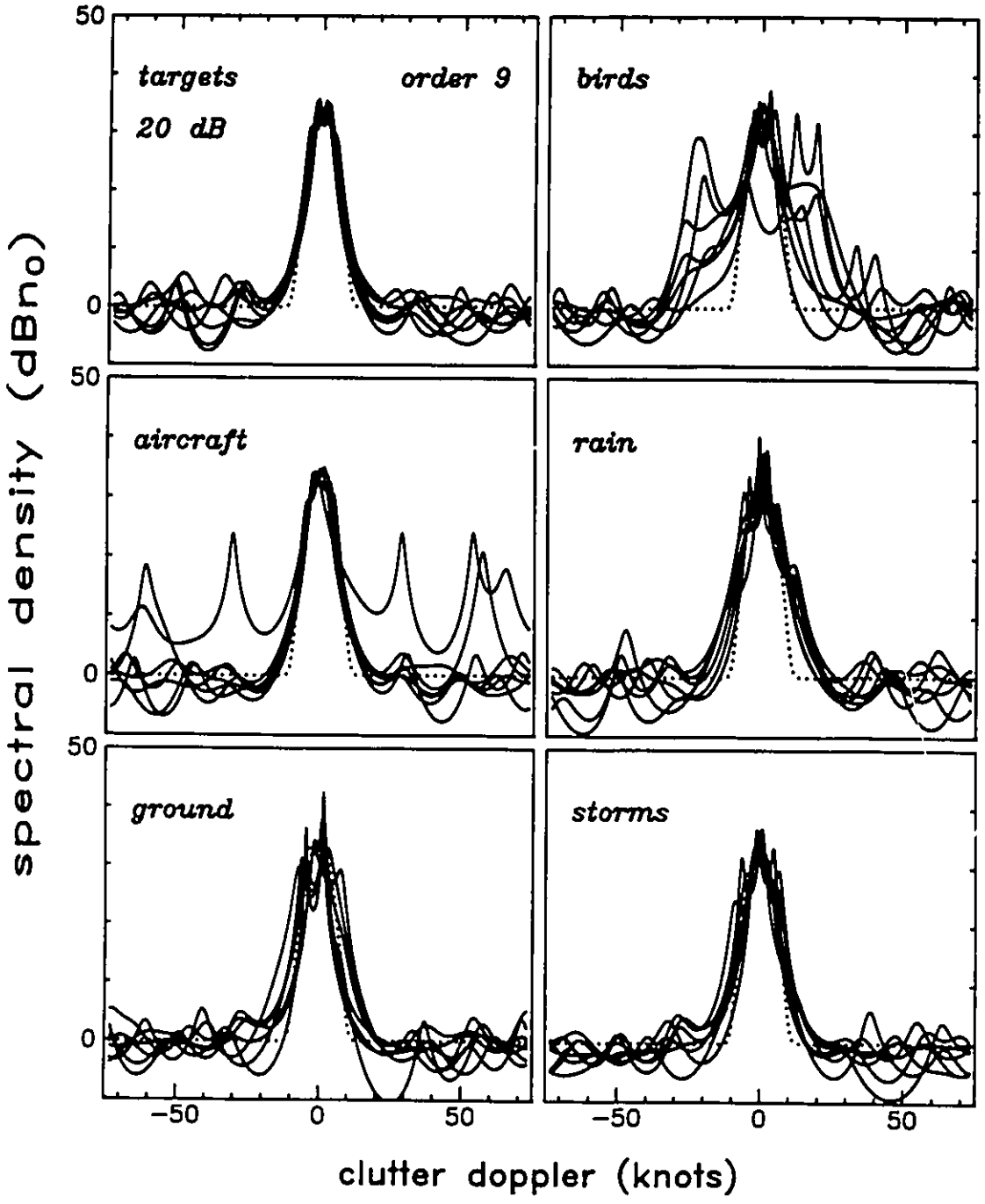


Figure 6.3b Comparative 9th-order MEM clutter spectra

The data used were the same as in Figure 6.3a.

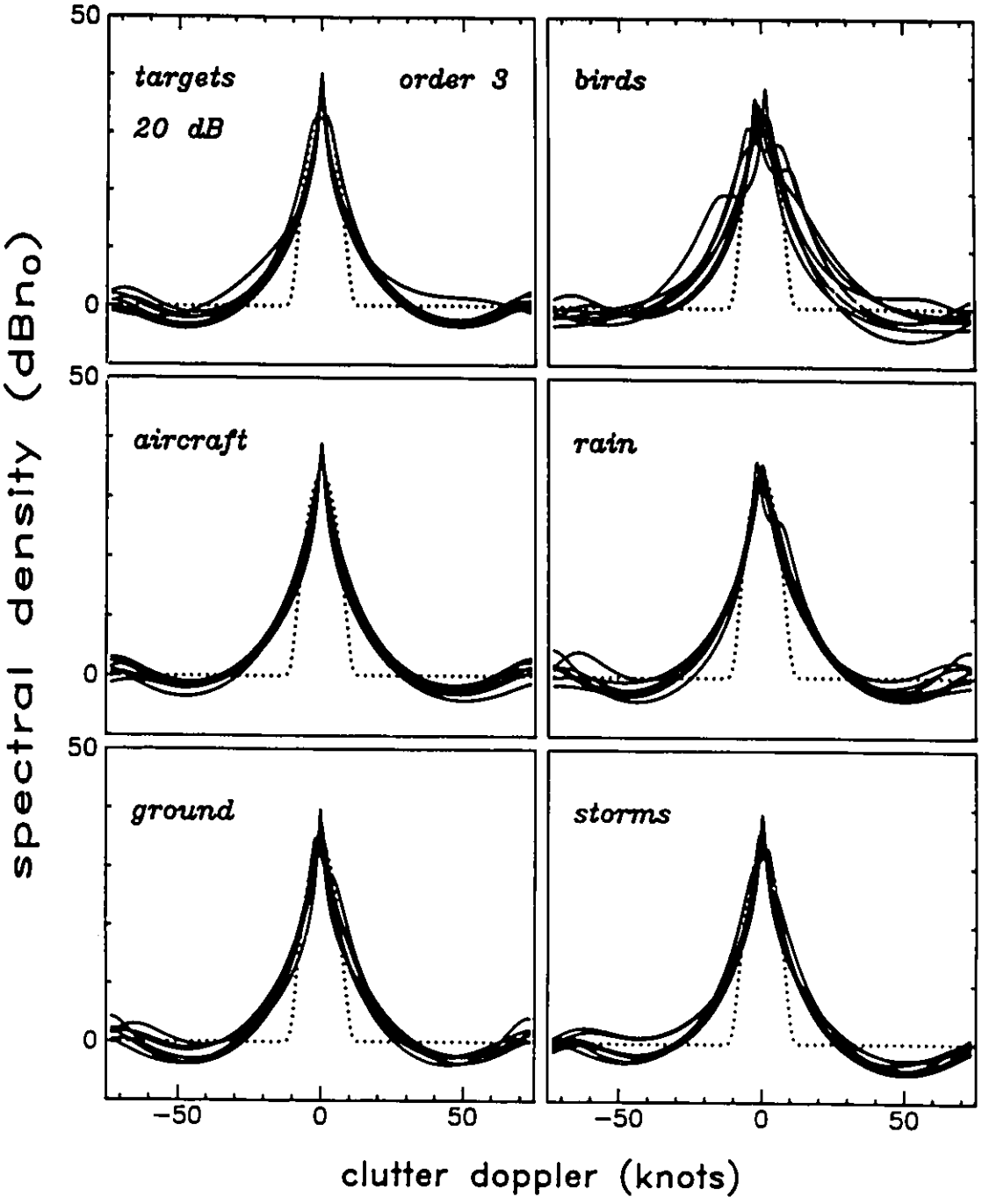


Figure 6.3c Comparative 3rd-order MEM clutter spectra

The data used were the same as in Figure 6.3a.

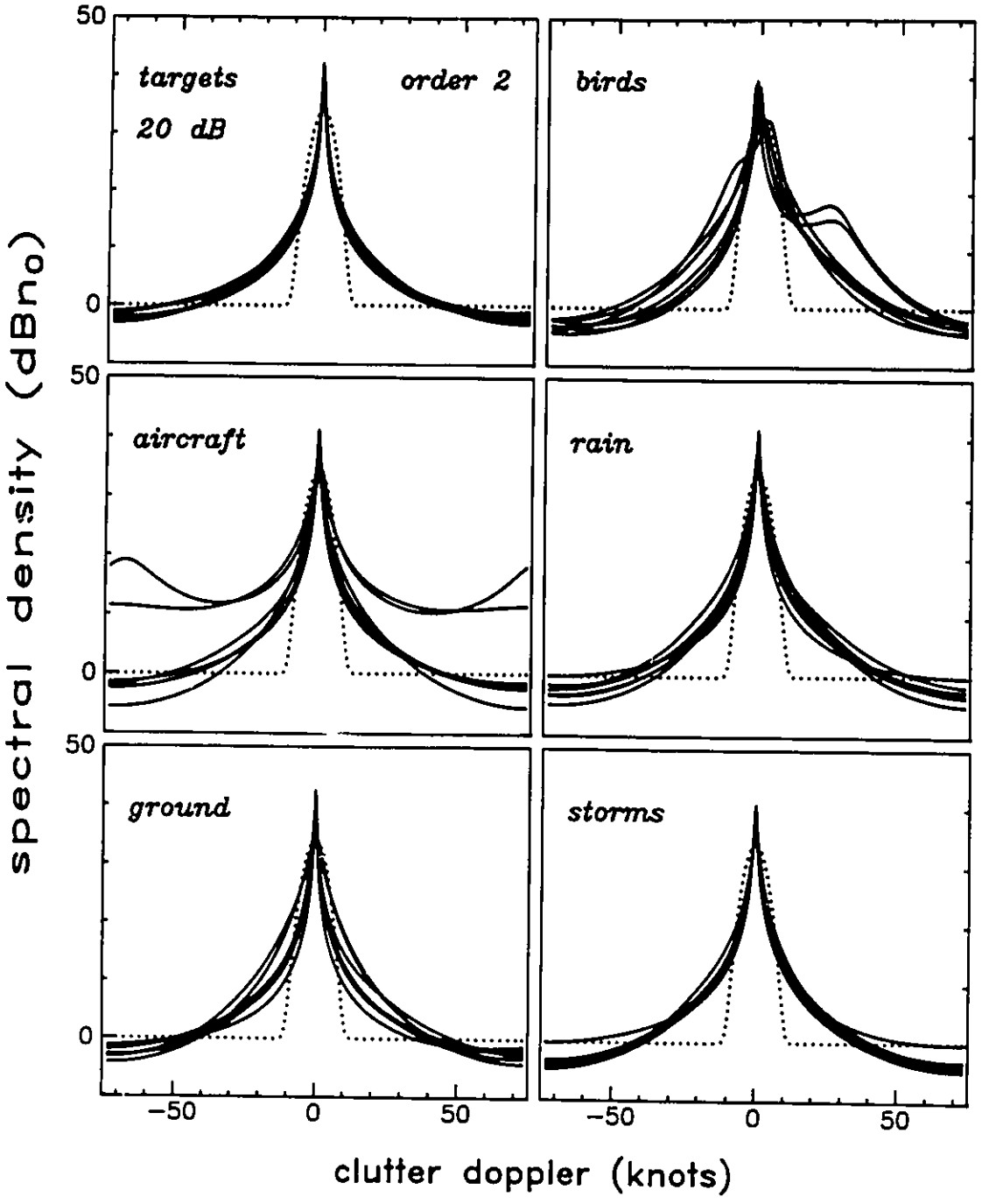


Figure 6.3d Comparative 2nd-order MEM clutter spectra

The data used were the same as in Figure 6.3a.

to residual ground clutter. Conversely, moving clutter is absent in the ground clutter class. The presence of multiple moving scatterers and the extra width of bird clutter compared to rain is evident. The storm class represents thunderstorm cells at a distance of 30 to 70 nautical miles from the radar, with each cell having a diameter from 1 to 5 miles. A plot of rain clutter spectra containing asymmetries due to windshear is not included in this comparison but may be found in Chapter 3 (Figure 3.6). It was observed previously that apparently little information remains in the reflection coefficients above order 5, and Figure 6.2b shows the 5^{th}-order spectra of the same data. The loss of high-order detail is quite noticeable, although the general characteristics are still evident. From this observation alone it is not clear whether a 5^{th}-order feature extractor will be sufficient.

A better comparison of the various classes is only possible with frequency heterodyned spectra. Figure 6.3a shows another set of 5^{th}-order spectra, but after heterodyning. The similarities amongst the classes are now the most obvious characteristic, any remaining difference is rather subtle. This implies that higher-order information would be required to bring out these differences, and this is indeed the case as shown in Figure 6.3b. Even in these high-order spectra there is no noticeable difference between rain and thunderstorms, although that should not be too surprising. Figures 6.3c and 6.3d show the detrimental effects of reducing the order below 5, which make the spectra practically indistinguishable.

Frequency polygons representing the marginal probability densities are shown in Figure 6.4. All features within the labelled sets were used in their computation. Deviations from normal behavior are clearly obvious in many cases. The statistics of receiver noise, which consists mostly of thermal noise, serve as a reference and provide experimental verification that the basic distribution of the reflection

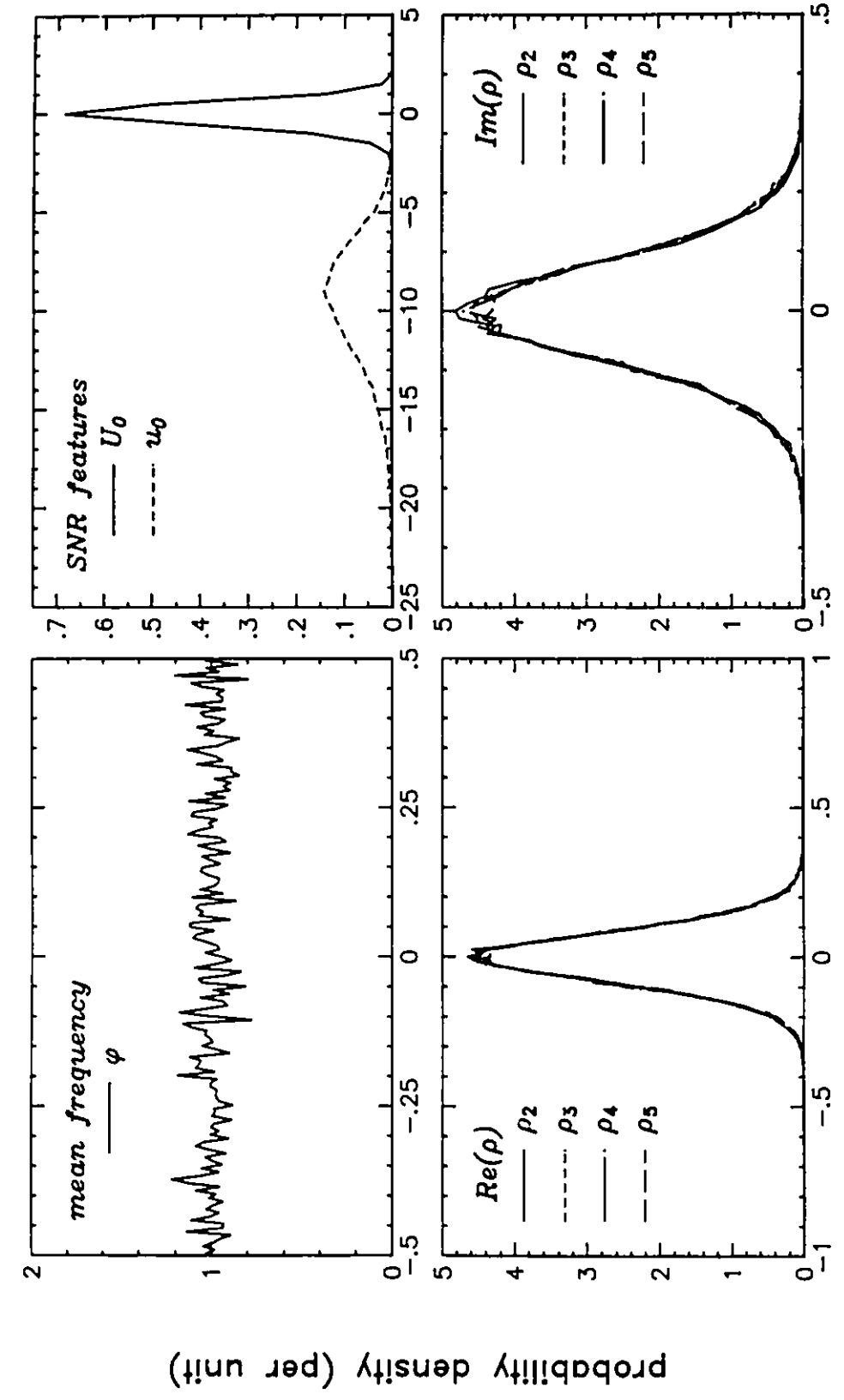


Figure 6.4a Marginal probability densities of the feature set of thermal noise

Feature value units are relative frequency for ϕ, dB for U_0 and u_0, and actual values for ρ_m.

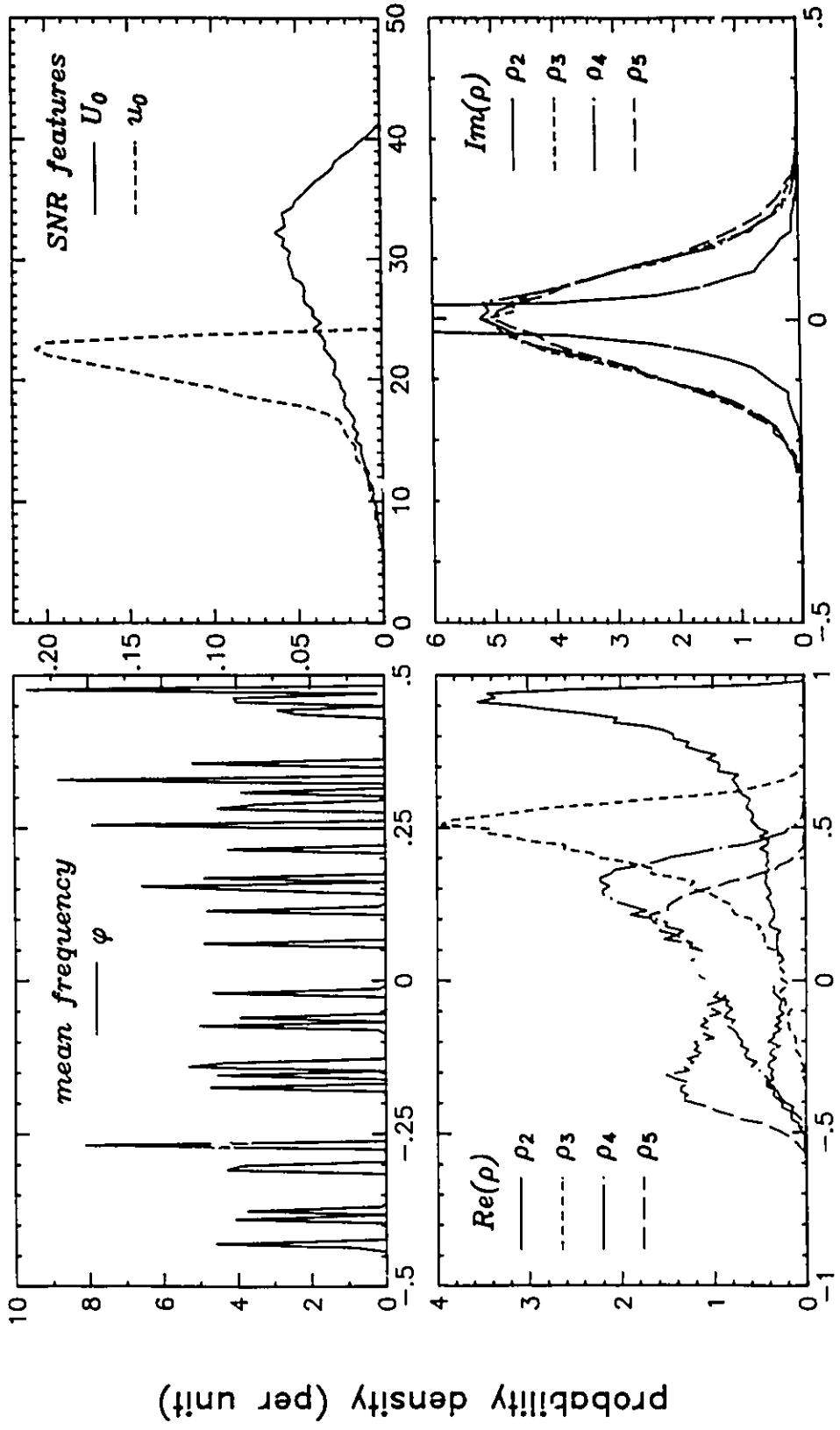


Figure 6.4b Marginal probability densities of the feature set of synthetic targets

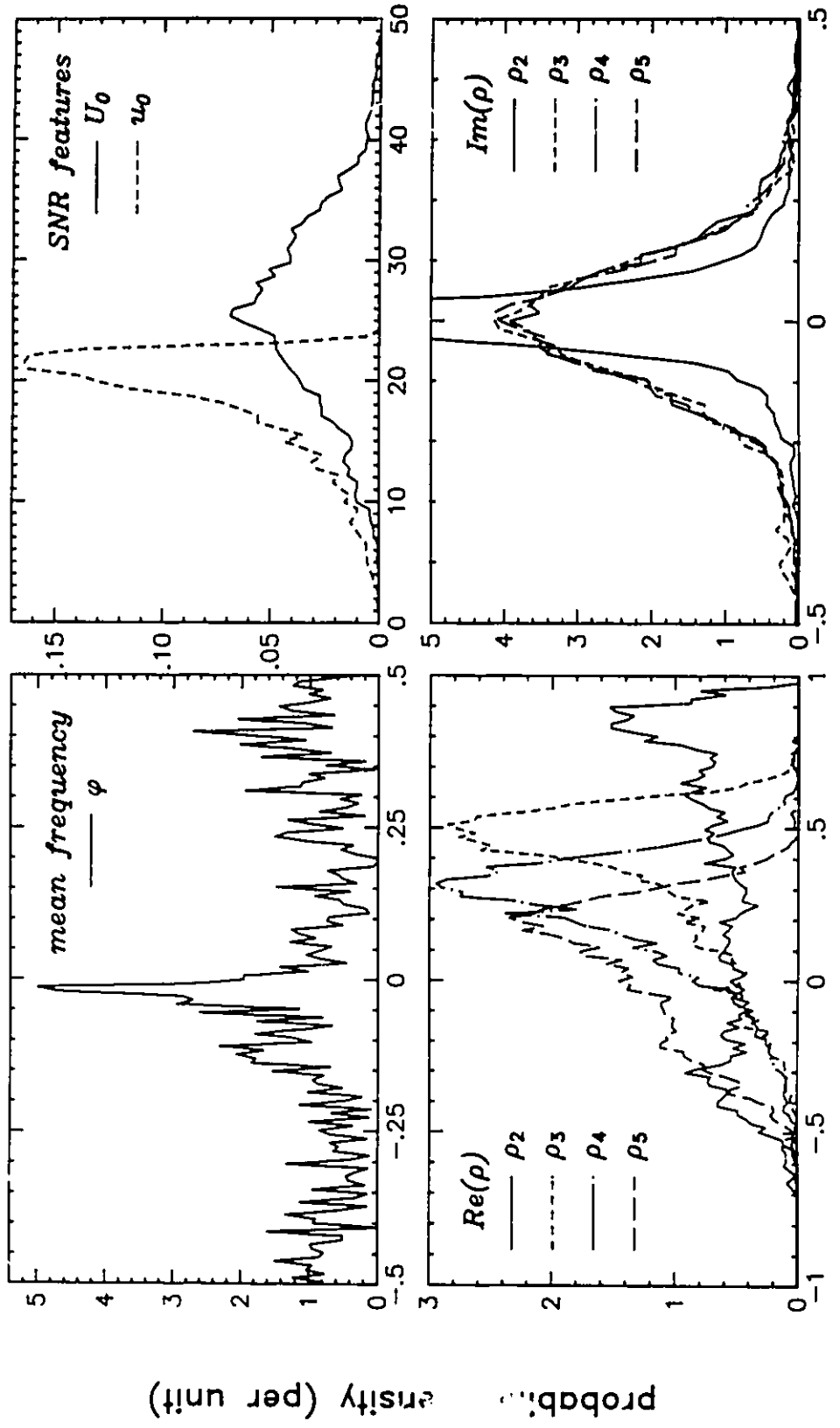


Figure 6.4c Marginal probability densities of the feature set of aircraft targets

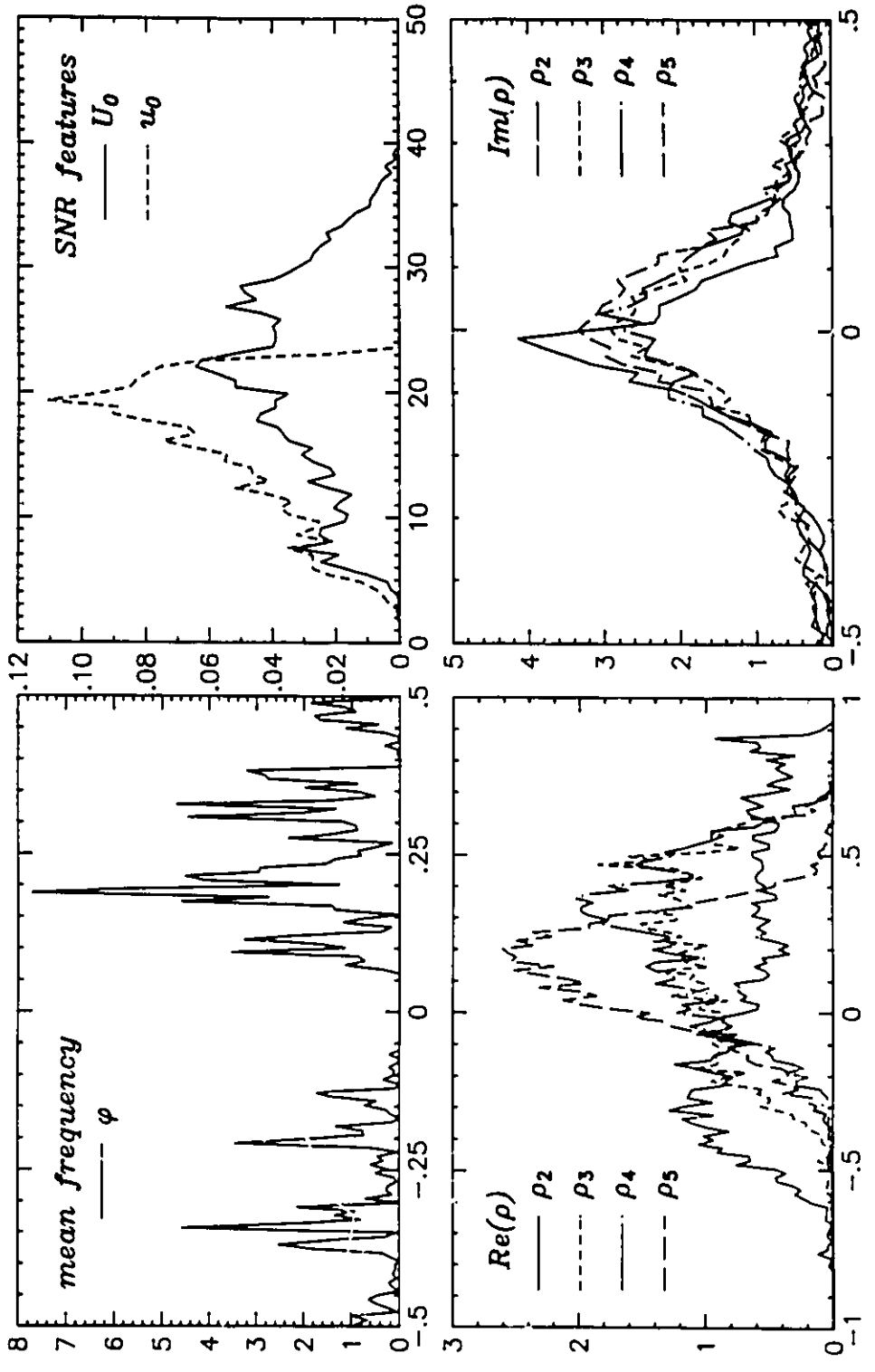


Figure 6.4d Marginal probability densities of the feature set of cars and trucks

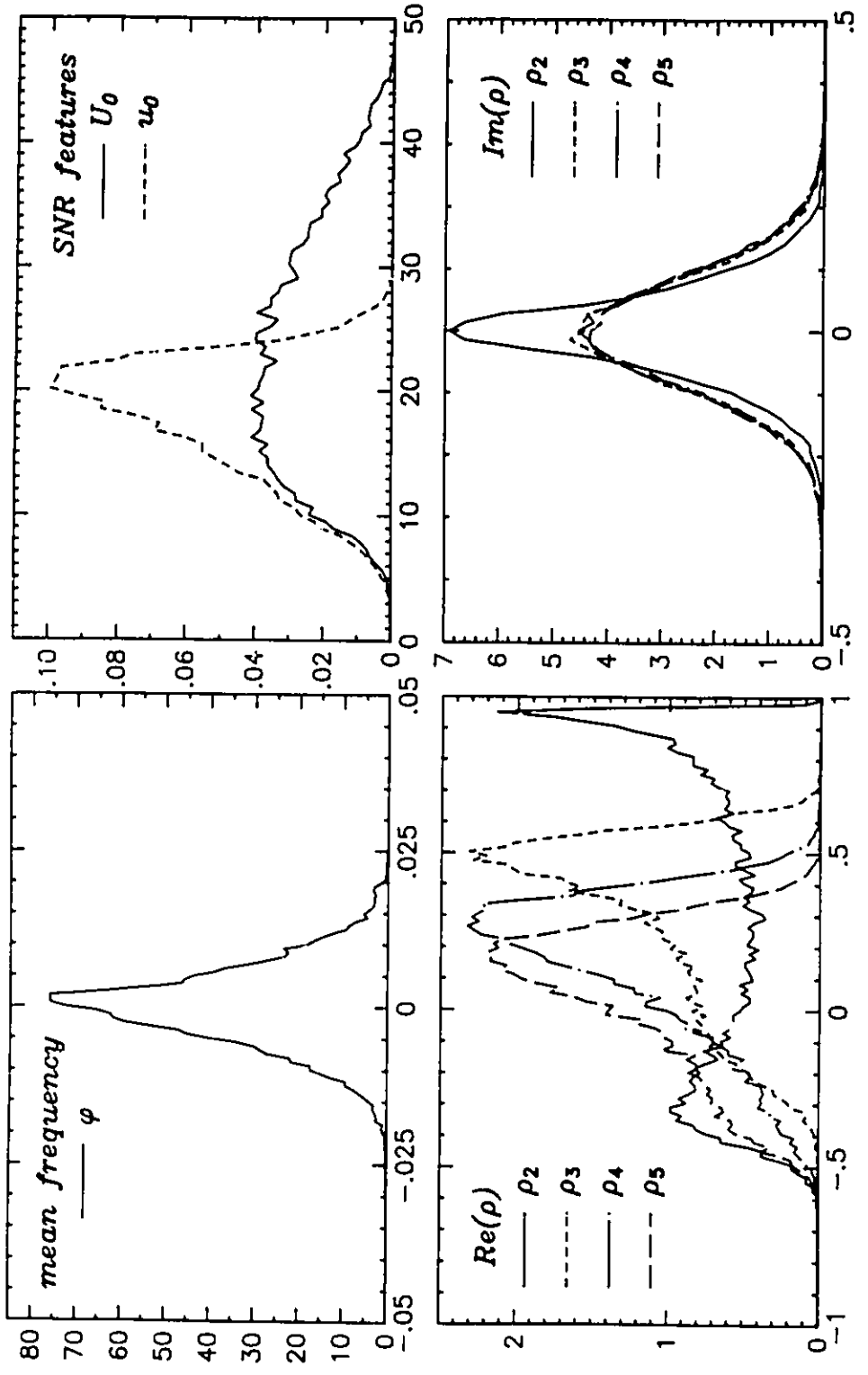


Figure 6.4e Marginal probability densities of the feature set of ground clutter

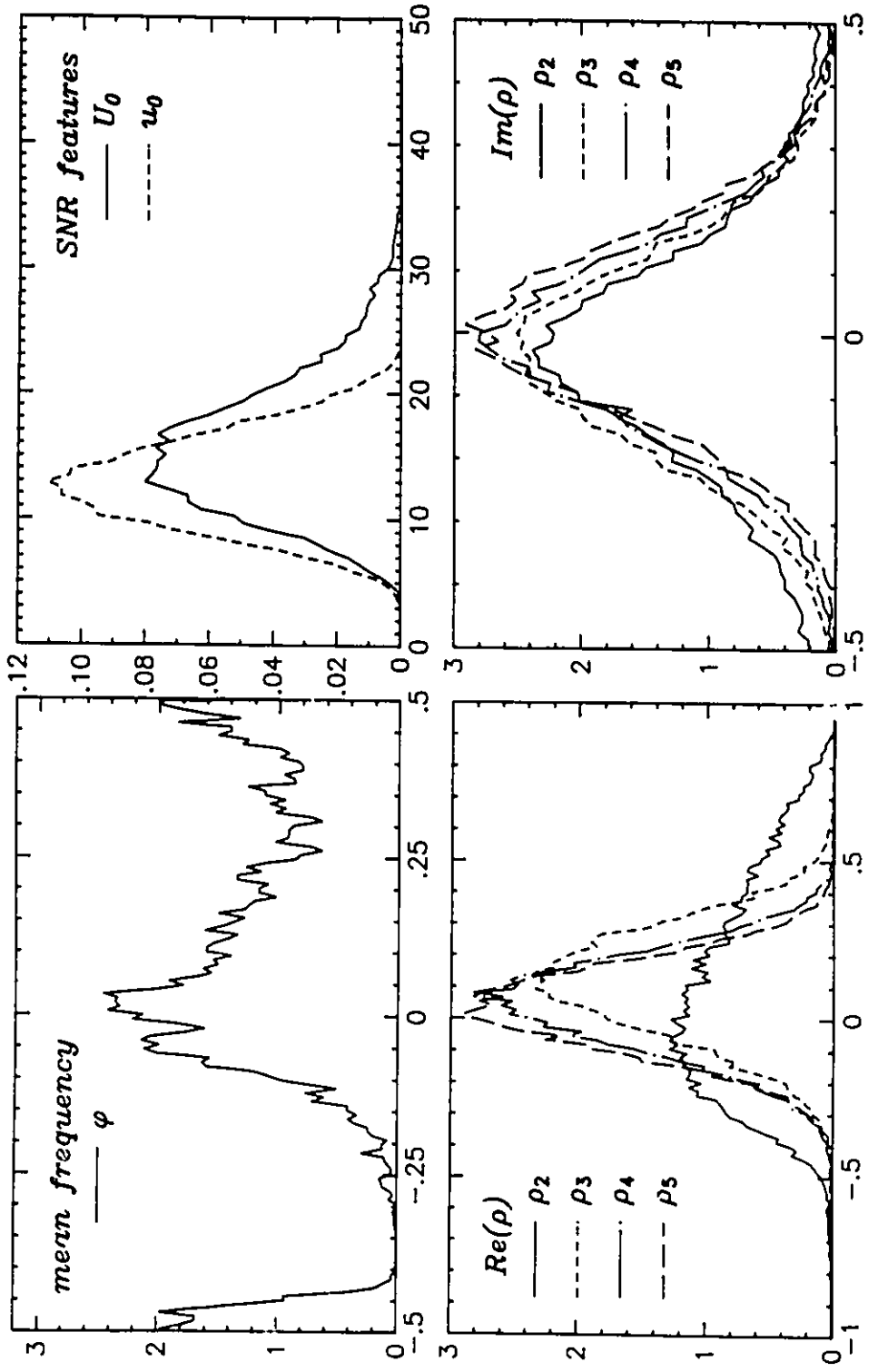


Figure 6.4f Marginal probability densities of the feature set of bird clutter

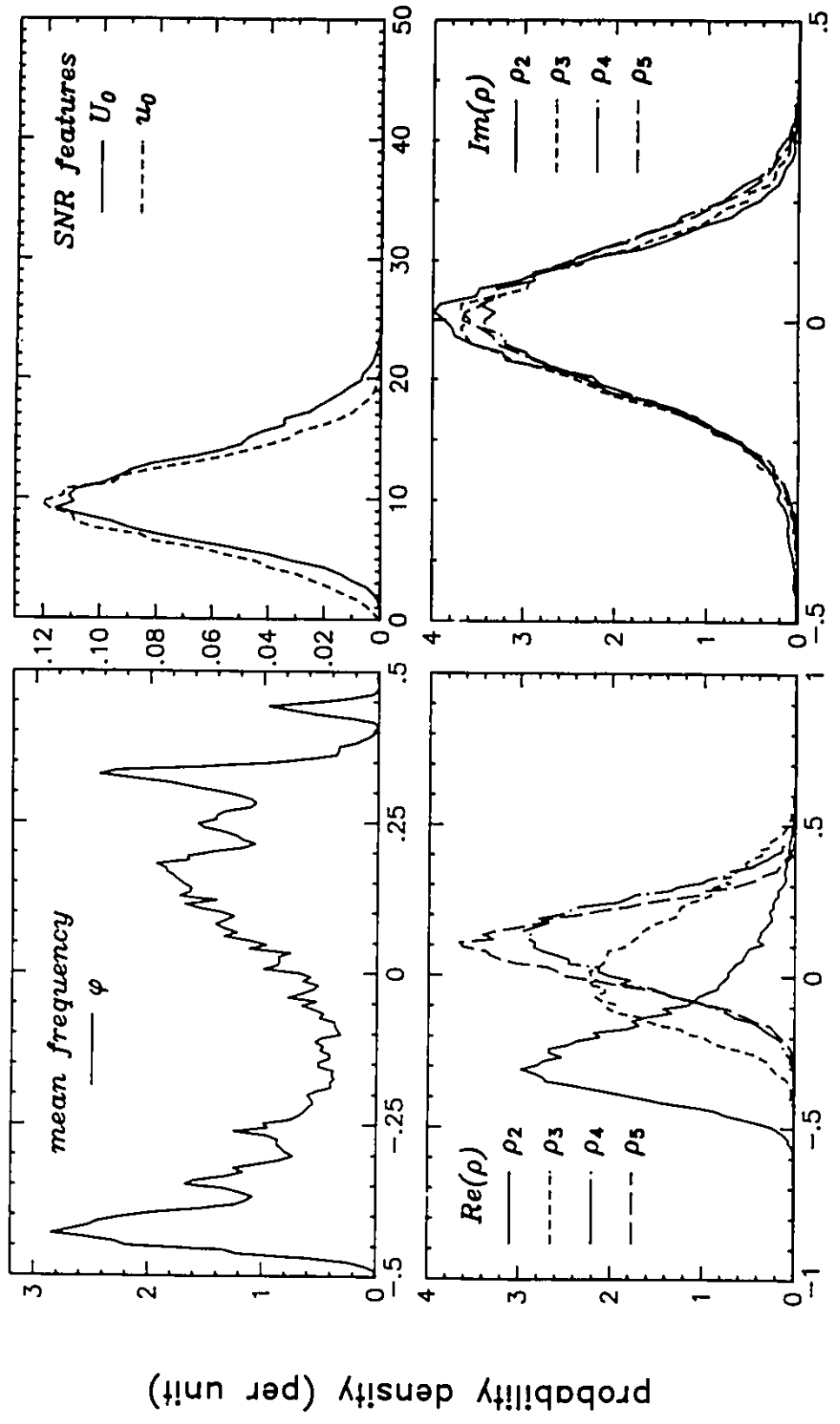


Figure 6.4g Marginal probability densities of the feature set of rain clutter

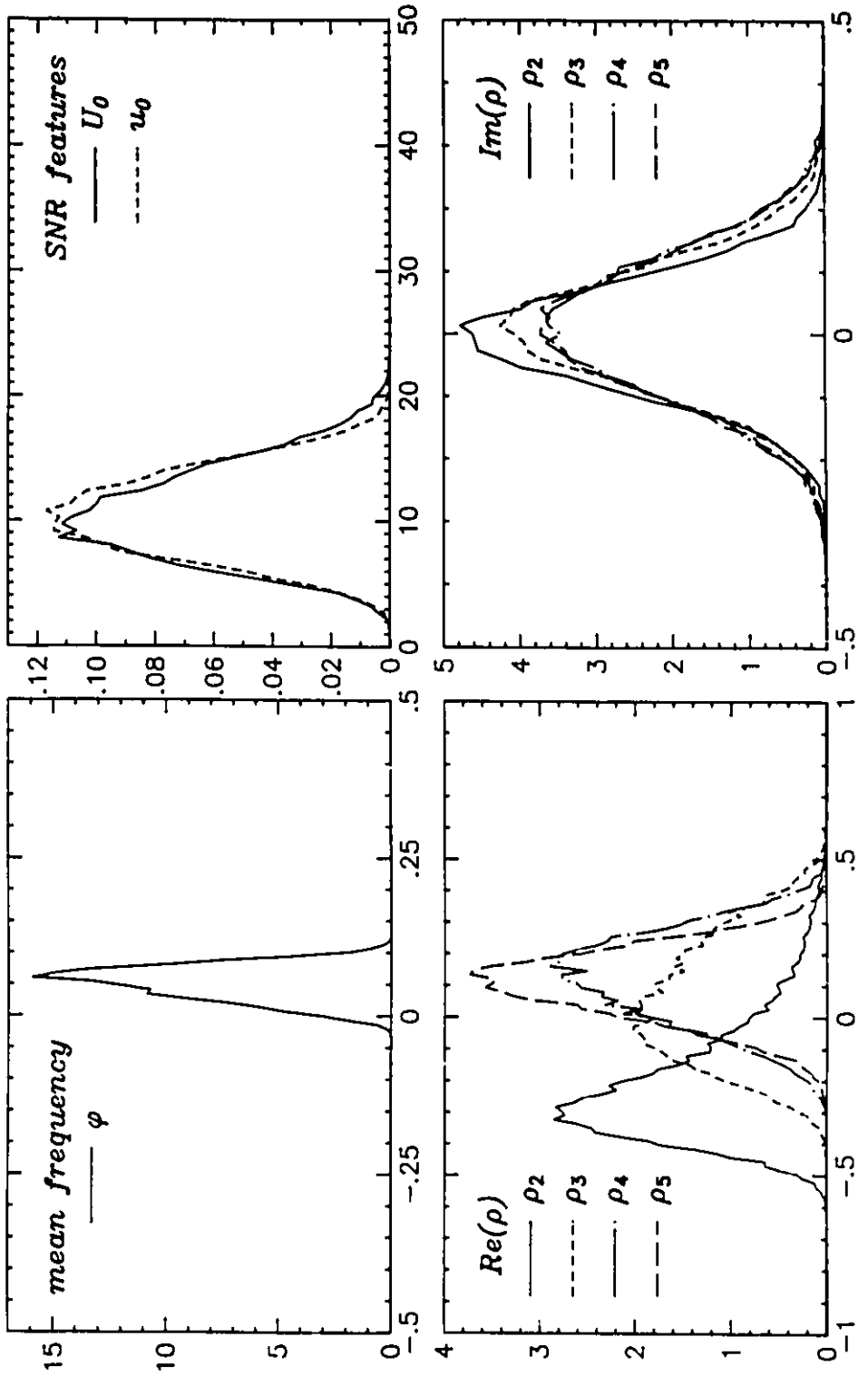


Figure 6.4h Marginal probability densities of the feature set of thunderstorms

coefficients is normal, and that of the Doppler frequency is uniform. Another reference, derived from the synthetically generated targets, shows the nature of the deviations from the normal density under well-controlled conditions. It should be noted here that the mostly triangular distribution of the SNR based feature U_0, and the multimodal distribution of ϕ are due solely to the way the targets were generated, and have little to do with the nature of the feature extractor.

In some cases, particularly with aircraft, cars and trucks, and ground clutter classes, the assumption of normal behavior for the reflection coefficients is clearly suspect. The normality transformation for $|\rho_1|$ into the feature u_0 appears to work quite well, but it also clear that such a transformation would not work as well for the bimodally distributed higher-order coefficients, although the marginal densities alone do not provide sufficient information to establish the reason for this.

Figure 6.5 shows the marginal densities after applying the orthonormal transformation discussed later in Section 6.3.1 (part of Equation 6.40). The number of features remains unchanged, but the physical meaning of the resulting features is lost. It is interesting to note the degree to which this transformation succeeds in producing marginal normal behavior. Some notable exceptions are the low-order features of the point target classes, that is the synthetic and actual aircraft targets, cars and trucks, and ground clutter. Again, there is insufficient information present in the marginal densities to assess the reason for this.

To examine the multivariate statistics, the first step was to look at the dependencies between features. A complete matrix of scatter diagrams, each plotting one feature against any and all others, would require $d(d-1)/2$ individual plots, or 190 if each of the 20 features shown in Figure 6.1 is to be examined. For presentation purposes, in the sequence of plots shown in Figures 6.6 to 6.8 only six features are

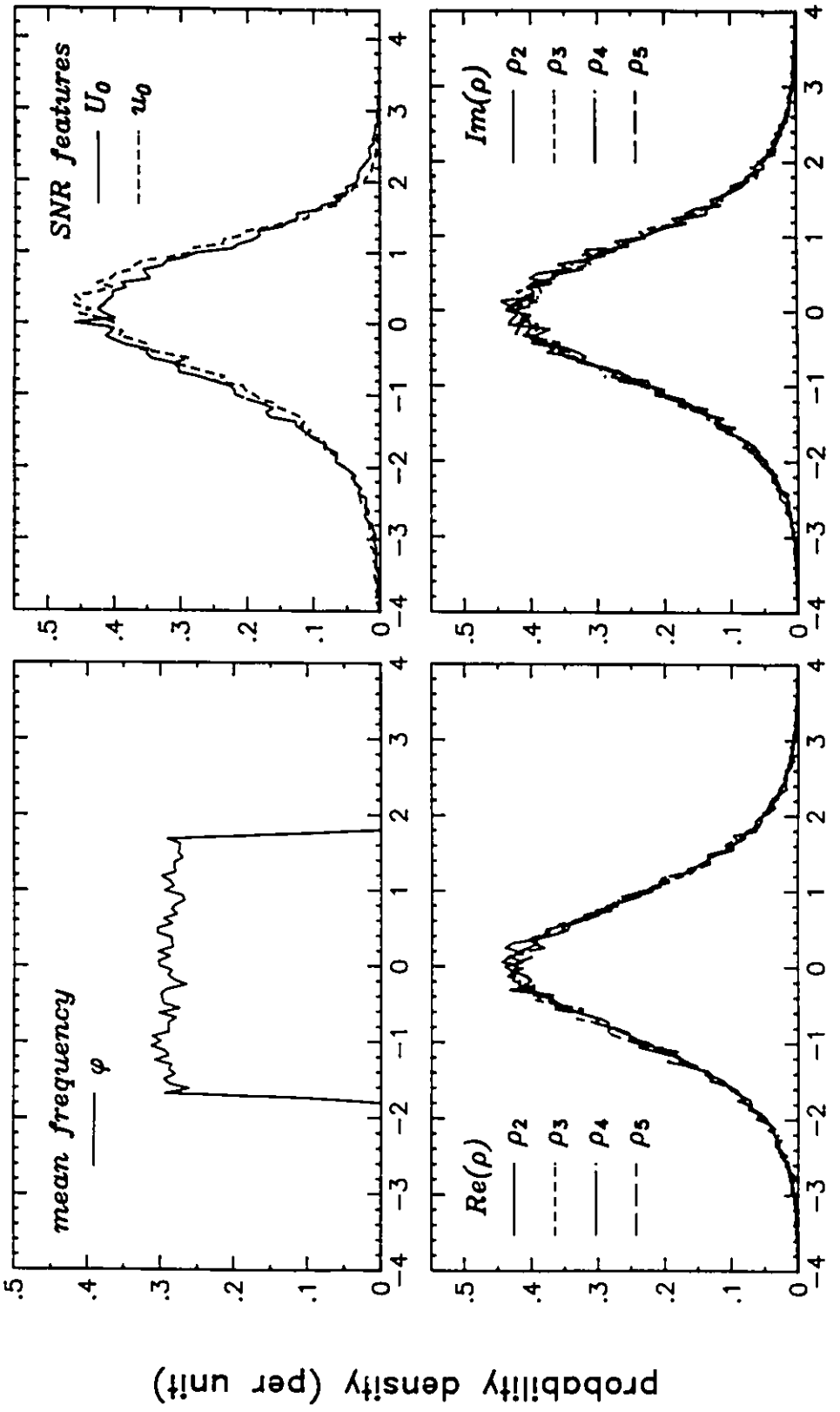


Figure 6.5a Marginal densities of the orthonormal feature set of thermal noise

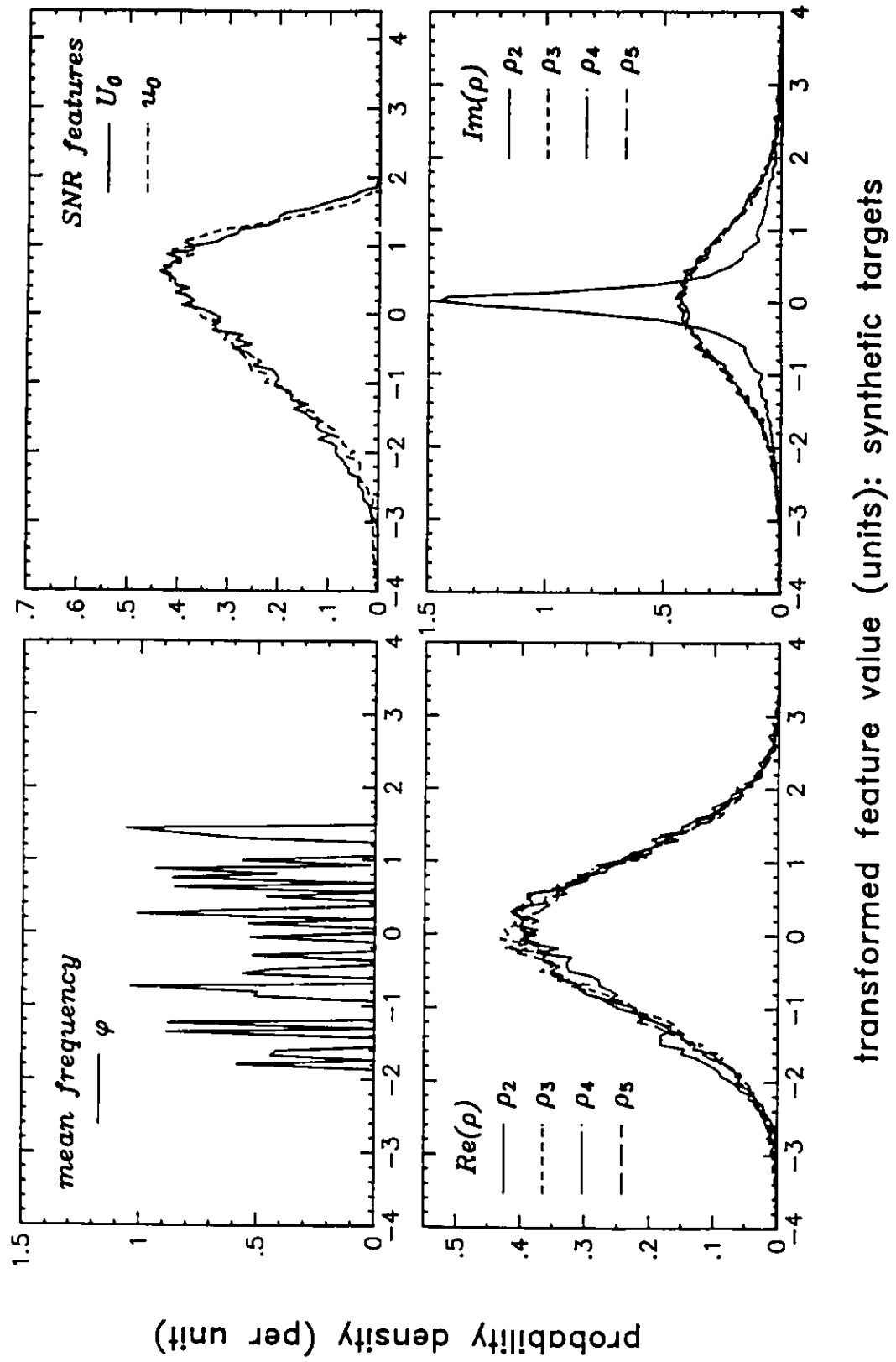


Figure 6.5b Marginal densities of the orthonormal feature set of synthetic targets

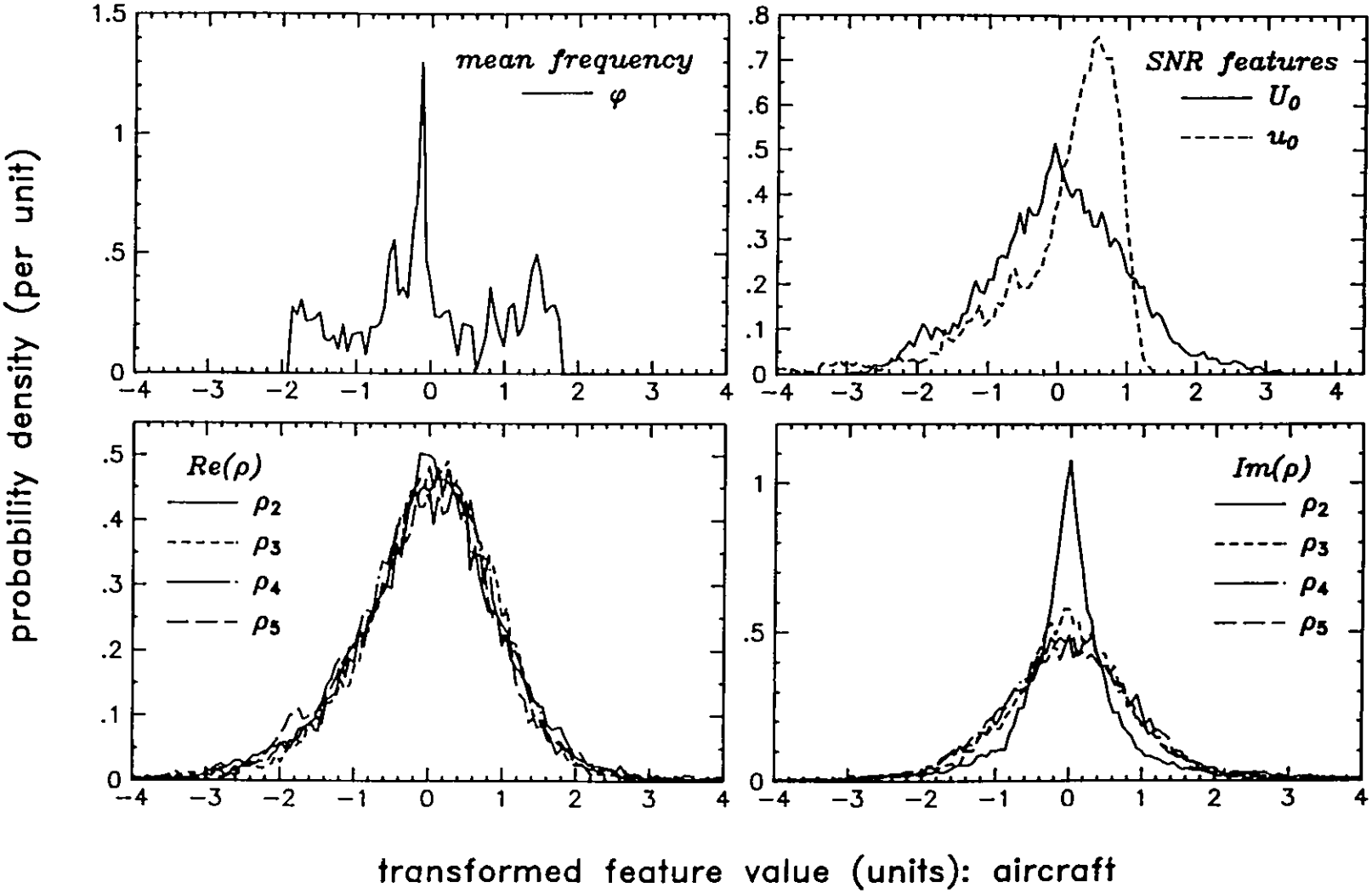


Figure 6.5c Marginal densities of the orthonormal feature set of aircraft targets

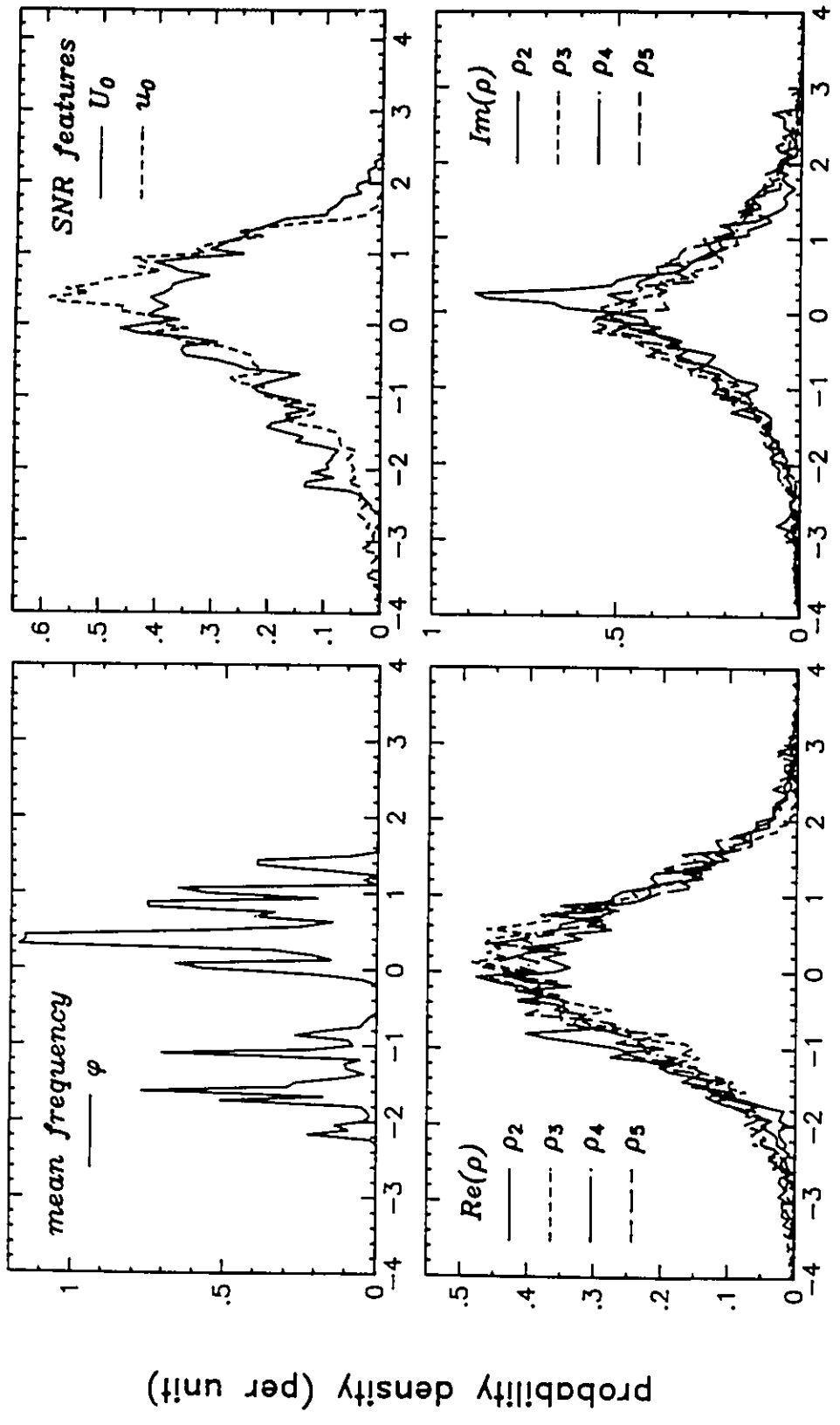


Figure 6.5d Marginal densities of the orthonormal feature set of cars and trucks

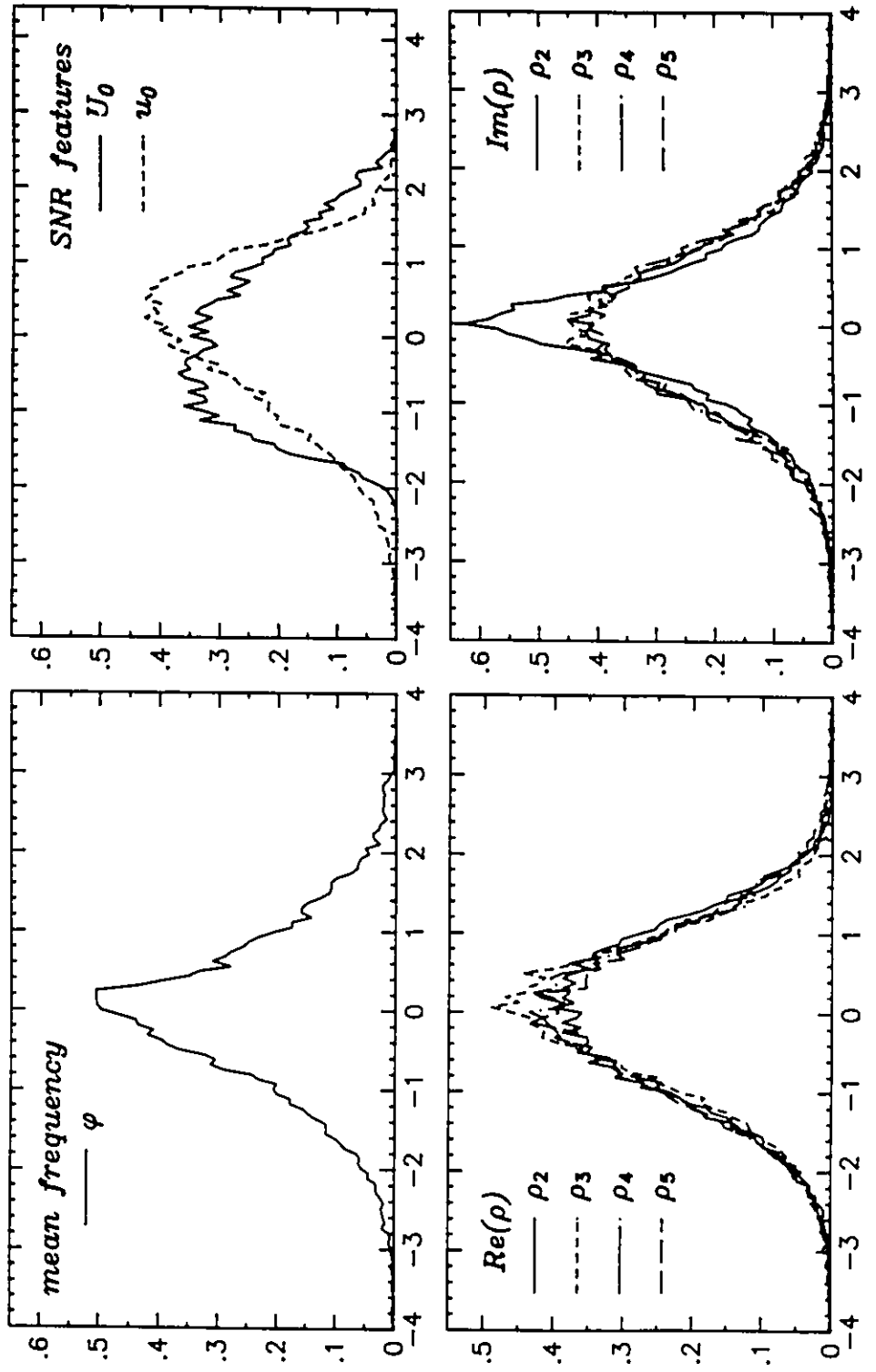


Figure 6.5e Marginal densities of the orthonormal feature set of ground clutter

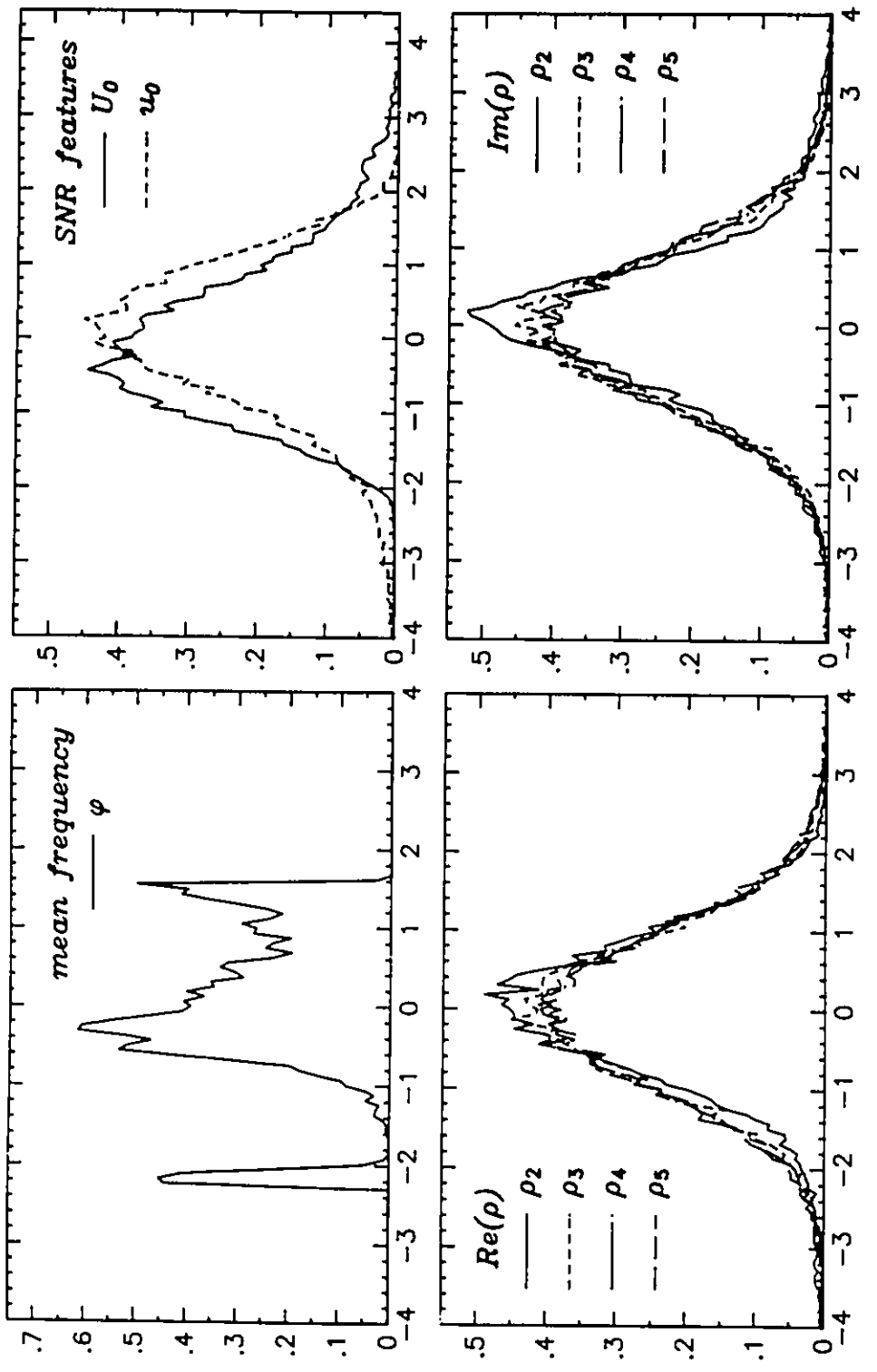


Figure 6.5f Marginal densities of the orthonormal feature set of bird clutter

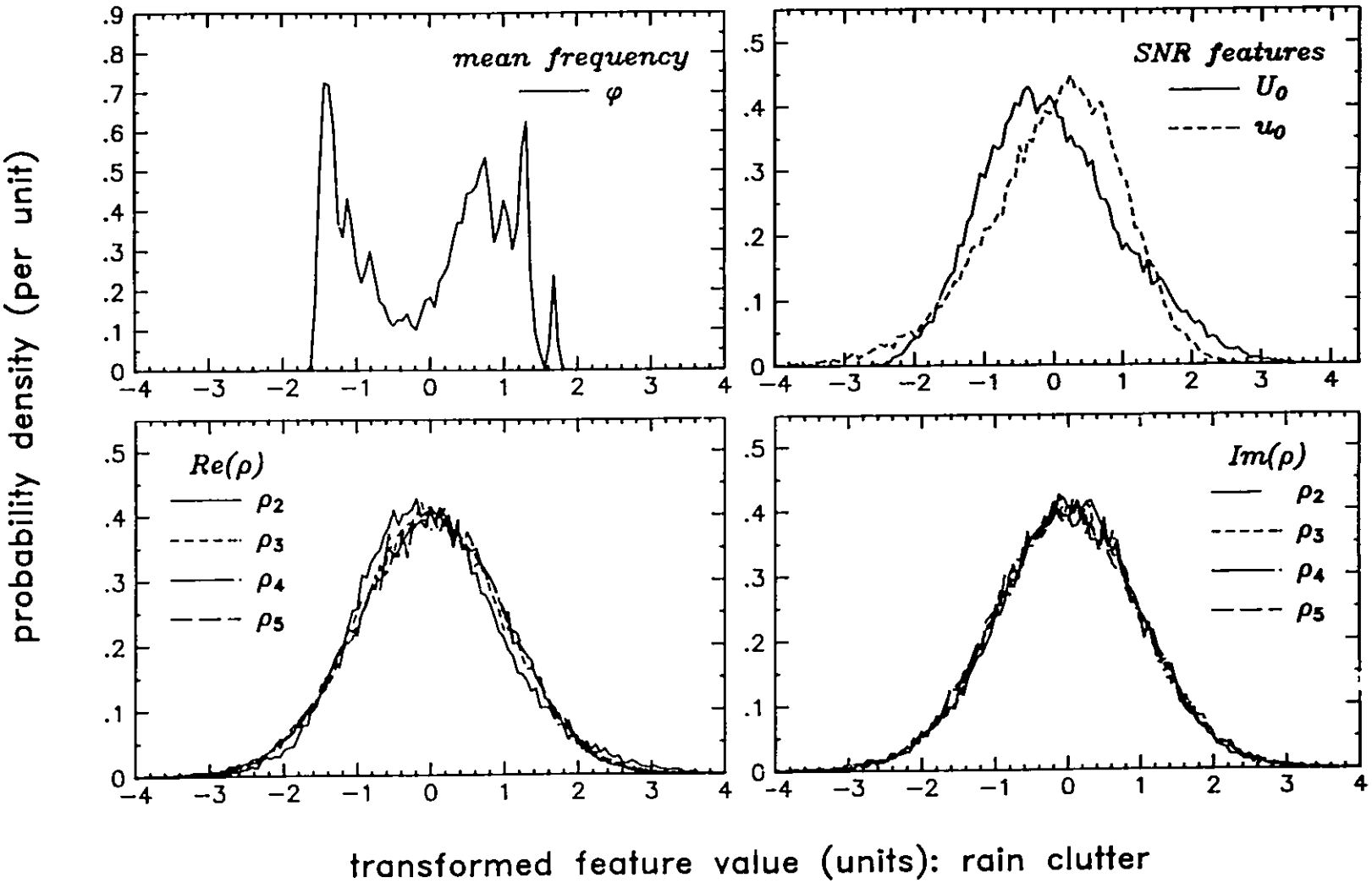


Figure 6.5g Marginal densities of the orthonormal feature set of rain clutter

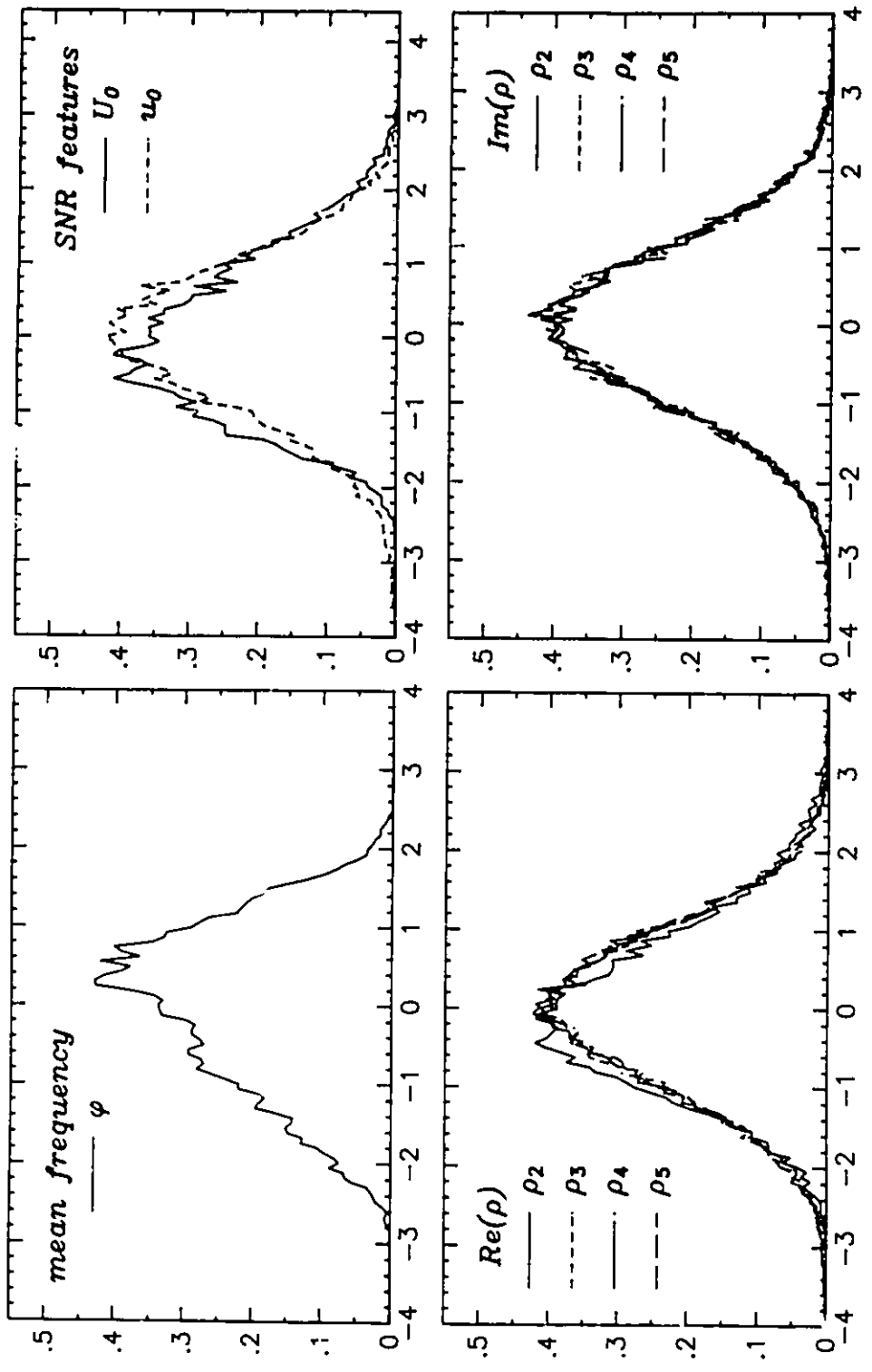


Figure 6.5h Marginal densities of the orthonormal feature set of thunderstorms

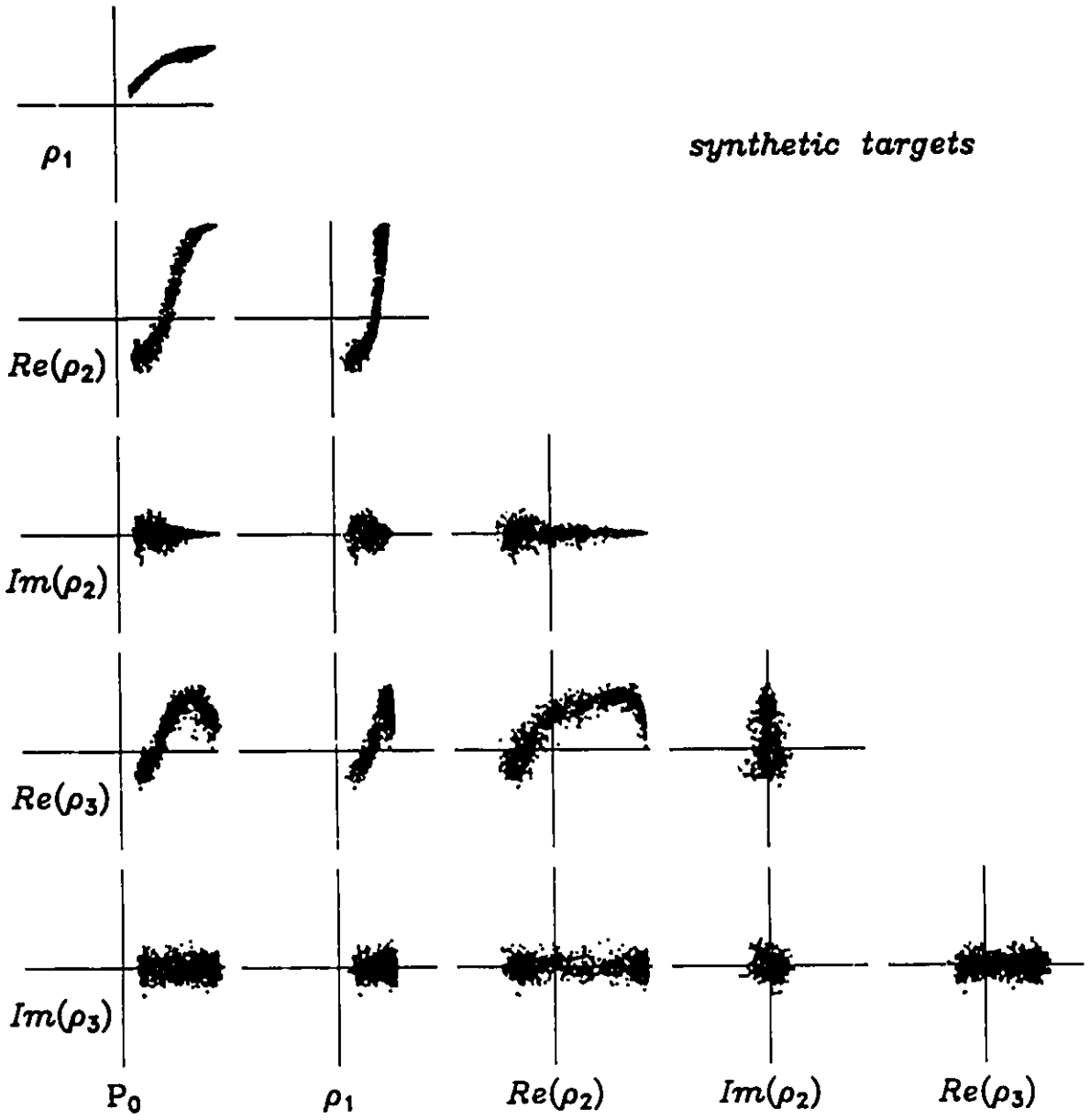


Figure 6.6a Partial matrix of scatter plots for synthetic targets

The matrix explores the correlations amongst the first three reflection coefficients and the signal strength P_0. The strongest correlations appear to be between the real parts of the coefficients, with very little between the imaginary parts.

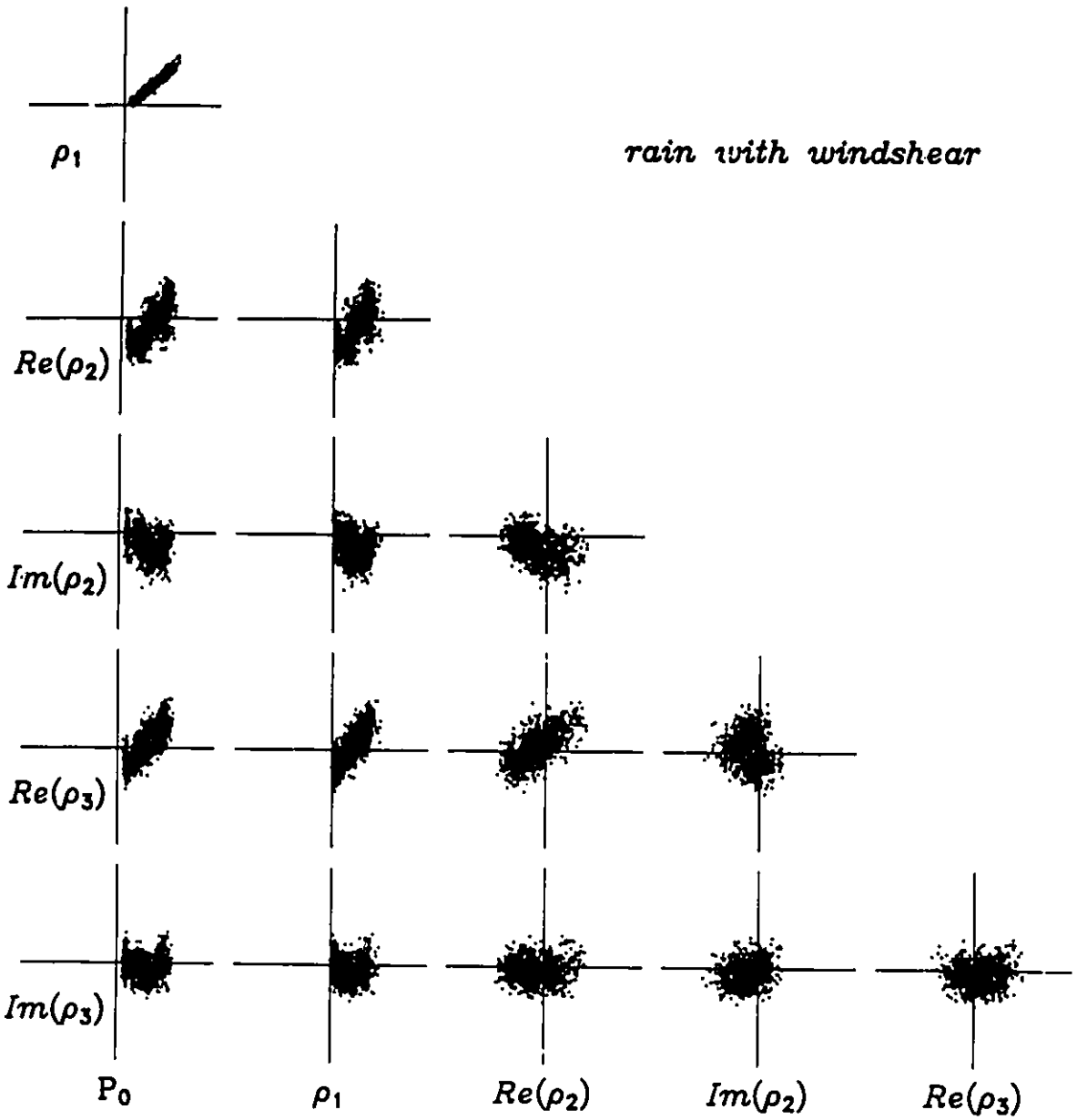


Figure 6.6b Partial matrix of scatter plots for rain with windshear

Note the greater variability of the clusters compared to Figure 6.6a. The asymmetry in the spectrum leads to more correlation between the imaginary parts and the signal strength P_0.

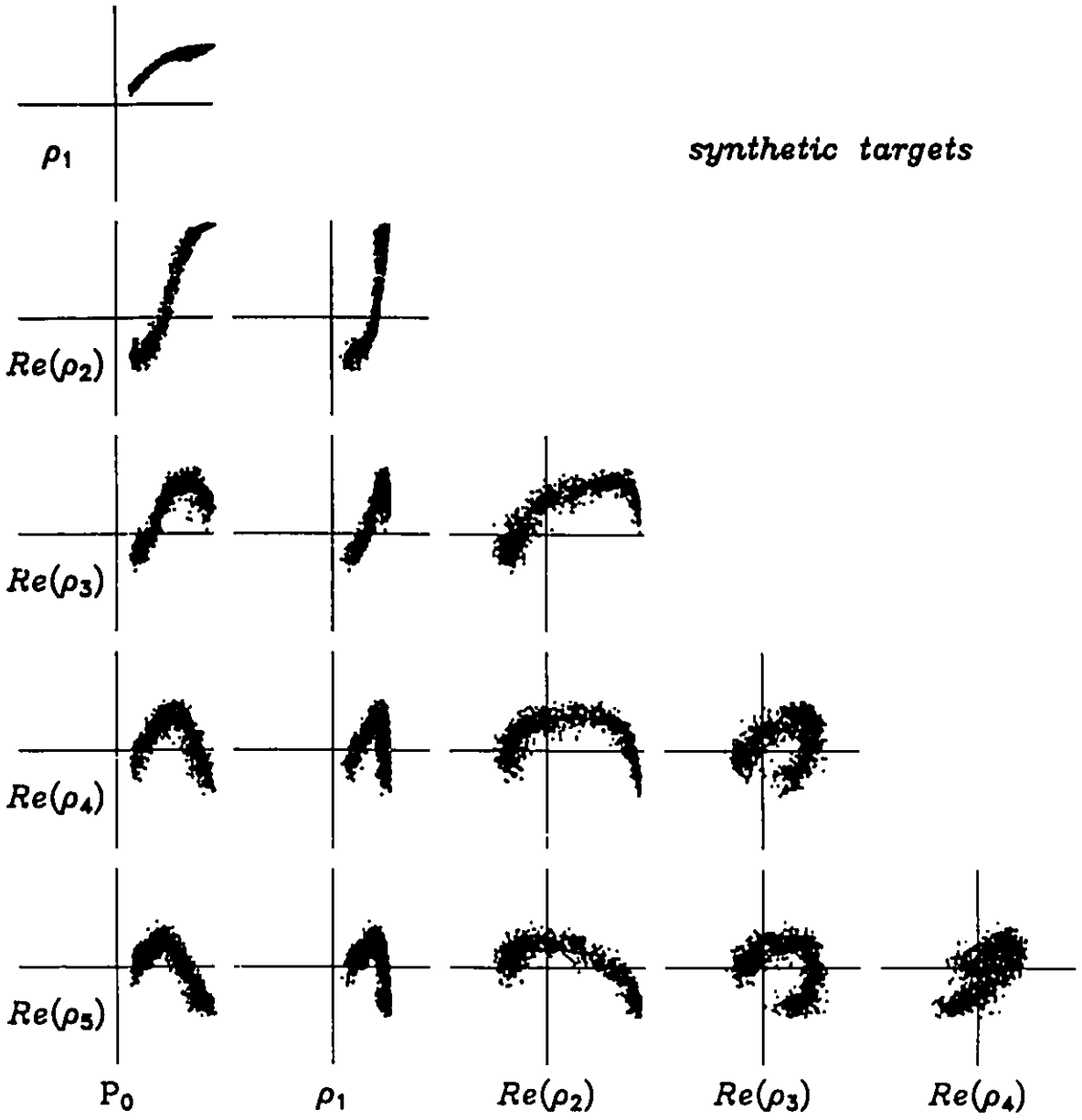


Figure 6.7a Matrix of scatter plots for synthetic targets

The matrix explores the correlations amongst the real parts of the first five reflection coefficients and the signal strength P_0. Note the considerable amount of information remaining in the coefficients ρ_4 and ρ_5.

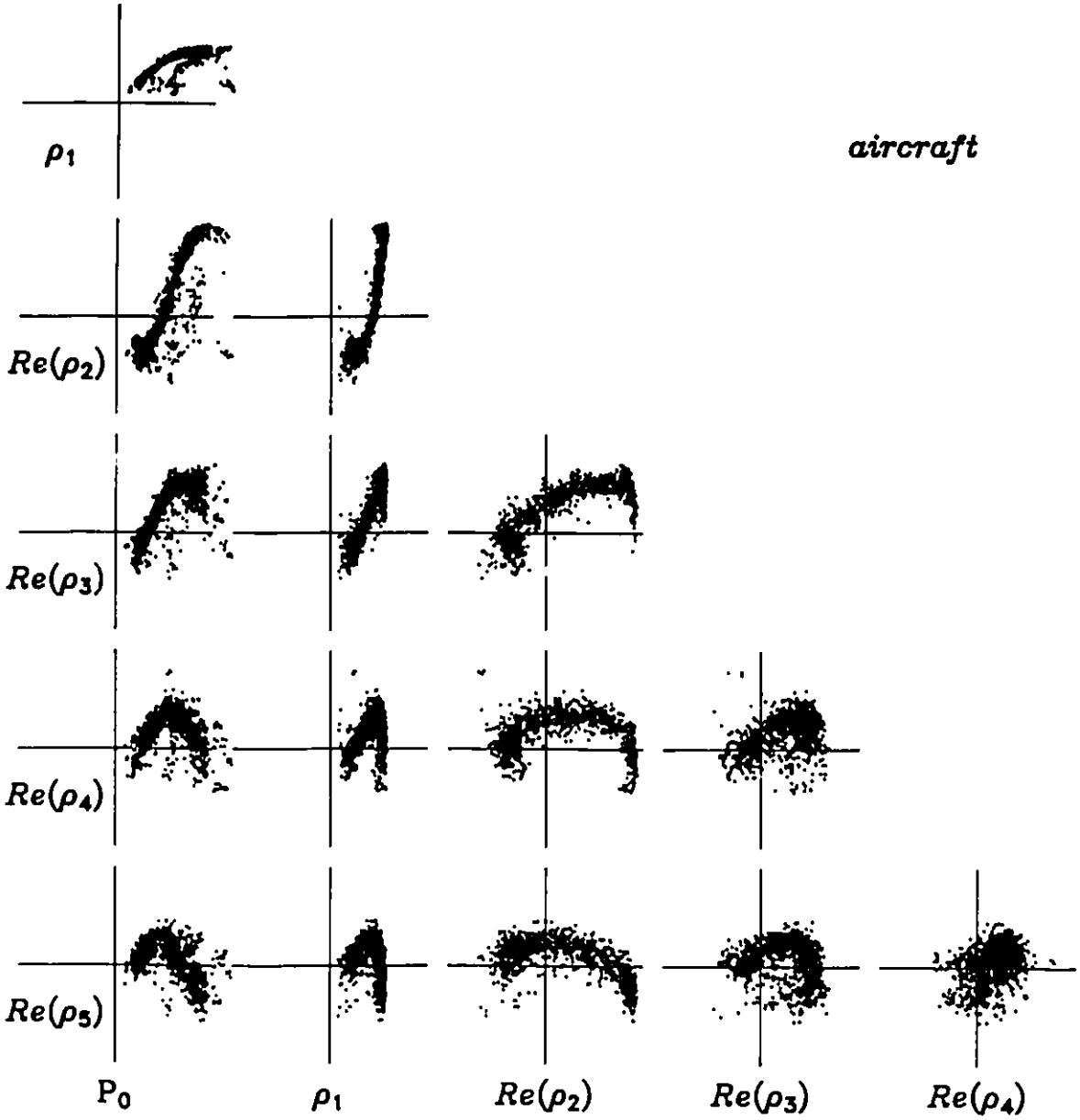


Figure 6.7b Matrix of scatter plots for aircraft targets

Note the similarity to the synthetic targets in Figure 6.7a. The greater variability stems from the internal motion of the aircraft, which was not modeled in the synthetic targets.

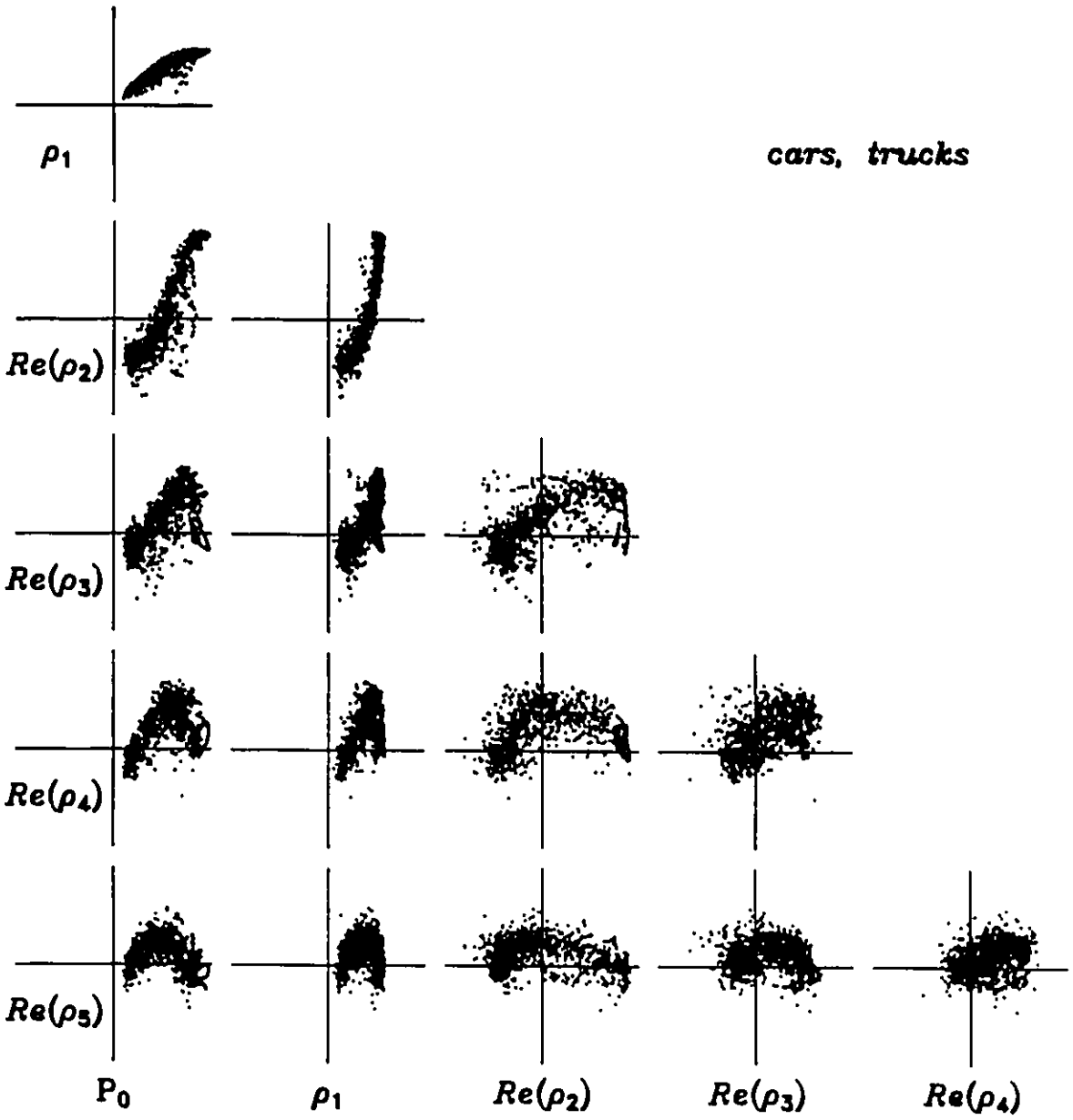


Figure 6.7c Matrix of scatter plots for cars and trucks

Larger deviations from the ideal target model are apparent in this class of targets, as is a significantly greater variability.

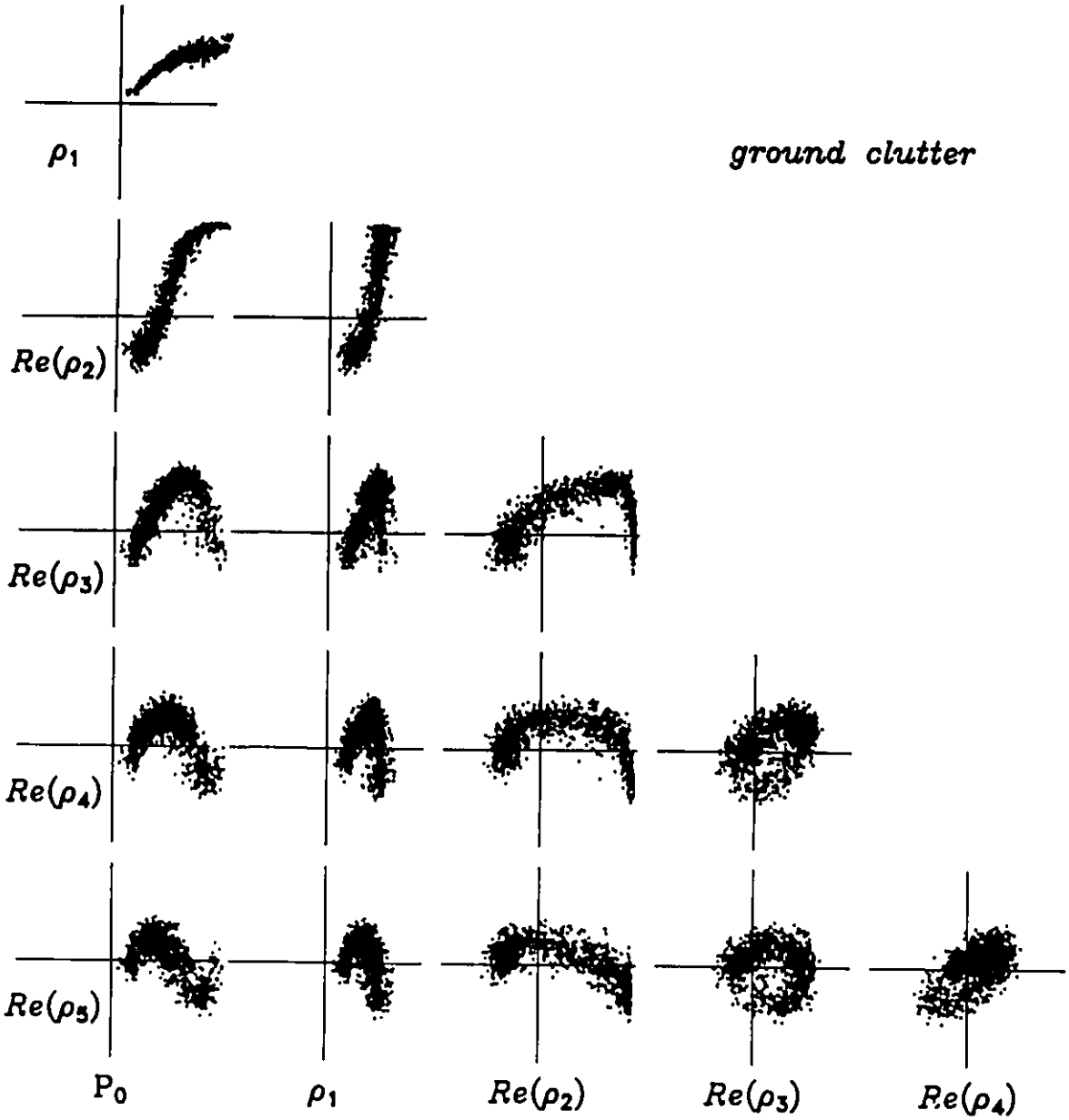


Figure 6.7d Matrix of scatter plots for ground clutter

Of the clutter classes, ground clutter displays the greatest similarity with targets.

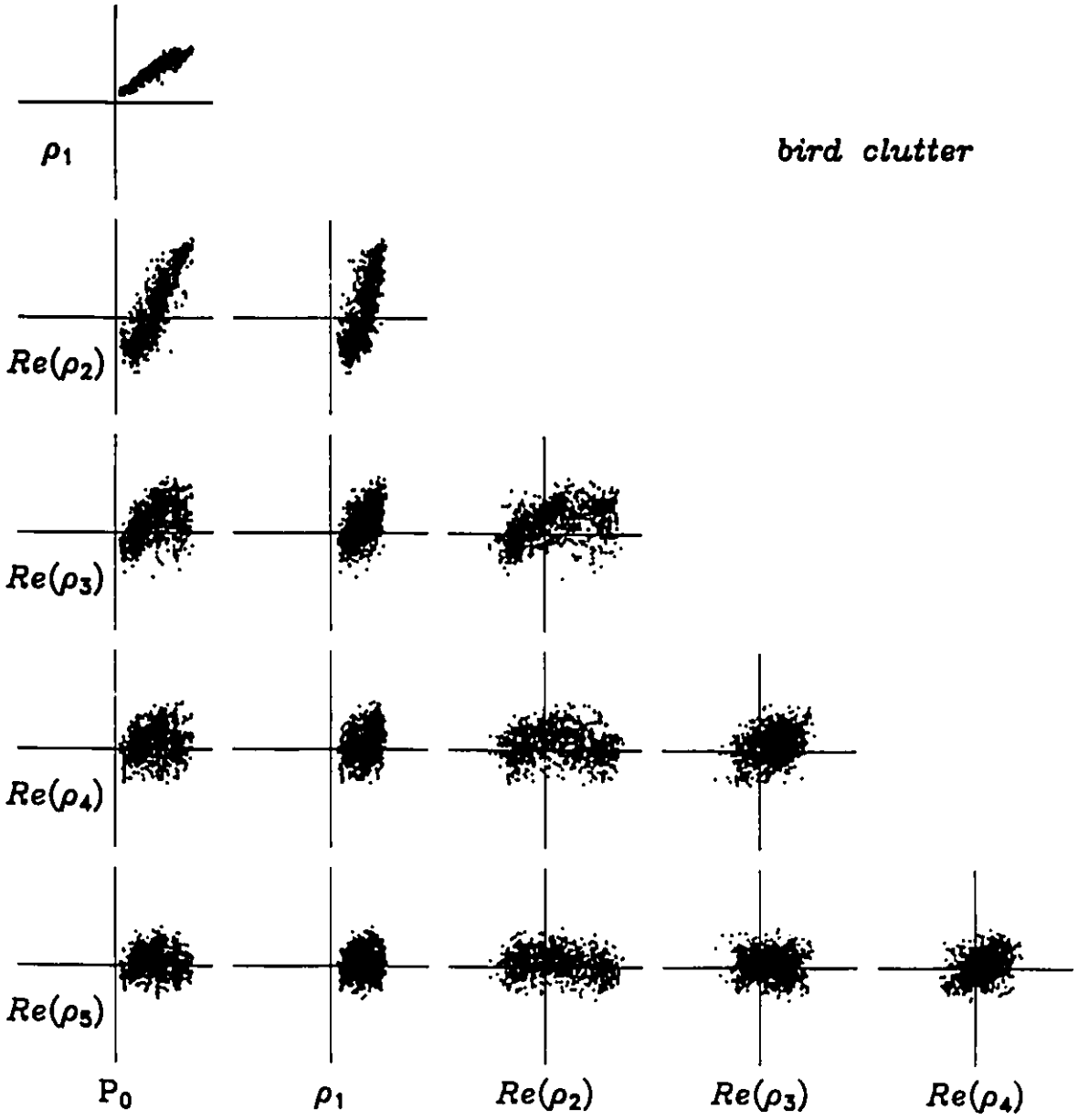


Figure 6.7e Matrix of scatter plots for bird clutter

The small amount of information remaining in the high-order coefficients ρ_4 and ρ_5 indicate a much lower AR model order is sufficient for birds.

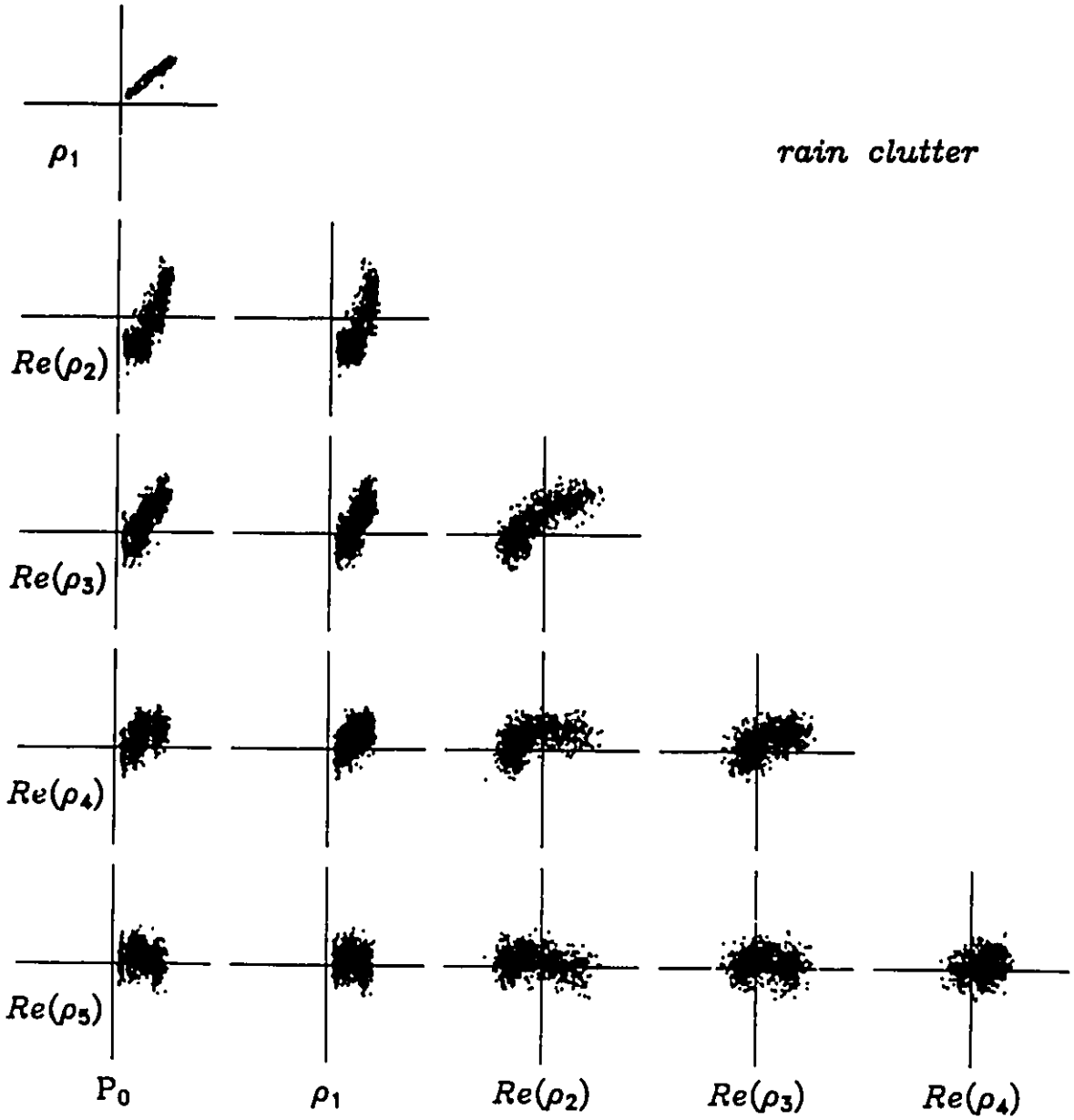


Figure 6.7f Matrix of scatter plots for rain clutter

Rain clutter appears similar to bird clutter but has less variability. The residual information remaining in ρ_4 and ρ_5 is thus much more obvious.

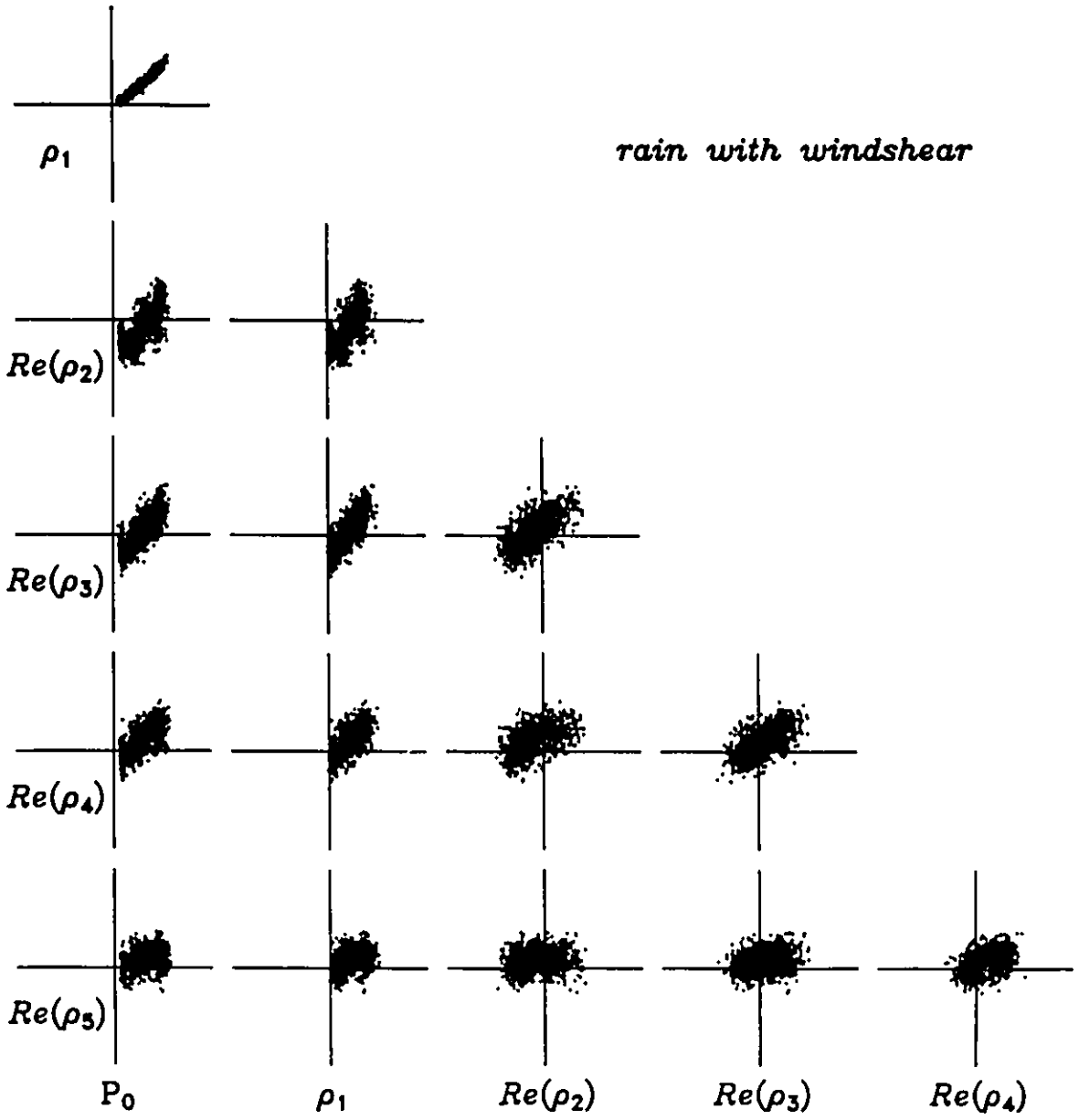


Figure 6.7g Matrix of scatter plots for rain clutter with windshear

The difference between this class and simple rain lies mainly in the imaginary component not shown here (see Figure 6.6b), but note the lower values for ρ_1.

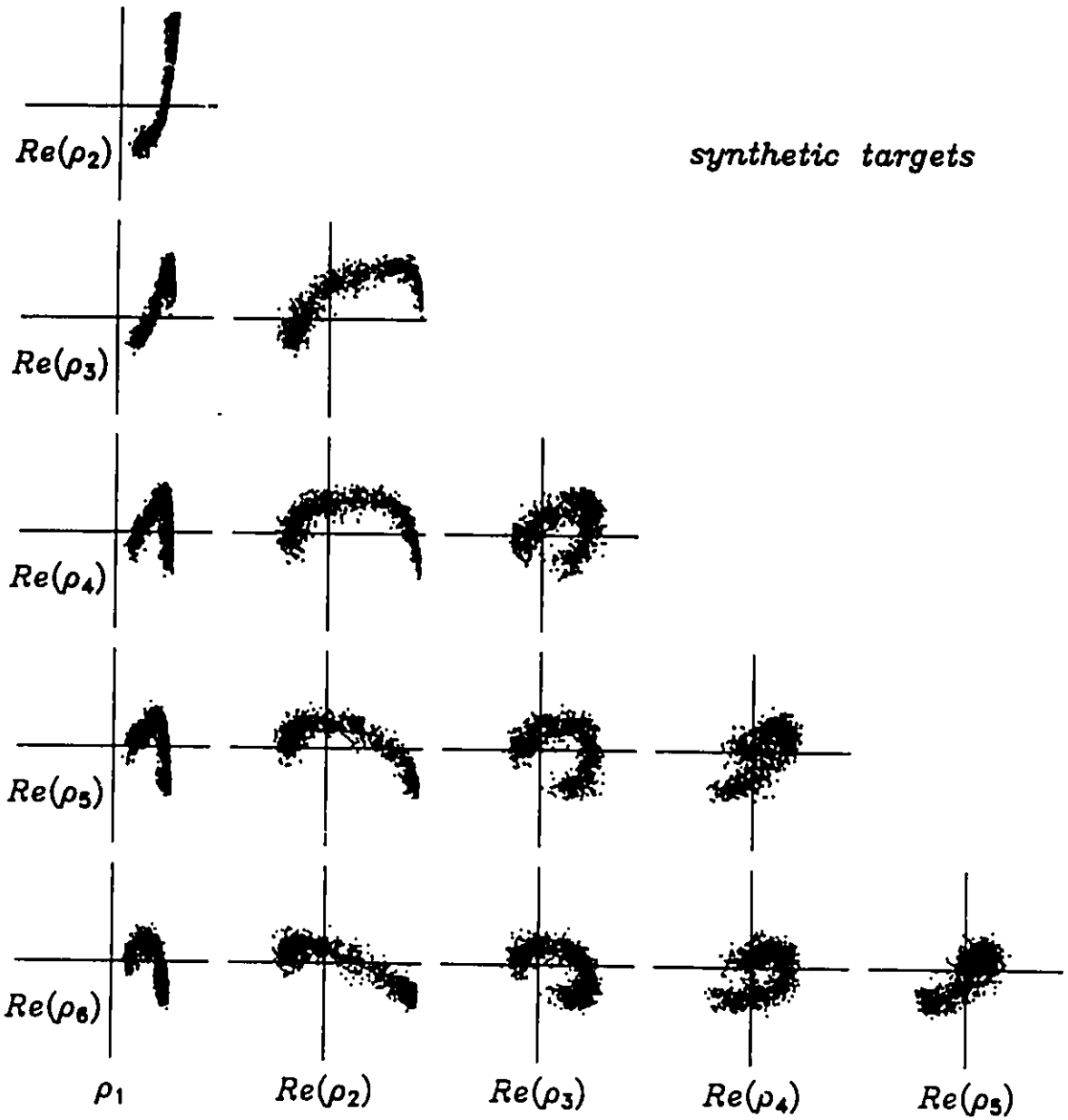


Figure 6.8a Matrix of scatter plots without P_0 for synthetic targets

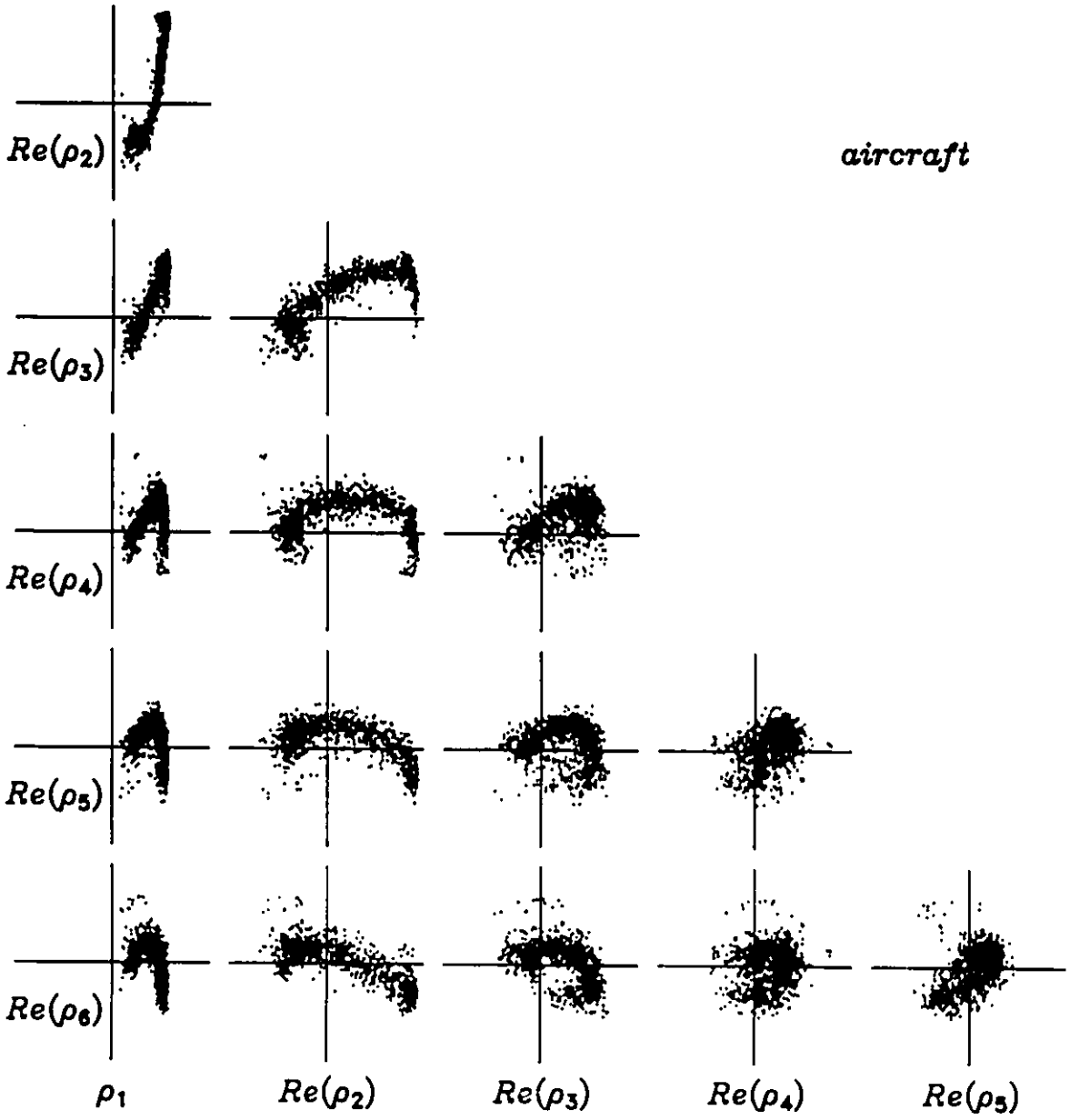


Figure 6.8b Matrix of scatter plots without P_0 for aircraft targets

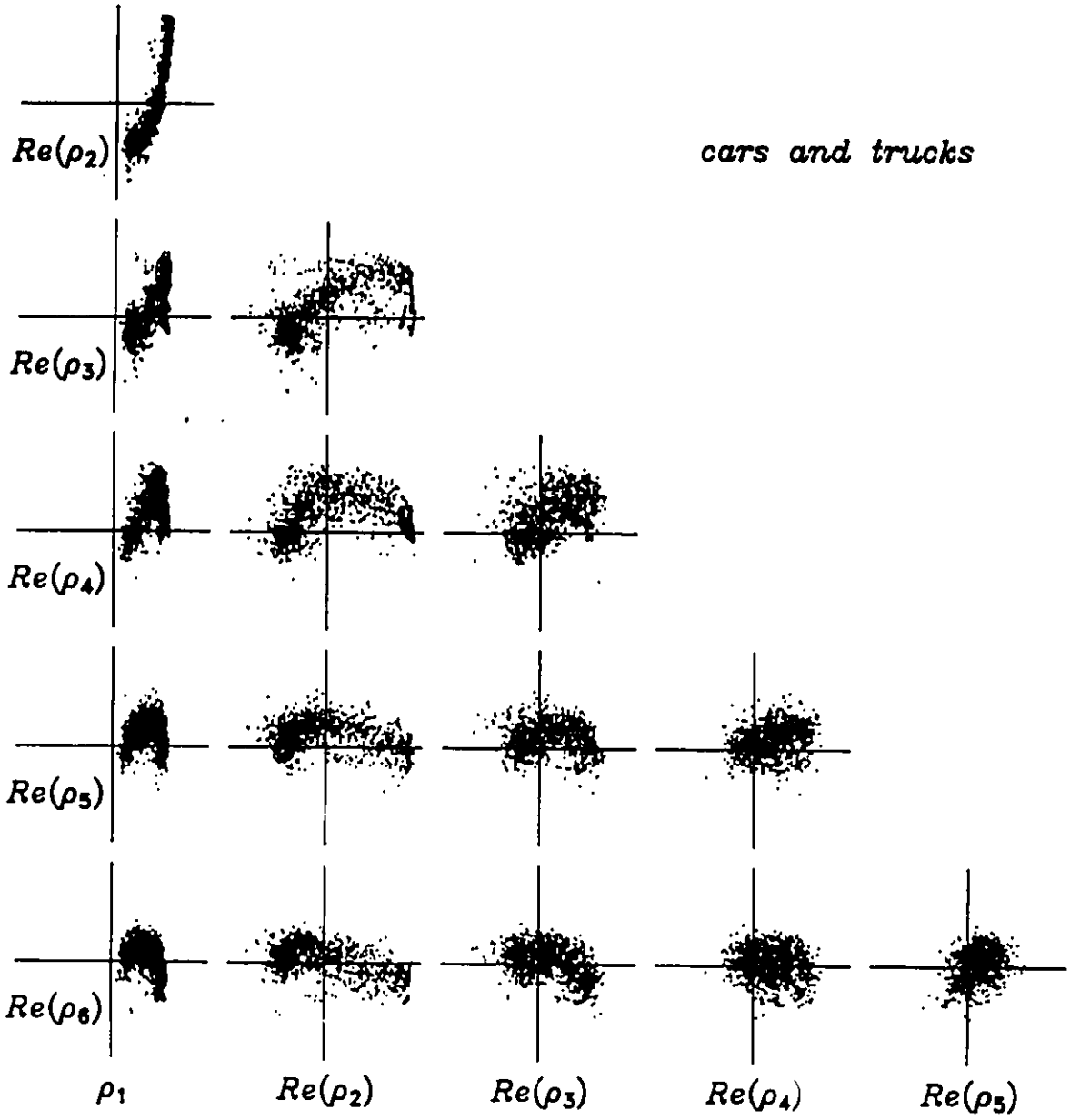


Figure 6.8c Matrix of scatter plots without P_0 for cars and trucks

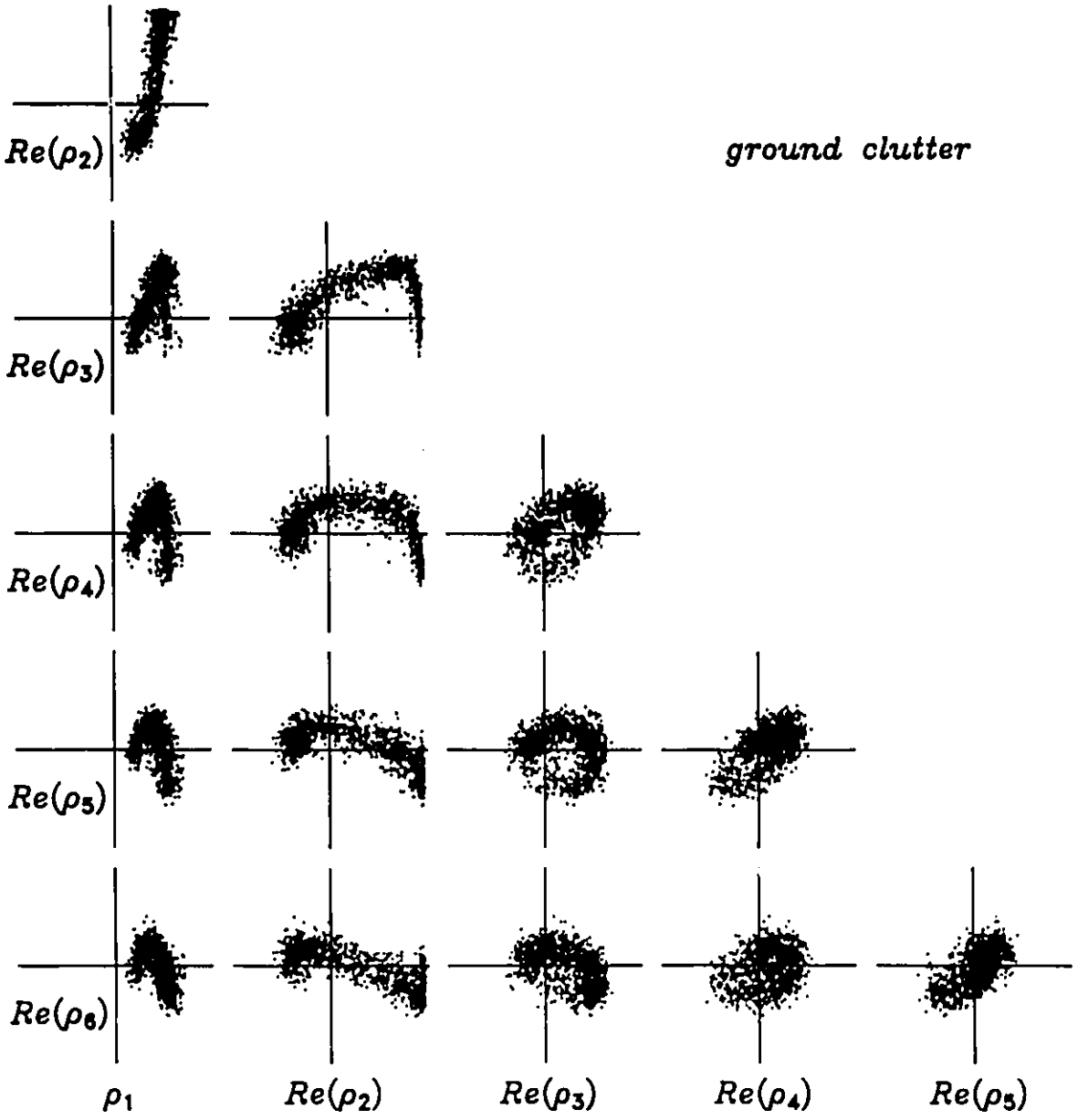


Figure 6.8d Matrix of scatter plots without P_0 for ground clutter

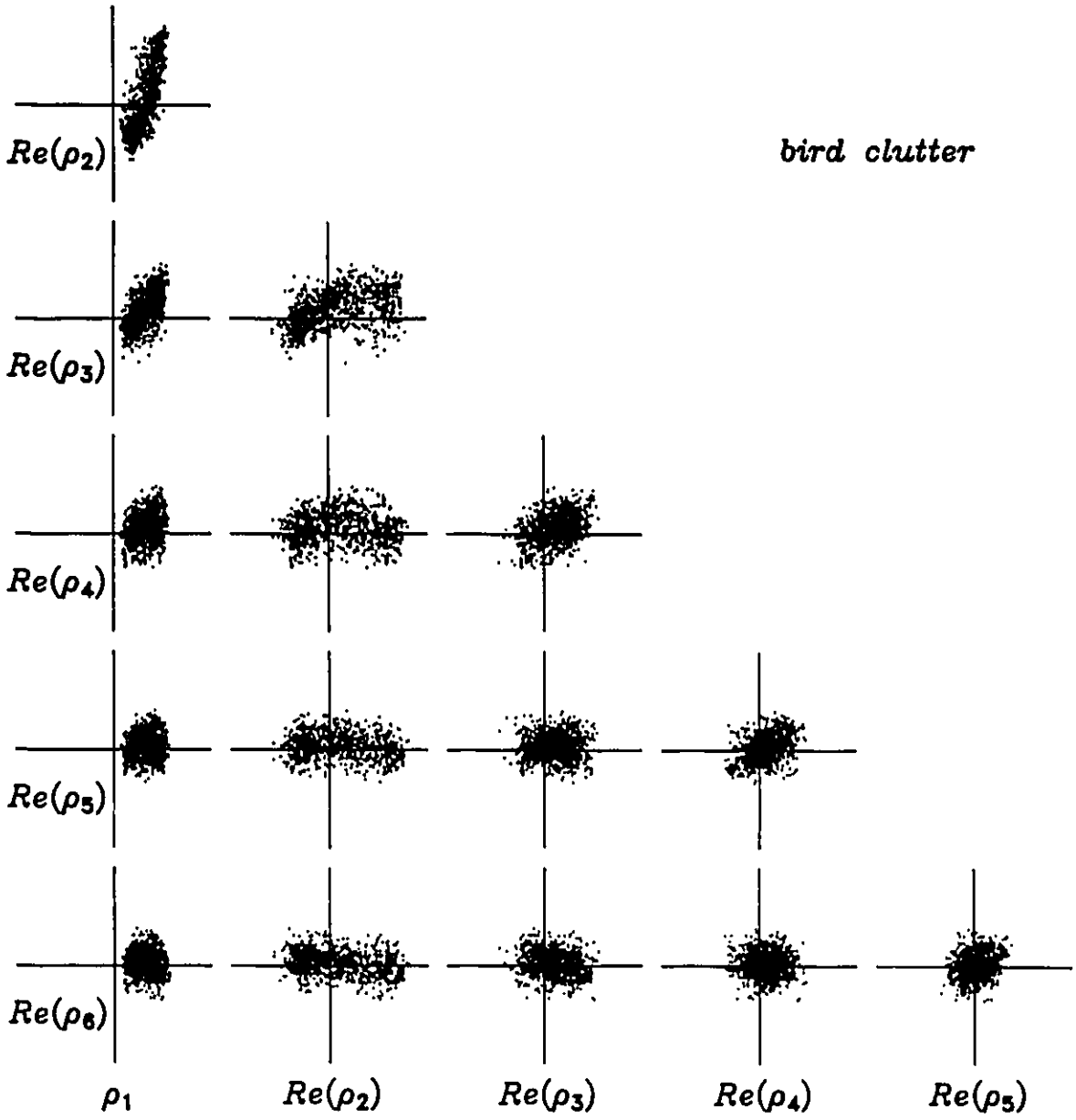


Figure 6.8e Matrix of scatter plots without P_0 for bird clutter

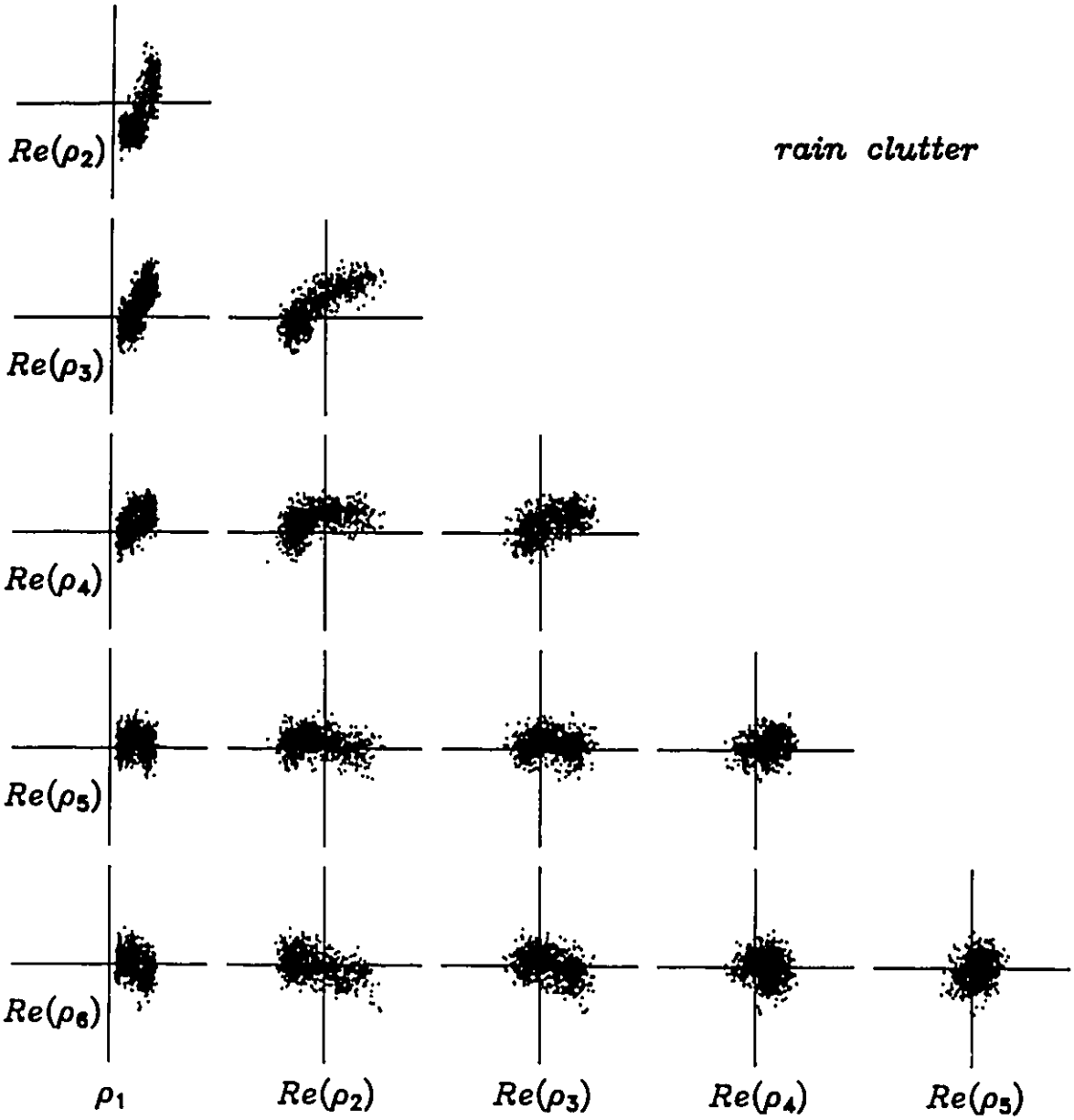


Figure 6.8f Matrix of scatter plots without P_0 for rain clutter

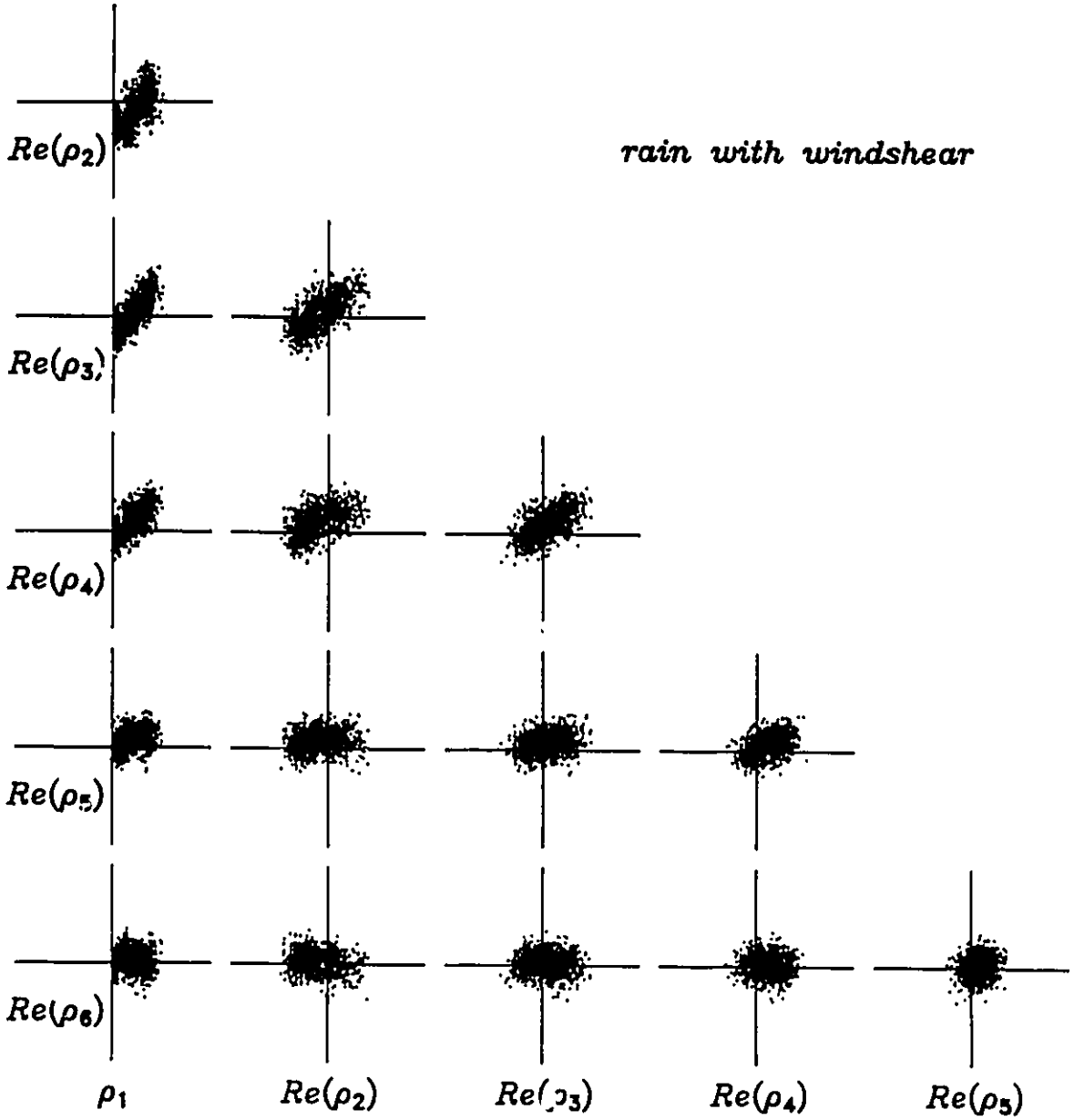


Figure 6.8g Matrix of scatter plots without P_0 for rain clutter with windshear

examined at one time.

In Figure 6.6, the low-order features up to and including ρ_3' are plotted. Some strong dependencies amongst the real parts of the coefficients and the SNR related features are immediately apparent. The imaginary parts seem to play less of a role, even in the case of an asymmetrical spectrum such as rain with a windshear-induced secondary peak (Figure 6.6b). While they clearly contribute invaluable information toward the identification (or rejection) of spectra with two peaks, the dependencies in the real parts are probably the most important discriminants between classes and should be examined more closely.

Figure 6.7 thus shows the matrix including only the real parts of the reflection coefficients up to and including $\Re(\rho_5')$. The nonlinear (hence non-normal) behavior is now very evident and it is almost possible to "see" the cluster of points winding its way through the multidimensional feature space as the signal strength is increased. Both differences and similarities between the classes can now be detected. What is also obvious in these plots is that the fifth-order reflection coefficient contains much more information than was apparent from Figure 6.1. This suggests that high-order information may still be useful in class discrimination, especially in separating point targets (aircraft, cars and trucks, and ground clutter) from the more distributed clutter types (bird and weather clutter). This observation is reinforced in Figure 6.8, where the feature U_0 is dropped and $\Re(\rho_6')$ is added. Clearly, even ρ_6' contains residual information.

The relationship between the signal strength and the coefficient pattern is now well recognized. This fact is emphasized by the plot of spectra from the same clutter classes, but with different signal strengths, in Figure 6.9. It is clear that, as the SNR is increased, more spectral poles move toward the peak to assist in fitting the

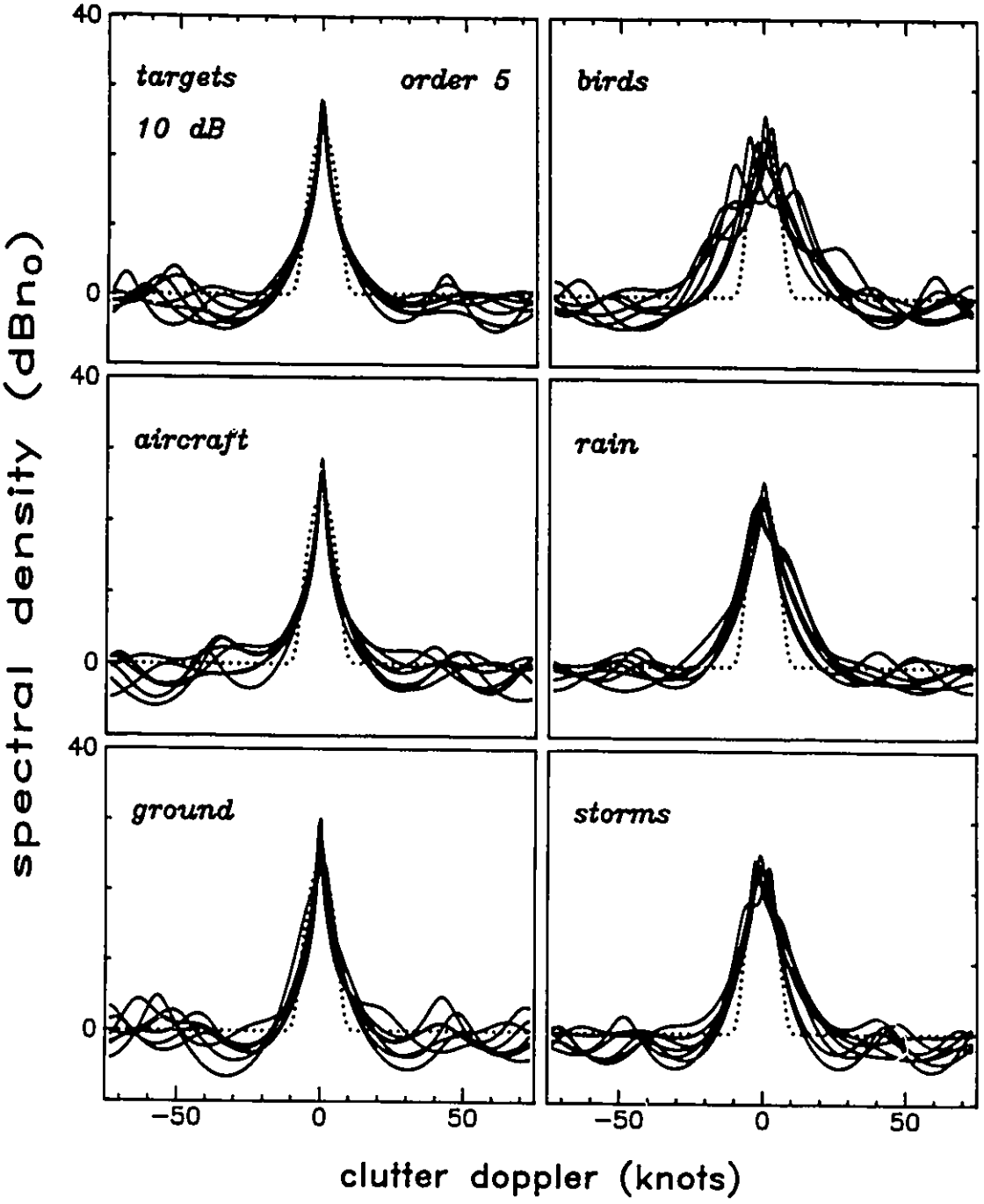


Figure 6.9a Comparative 5th-order clutter spectra for a 10 dB signal strength

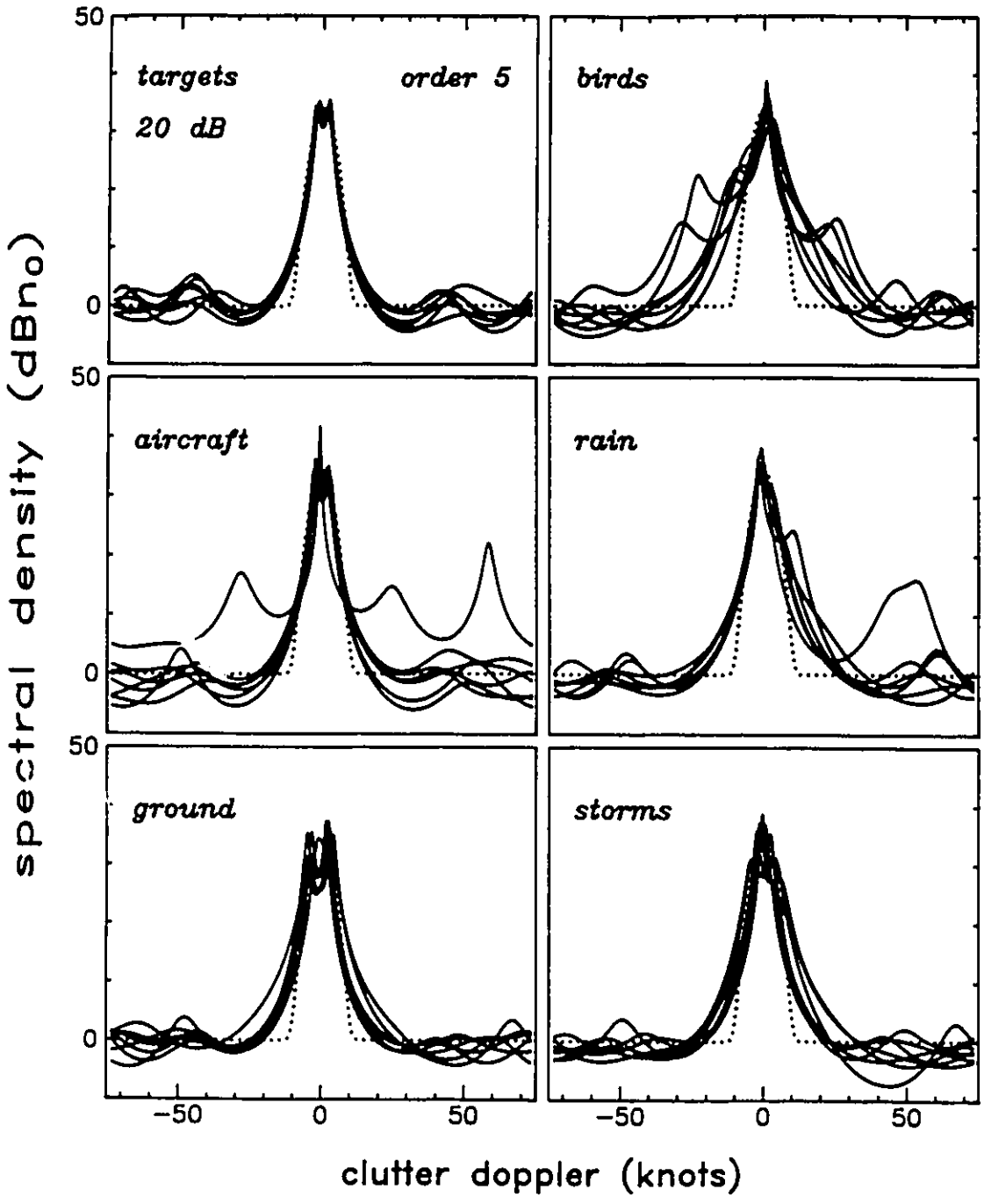


Figure 6.9b Comparative 5th-order clutter spectra for a 20 dB signal strength

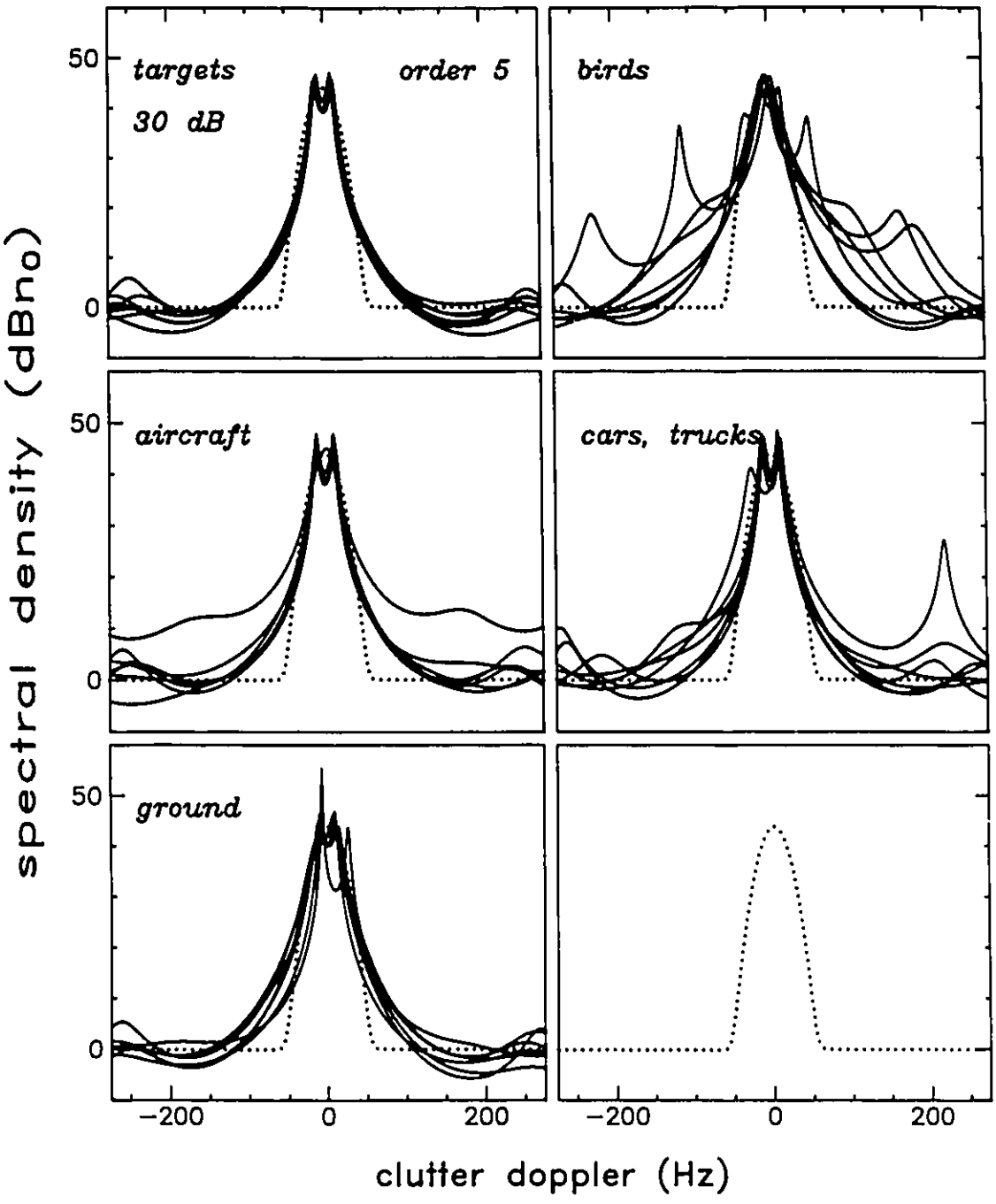


Figure 6.9c Comparative 5th-order clutter spectra for a 30 dB signal strength
No 30 dB data were available for rain and thunderstorm clutter.

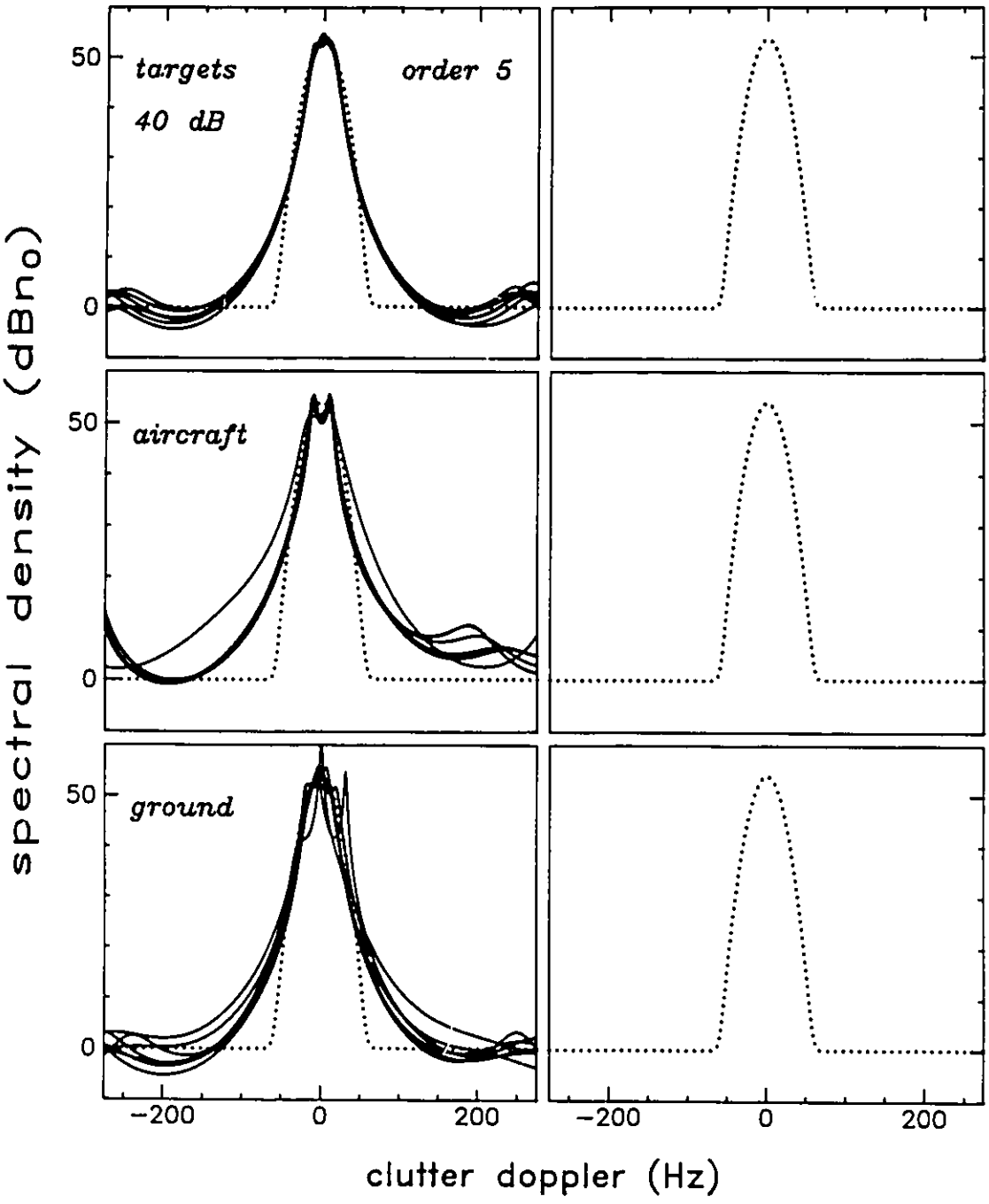


Figure 6.9d Comparative 5th-order clutter spectra for a 40 dB signal strength
No 40 dB data were available for bird and weather clutter.

Gaussian beamshape. The 10 dB spectra appear to be of a very low order with at most two coincident poles forming a well defined peak at a single location. At 20 and 30 dB, most peaks have split into two in an attempt to spread the energy and satisfy the spectral width requirements. The smaller amount of ripple at the 0 dB noise floor indicates the presence of fewer remaining poles. At 40 dB most of the five poles are concentrated at the peak. The absence of frequency and amplitude modulation, other than beamshaping, allows the poles to arrange themselves to almost perfectly fit the shape of the true spectrum for the synthetic target and ground clutter classes. The presence of two peaks in the aircraft spectra indicates that the spectral shape is not exactly Gaussian, which may be expected from such complex targets.

A more quantitative presentation of the relationship of the reflection coefficients with the signal strength is given in Figure 6.10. All sample feature vectors were separated into 1 dB bins according to U_0 and averaged. The averages of the real and imaginary parts of the reflection coefficients were then plotted against the average values for U_0 in each bin. For the reference class of synthetic targets there is no clear filter order for which all information will have been extracted, even at low signal levels. The pure symmetry of the spectra is evident from the absence of information in the imaginary parts: estimation errors due to the low number of samples contribute to the deviation from zero at both the low and high ends of the plots. The same general observations hold for aircraft and ground clutter classes and, presumably, the class for cars and trucks, although there was an insufficient number of samples available to produce such a plot for it.

The difference in the high-order coefficients of the bird and weather clutter classes is now dramatically visible. Bird clutter appears to have little meaningful information in the real part past ρ_3', although the increase in the values of the imaginary

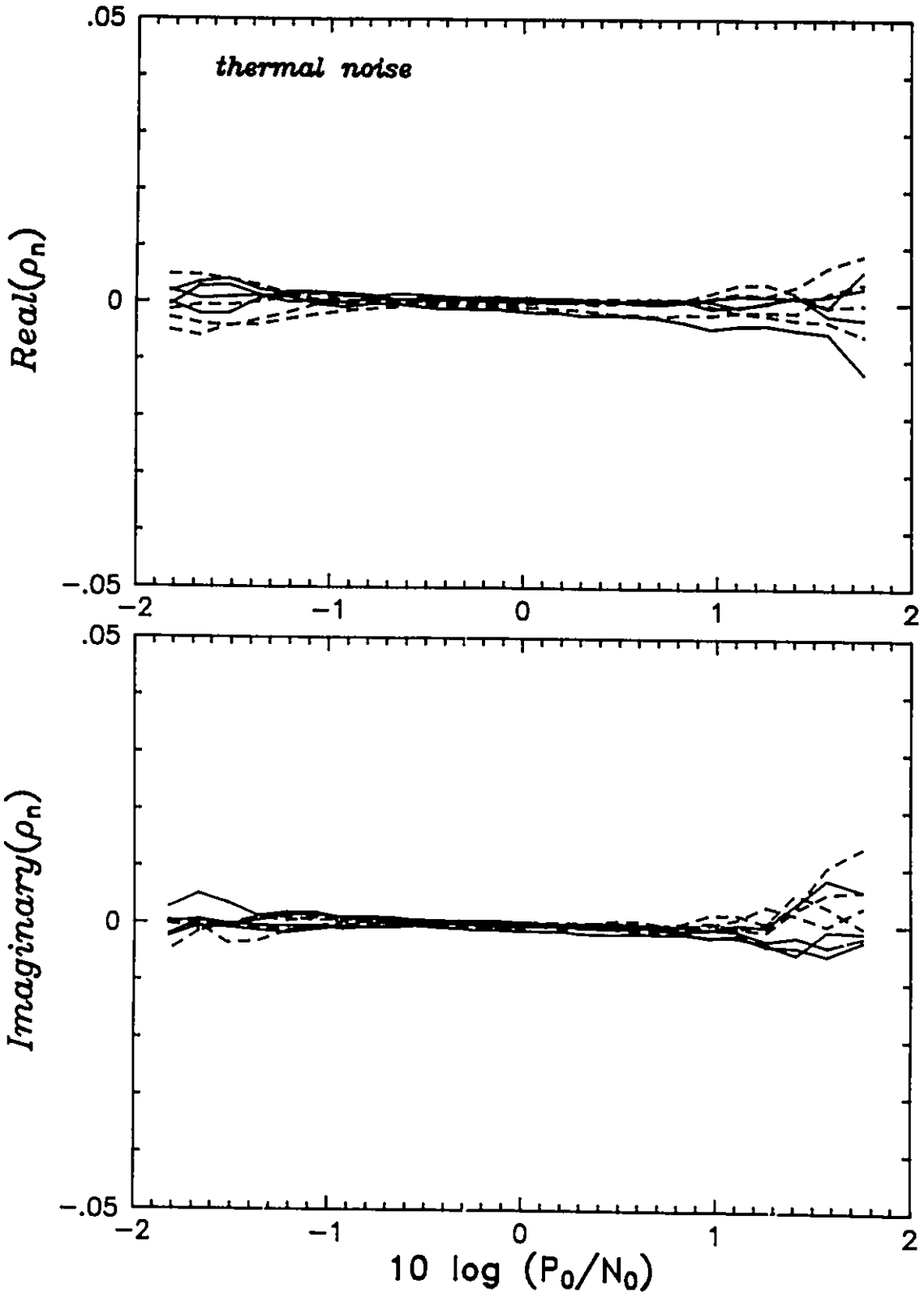


Figure 6.10a Mean feature values for thermal noise

The abscissa represents the feature U_0, with units of dB.

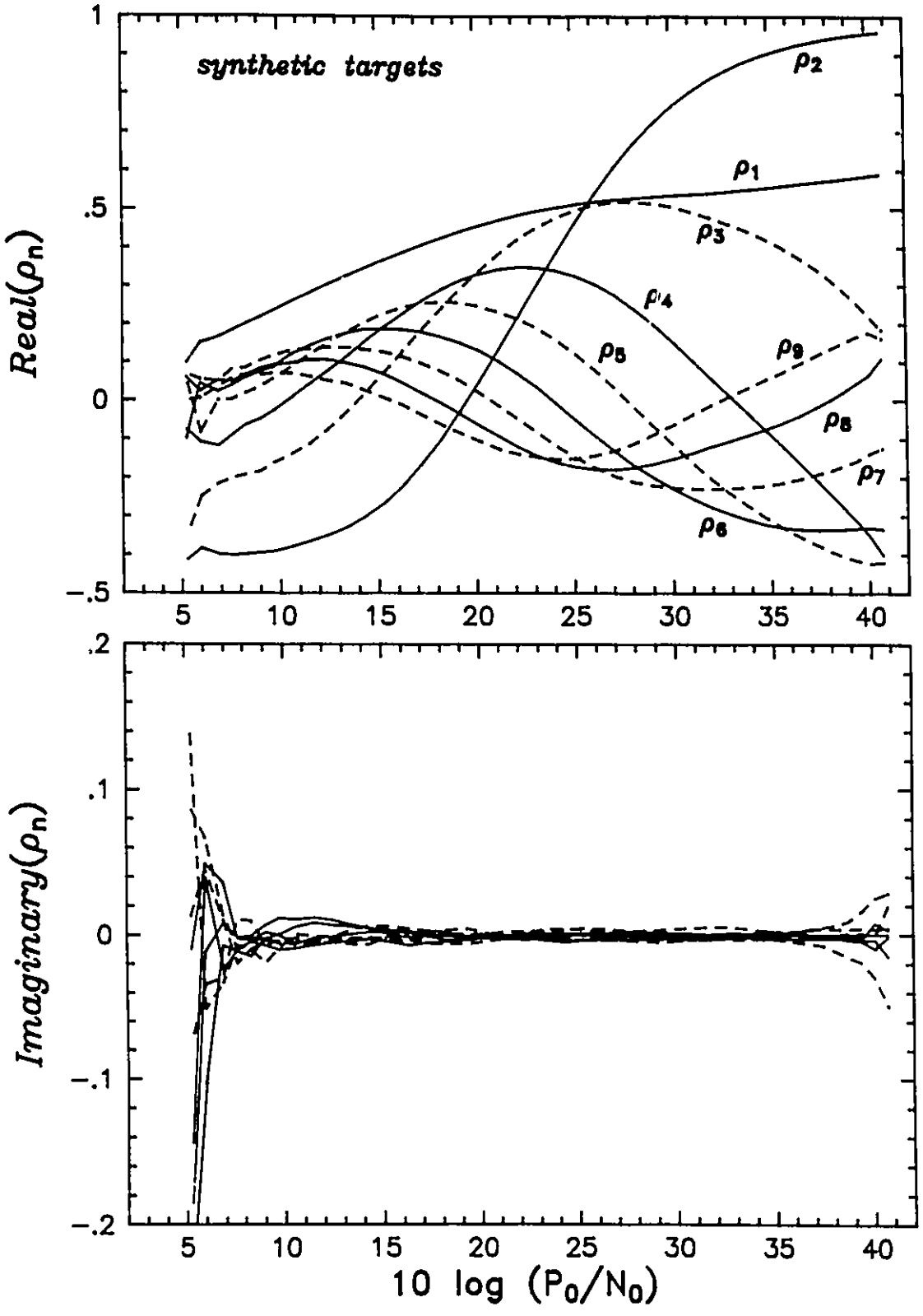


Figure 6.10b Mean feature values for synthetic target

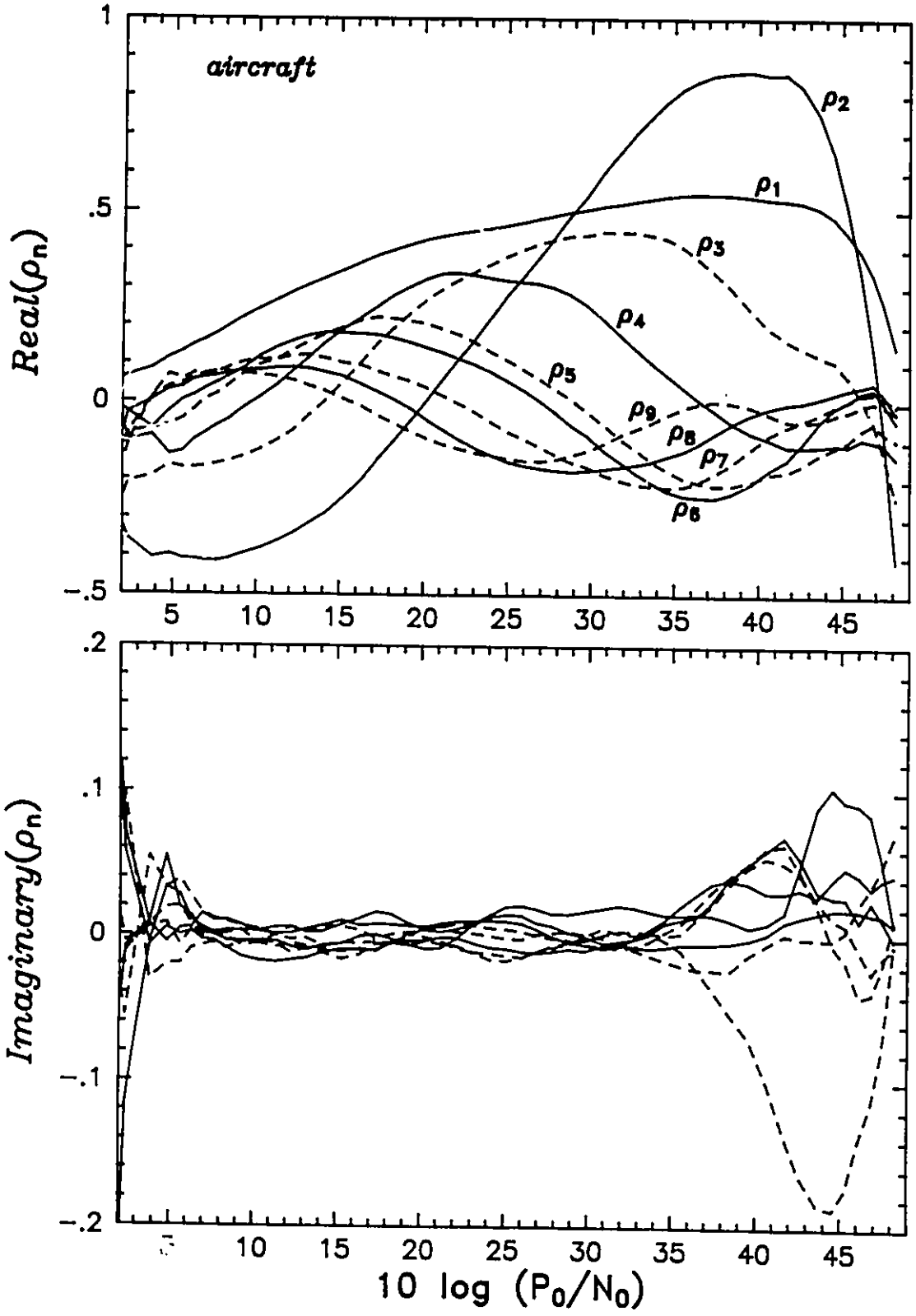


Figure 6.10c Mean feature values for aircraft target

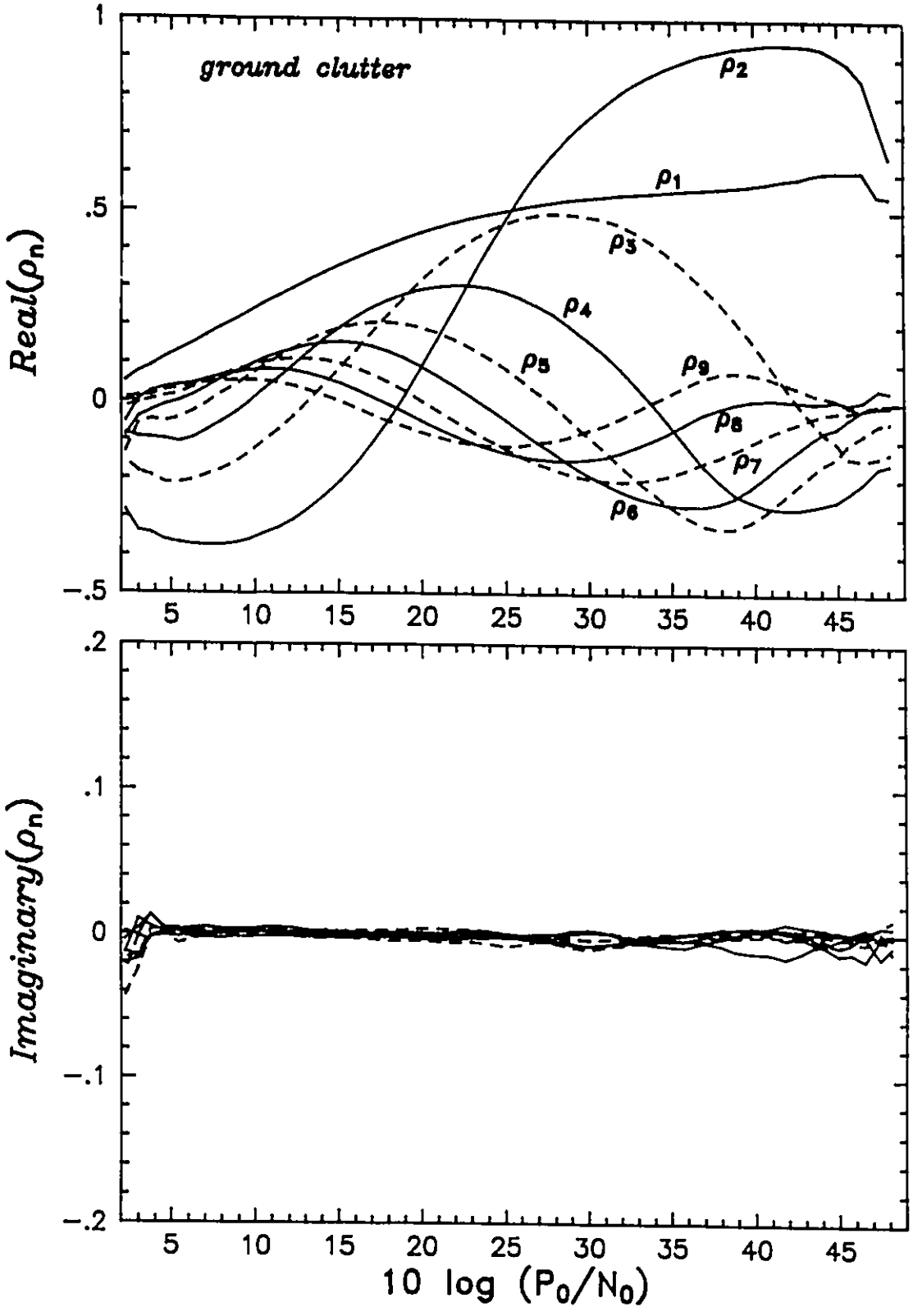


Figure 6.10d Mean feature values for ground clutter

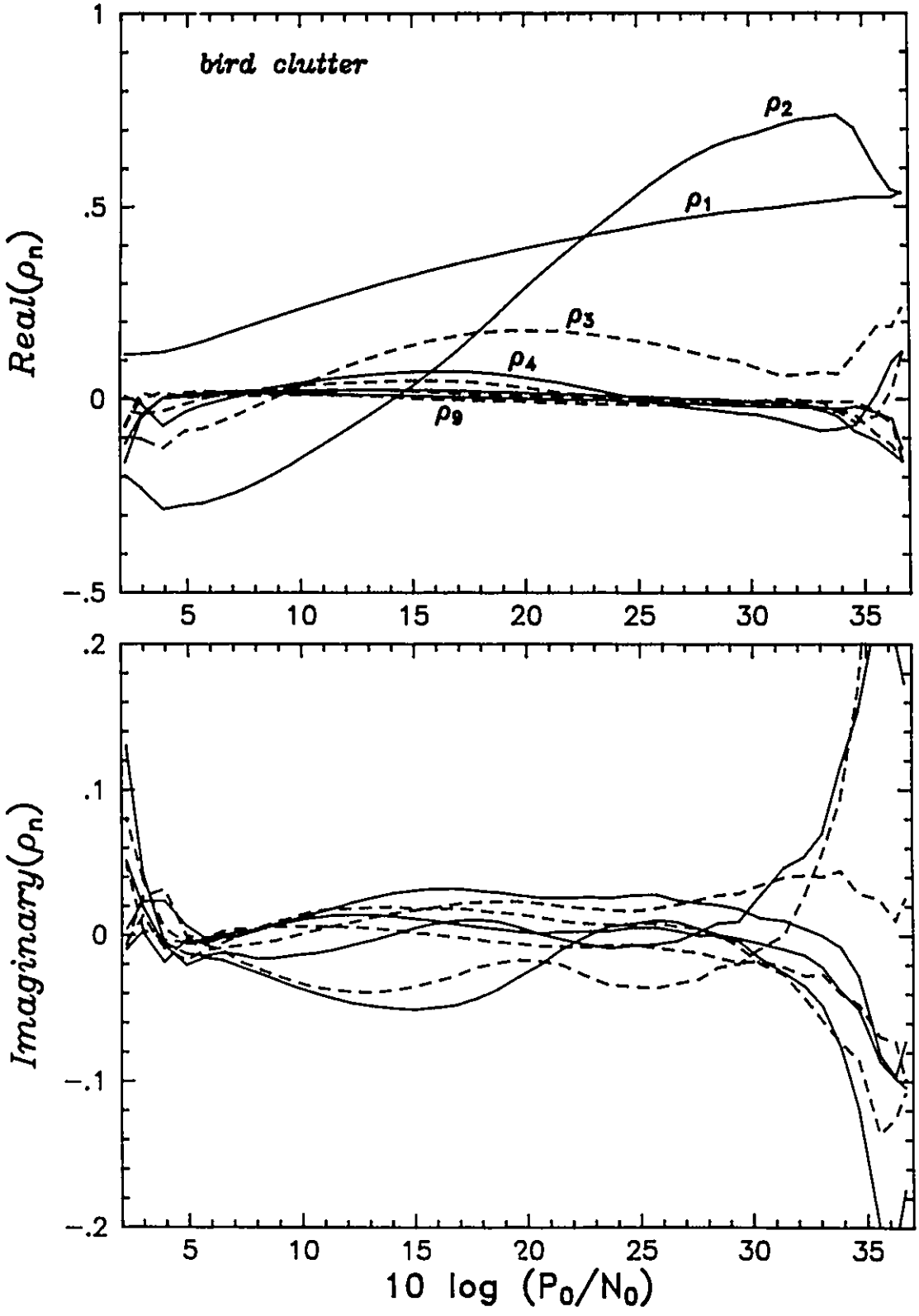


Figure 6.10e Mean feature values for bird clutter

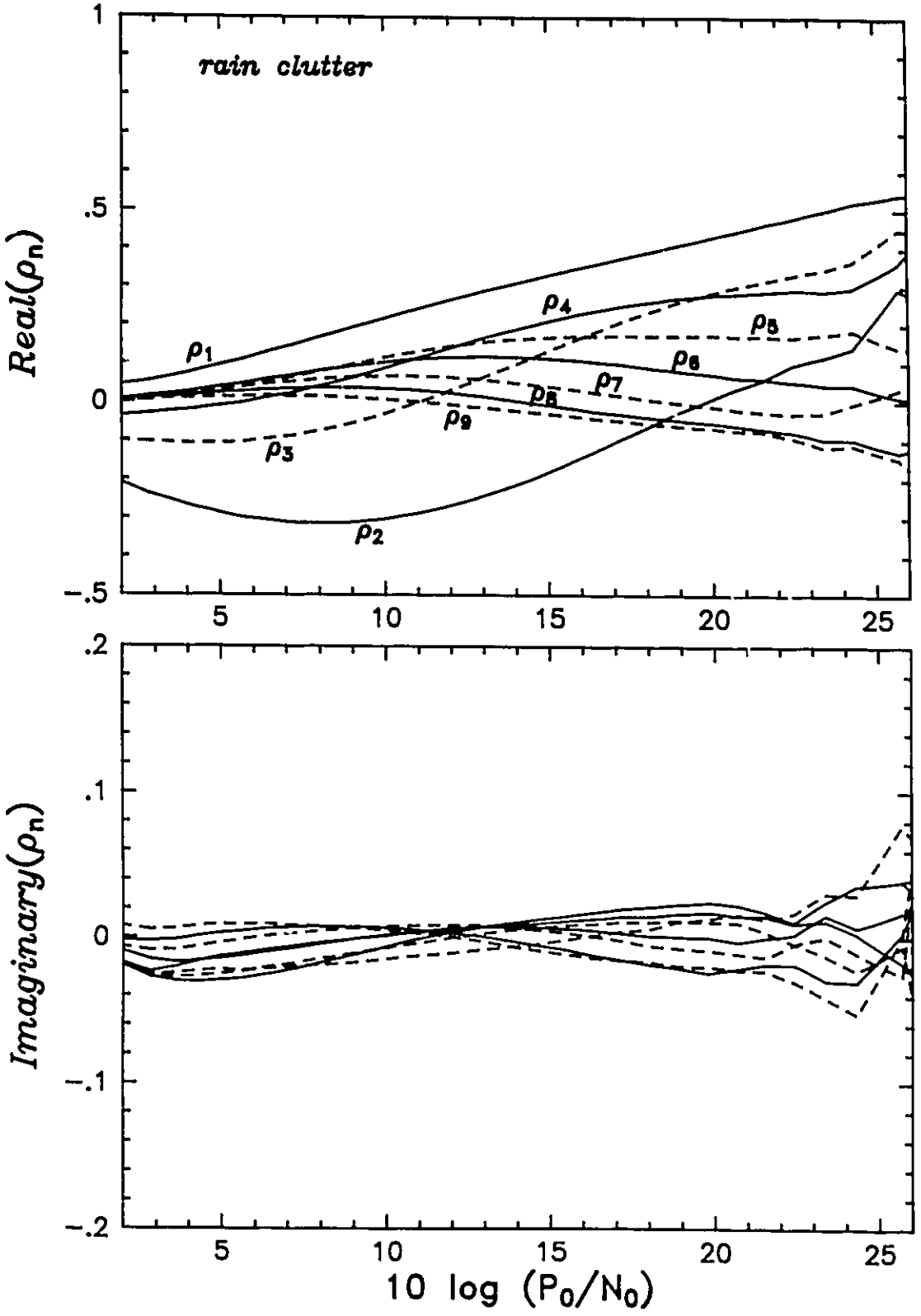


Figure 6.10f Mean feature values for rain clutter

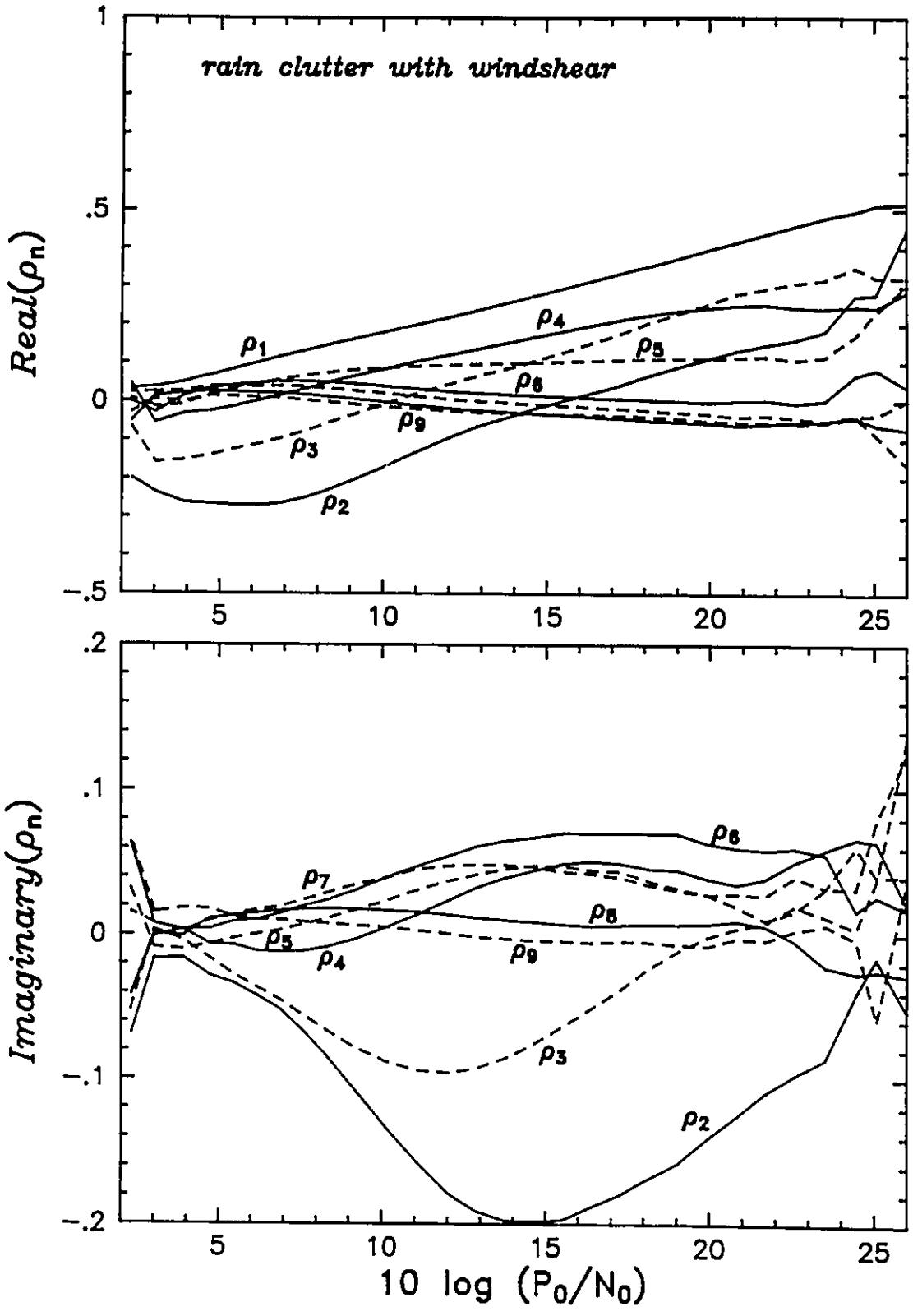


Figure 6.10g Mean feature values for rain clutter with windshear

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.

The dependencies discovered to this point lead to the inescapable conclusion that the features, if taken over the entire range of signal strengths, do not exhibit multivariate normal behavior. Before proceeding to solving this problem, it is instructive to take a look at the probability densities of the resulting squared Mahalanobis distances D_i^2, assuming a class has been defined for the entire labelled feature set. For reference, the probability plot - in fact quantile or Q-Q plots in keeping with the definition offered by Wilk and Gnanadesikan [1968] - of thermal noise, which is likely to behave as near to multivariate normal as can be expected, is shown in Figure 6.11a. The Mahalanobis distance is computed from a 13-dimensional feature set, which includes ϕ, U_0, u_0, ρ_2', ρ_3', ρ_4', ρ_5', P_{dif}, and P_{var}. The actual frequency ϕ has been replaced with a uniformly distributed random value between the normalized Doppler frequency limits of ± 0.5. The ordered squared Mahalanobis distances are plotted against the χ_{13}^2 variates expected given the same cumulative probability. If the D_i^2 are in fact χ_{13}^2 distributed, then a straight line should result. (The actual distribution is in fact $F_{d,N}\,[d(N+1)(N-1)]/[N(N-d)]$, where d is the length of the feature vector and N is the number of vectors used for the prototype estimate [Young, 1978]. This distribution converges toward χ_d^2 in the limit as $N \longrightarrow \infty$. For large values of N, the difference may be ignored; see Section 4.4.) This line is shown dotted, and the actual

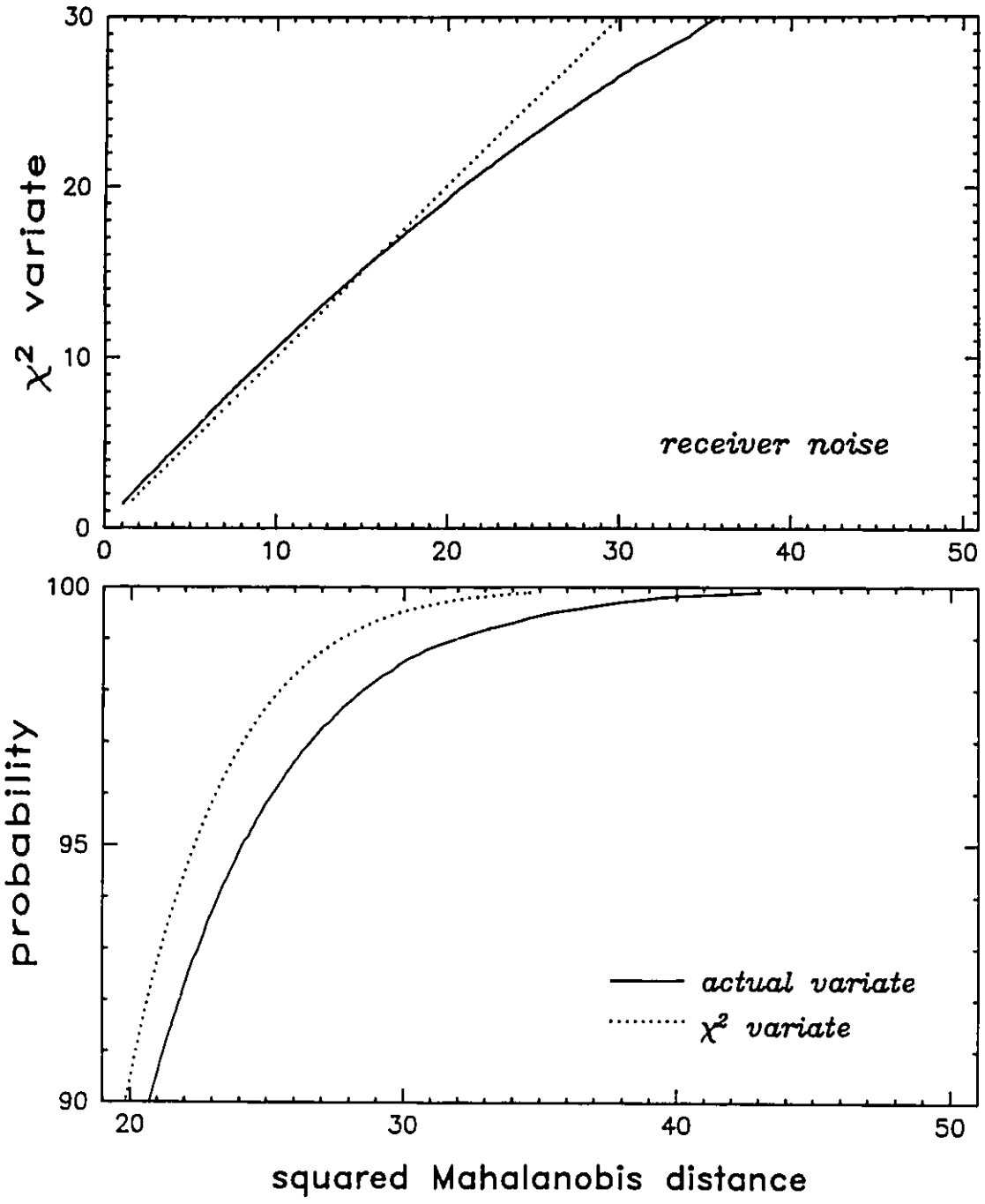


Figure 6.11a Quantile probability plots for D^2 of thermal noise

plot is seen to deviate only slightly until D_i^2 becomes large. The deviation in the tail of the distribution is shown more clearly in the lower plot, where the actual cumulative distribution is compared with that of the χ^2_{13} variates. For thermal noise, the deviation is mostly due to the uniform distribution of the feature ϕ and is not major; hence the approximate χ^2_{13} distribution can be considered confirmed. For the point-target classes (Figures 6.11b to 6.11e), however, it is major and not only in the tail. The deviation of the lower values indicates a closer clustering of the feature set to its mean values than would be expected for a multivariate normal variate. This is, in principle, helpful to the classifier in that separability will be enhanced for that percentage of feature samples. On the other hand, a larger percentage is also found in the tail of the distribution, increasing the expected probability of error given the same class separation, particularly if that probability is expected to be less than 10%. The strength of the tail also indicates that increasing separability through variance reduction will not lower the probability of error as rapidly as indicated by the performance predictors discussed in Chapter 4. Using rejection techniques to control the error rates under these conditions may lead to an unacceptably large number of rejections. As seen in Figures 6.11g and 6.11h, this problem is much less acute in the distributed clutter classes including rain and thunderstorm. These all appear to follow the χ^2 distribution very well.

6.2.3 The Class Prototypes

The question to be addressed is, then: how can class prototypes be defined in order to mitigate the above mentioned problems? Short of finding some nonlinear transformation which entirely removes the dependence of the reflection coefficients on the signal strength, comparisons amongst classes must proceed only for equal values of

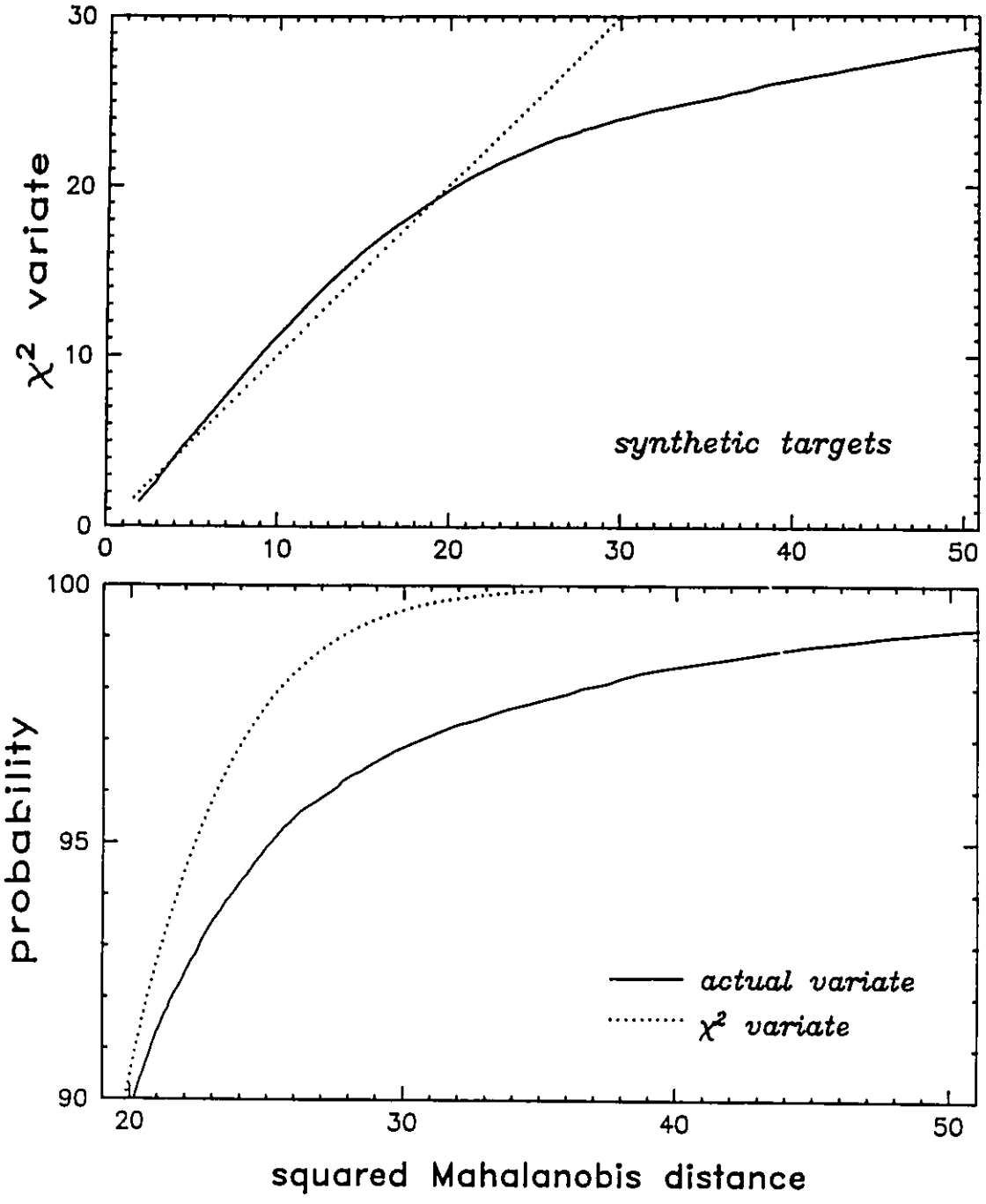


Figure 6.11b Quantile probability plots for D^2 of synthetic targets

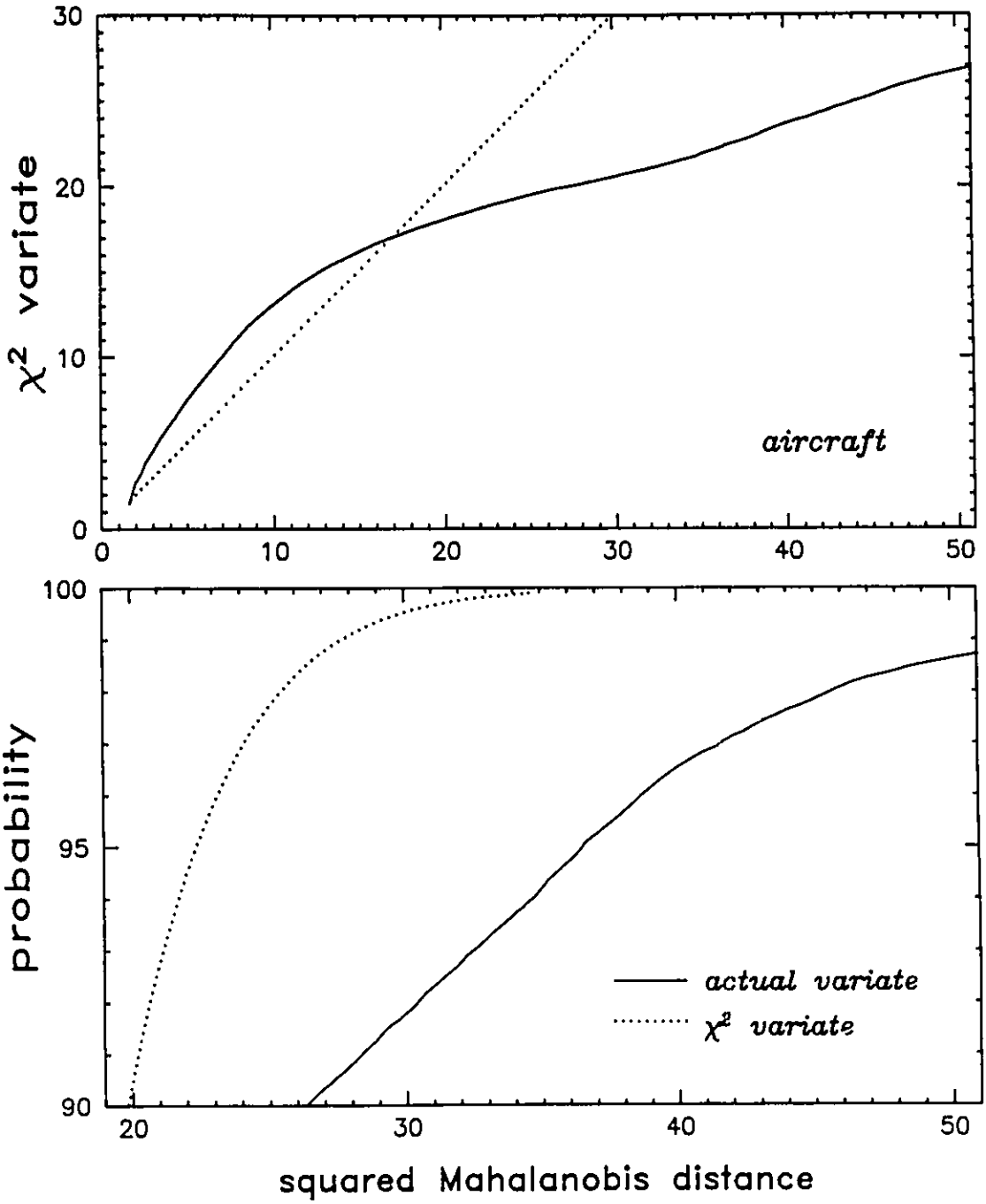


Figure 6.11c Quantile probability plots for D^2 of aircraft targets

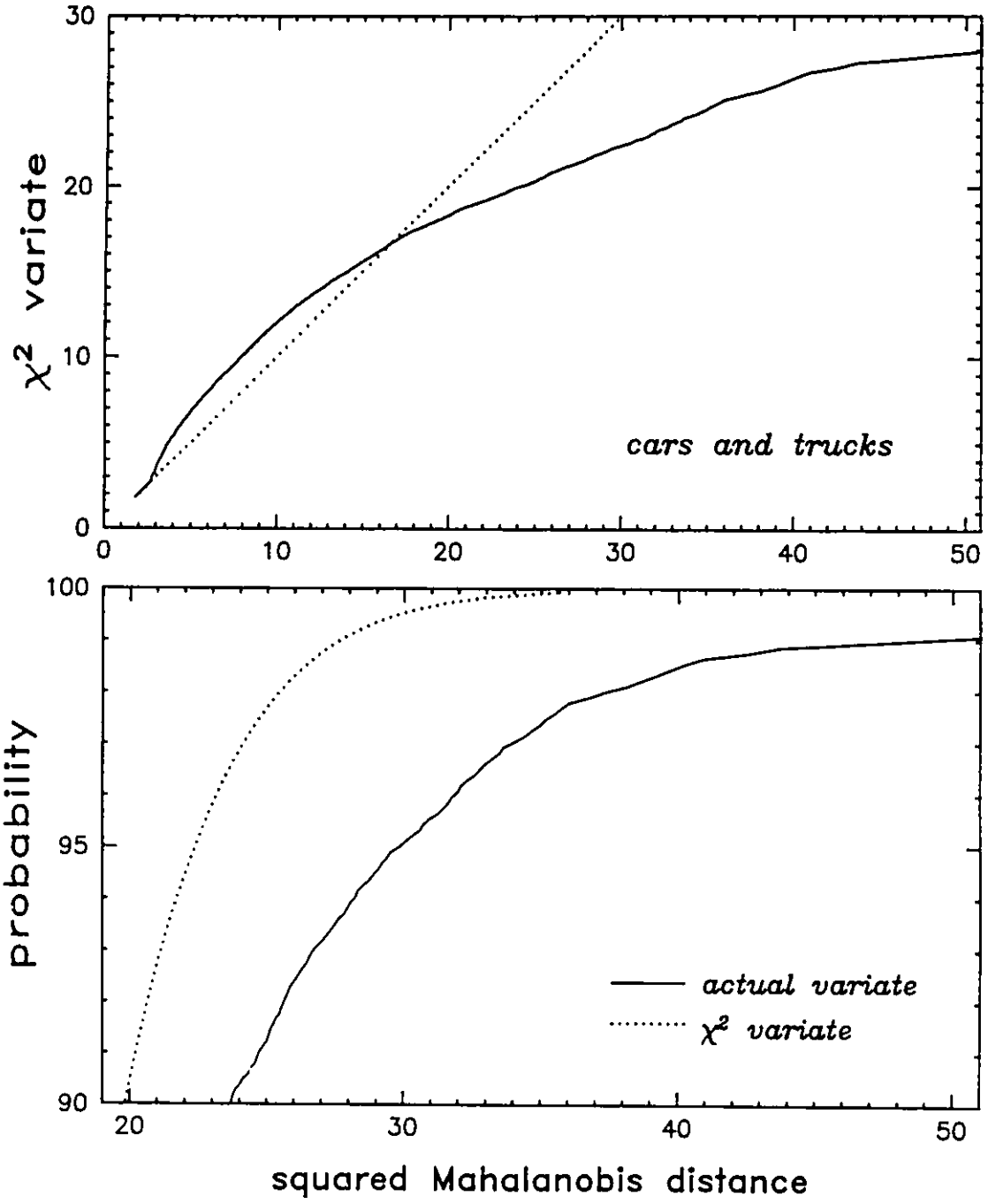


Figure 6.11d Quantile probability plots for D^2 of cars and trucks

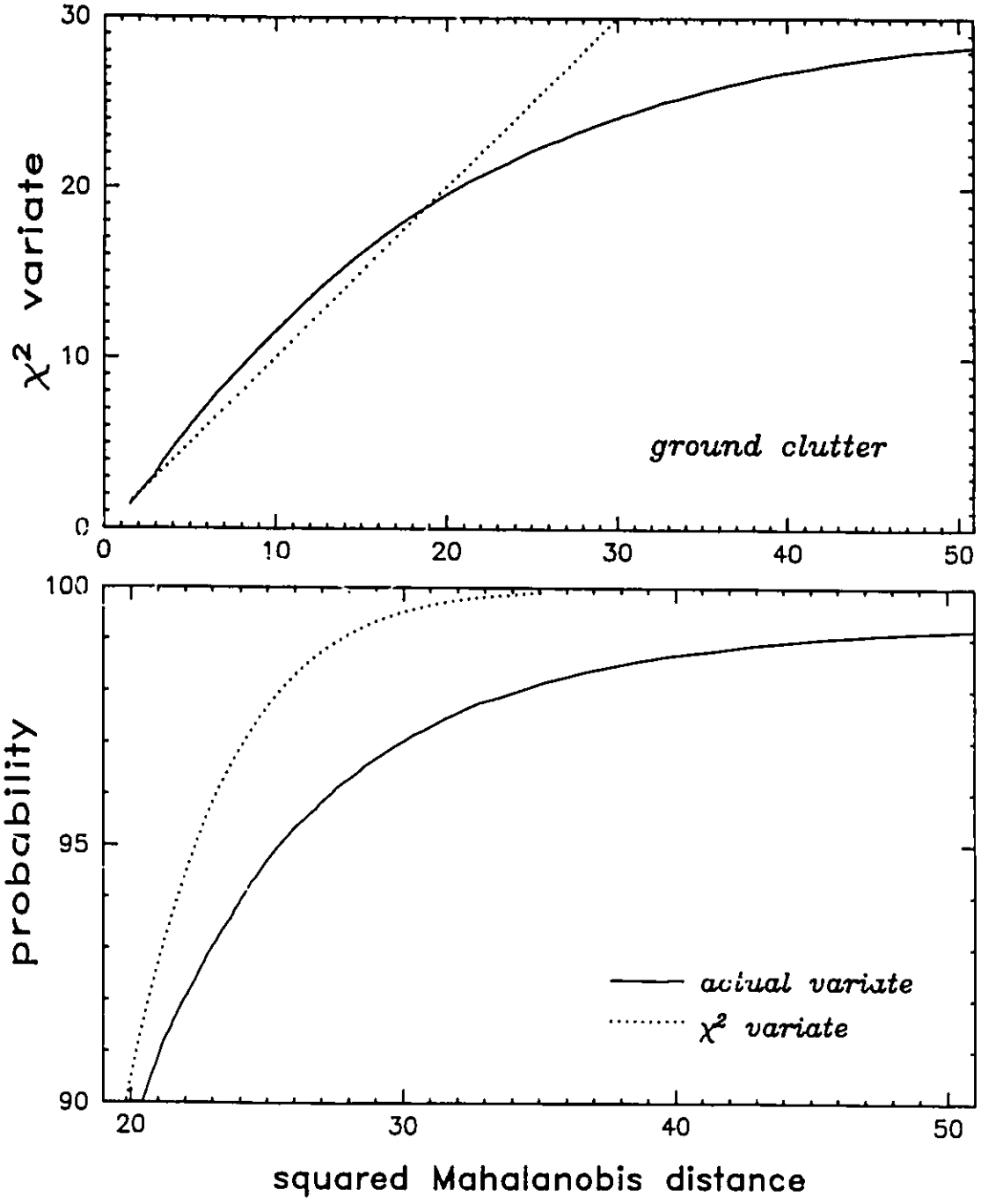


Figure 6.11e Quantile probability plots for D^2 of ground clutter

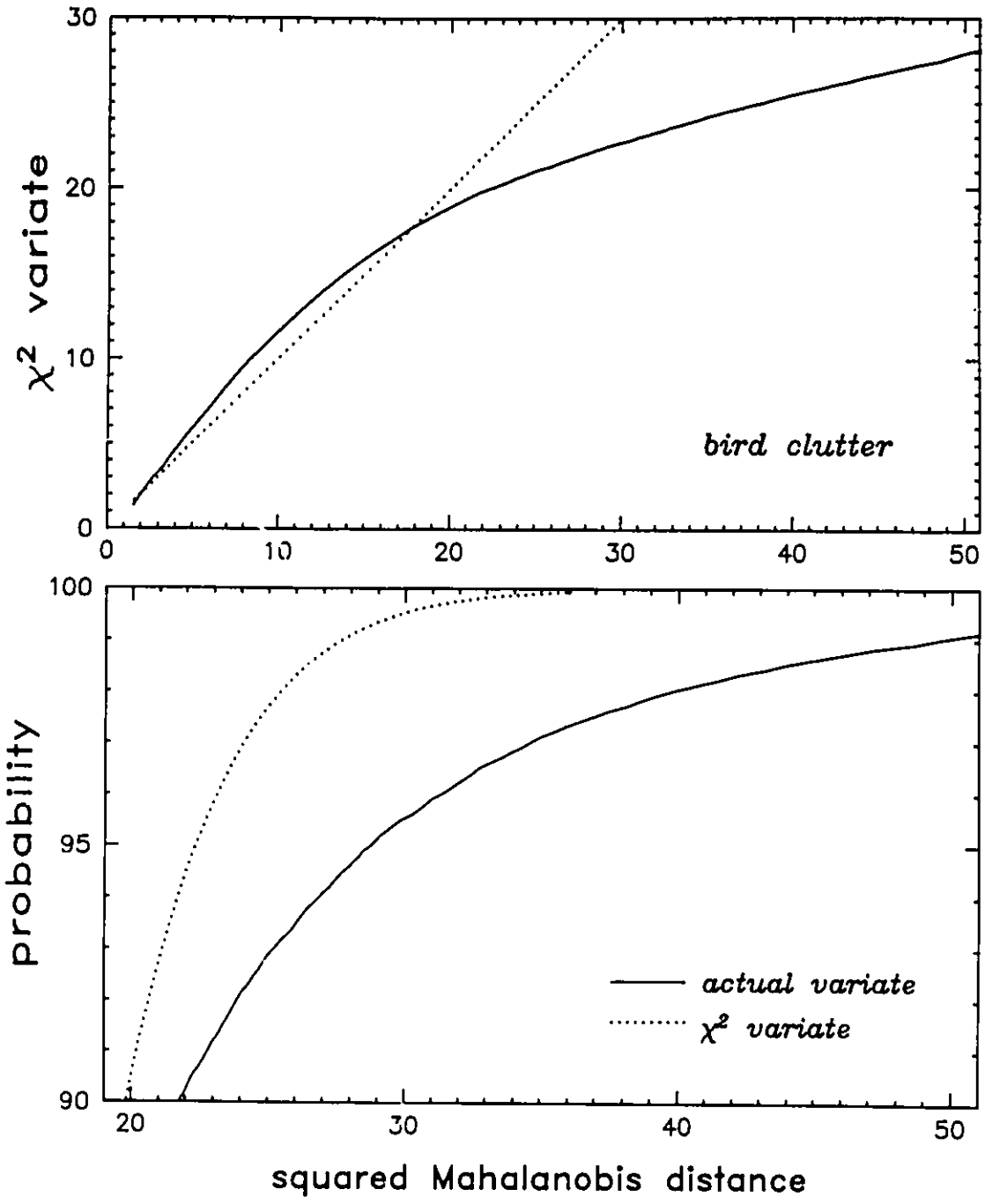


Figure 6.11f Quantile probability plots for D^2 of bird clutter

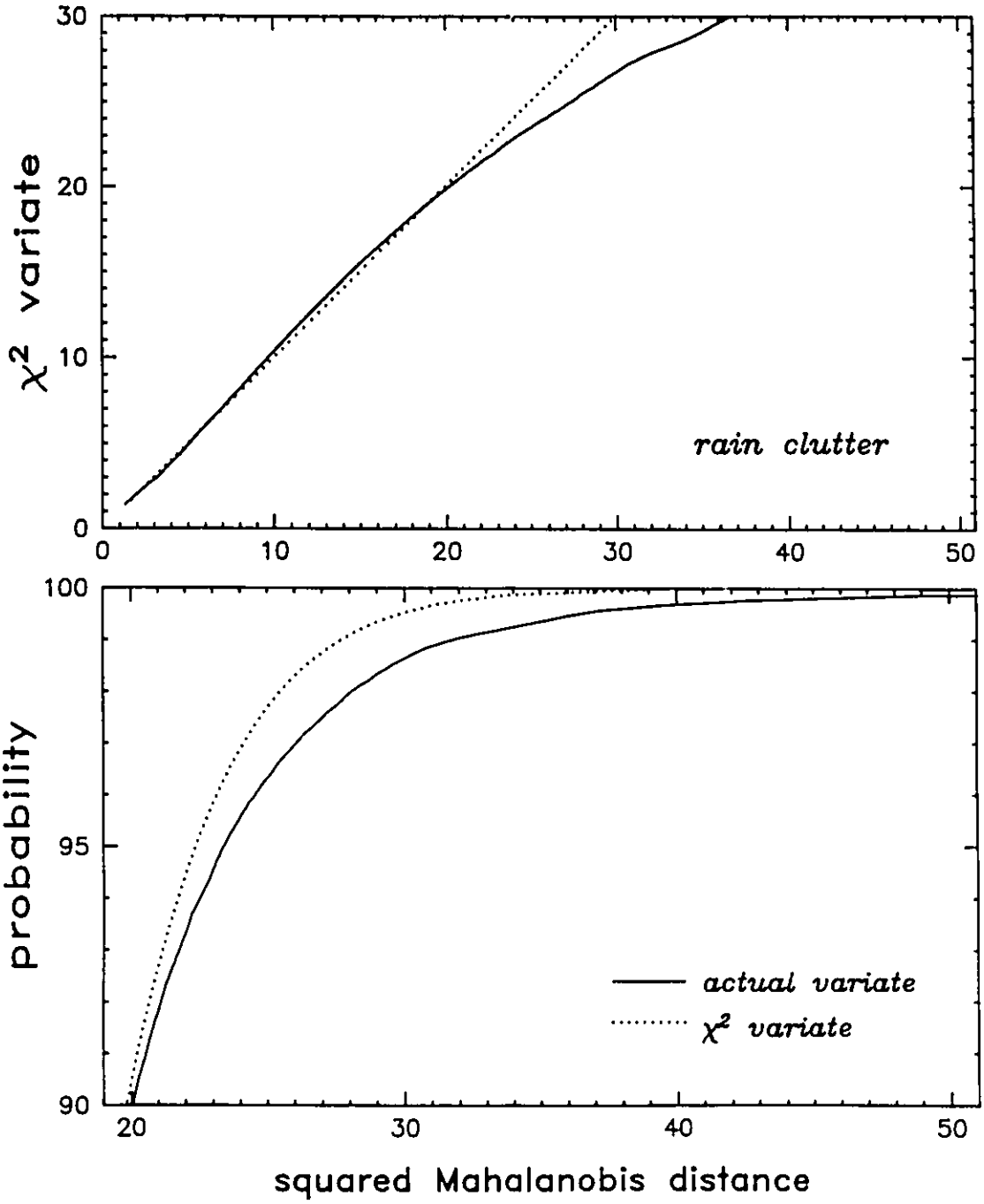


Figure 6.11g Quantile probability plots for D^2 of rain clutter

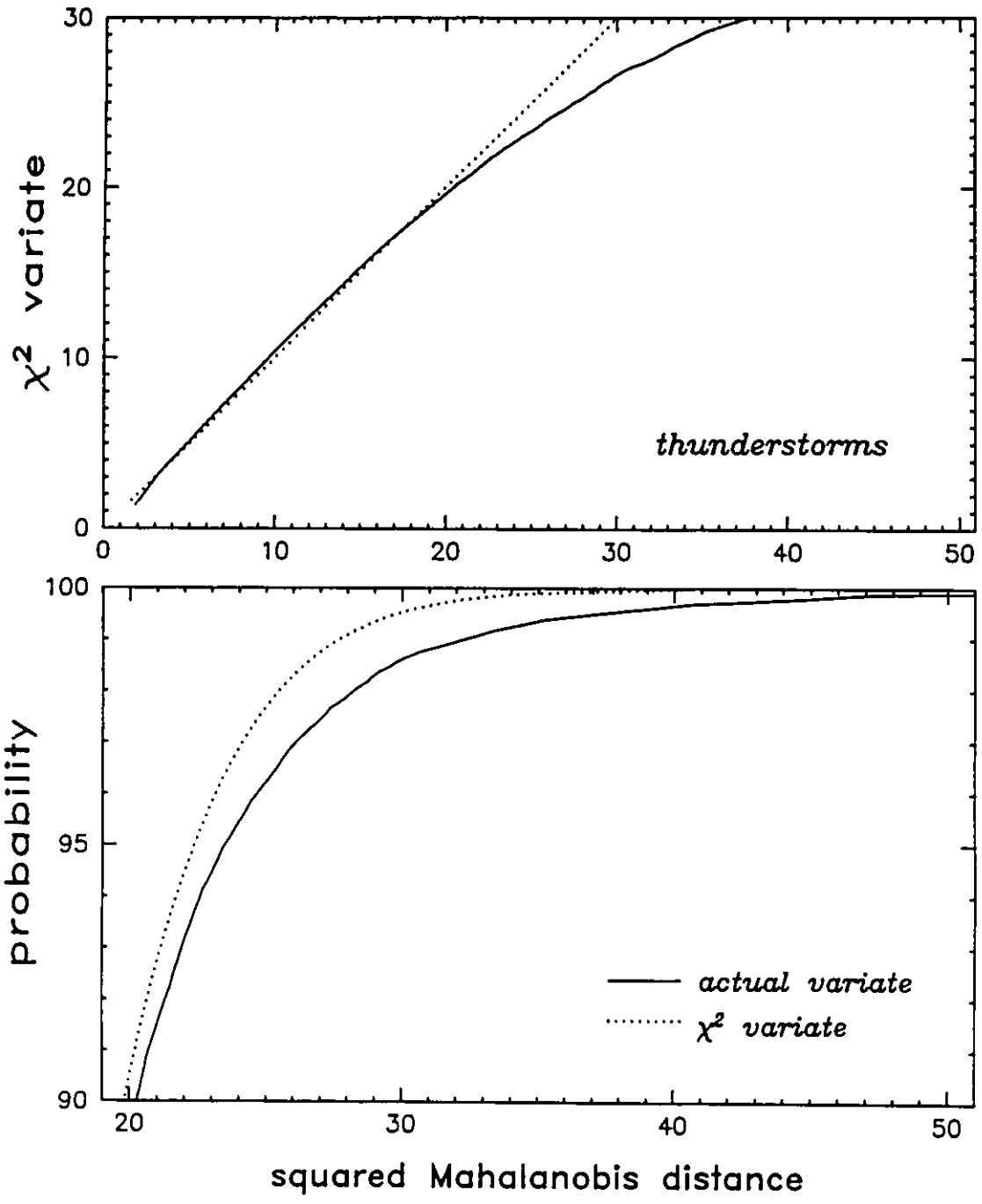


Figure 6.11h Quantile probability plots for D^2 of thunderstorms

U_0. That, of course, requires a prototype definition for every possible value of U_0, 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 since nature rarely provides logical breakpoints for such efforts. This case is no exception, and Figure 6.10 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 thesis work are therefore not likely those which 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:

A	B	C	D	E	F
min	4 dB	10 dB	17 dB	24 dB	32 dB	max

As is apparent from Figure 6.4, 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.

Figure 6.12 shows the scatter-plot matrices for the synthetic target classes. Little nonlinear behavior is evident at low signal strengths, and residual curvatures appear only for ranges E and F. The curvatures are, however, very slight and should be manageable. Since little distributed clutter will be found in these ranges, discrimination will only be required between the various point target classes which is desirable, but less important for the purpose of clutter classification. Certainly, reducing the width of these ranges may improve the classifier performance, but considerably more data than was available here would then be necessary for reliable prototype estimation.

The validity of this partitioning was also tested on aircraft and ground clutter, which demonstrated the most nonlinear behavior according to Figure 6.10. Both of them, aircraft in Figure 6.13 and ground clutter in Figure 6.14, show slight residual curvatures for ranges E and F, and virtually none for range B, C, and D. No scatter plots were generated for the other classes since, as is already evident from Figure 6.10, there will be even less residual nonlinear behavior for them.

Probability plots of the squared Mahalanobis distance for the class definitions are shown in Figure 6.15. The deviation from the χ^2 distribution is much less than that of Figure 6.11, especially for ranges A, B, and C (U_0 less than 17 dB). Residual nonlinearities present in the high-SNR classes suggest that the width of the SNR ranges could be reduced further, although some of them are undoubtedly due to the inclusion of the now mostly uniformly distributed feature U_0. The characteristics of the rain-with-windshear class are notable for their incredibly close adherence to the χ^2 distribution, especially in range C. It is unlikely, however, that this is due to exact multivariate normal behavior, since at least two features are uniformly distributed. It

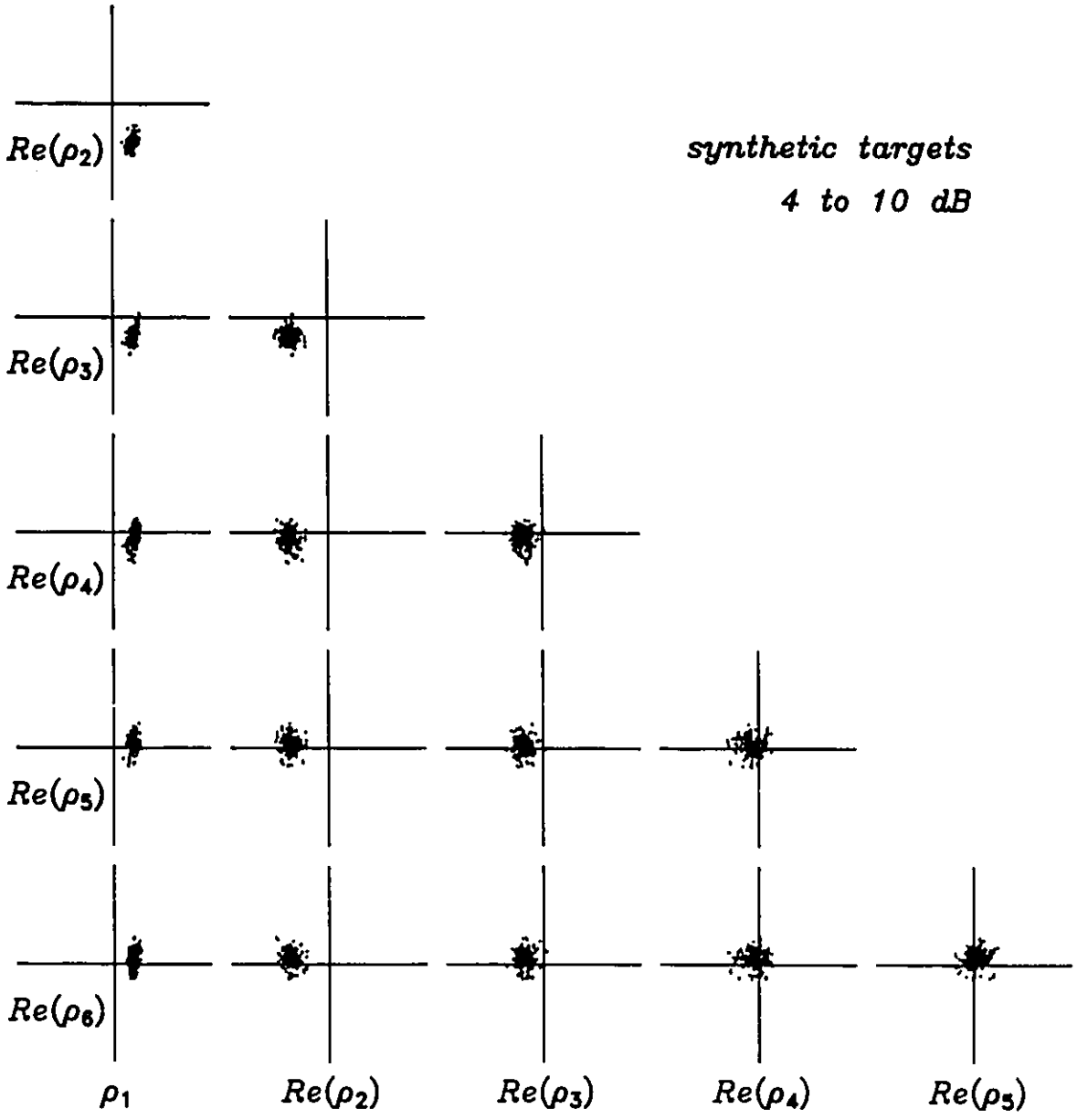


Figure 6.12a Matrix of scatter plots for synthetic targets in SNR range B

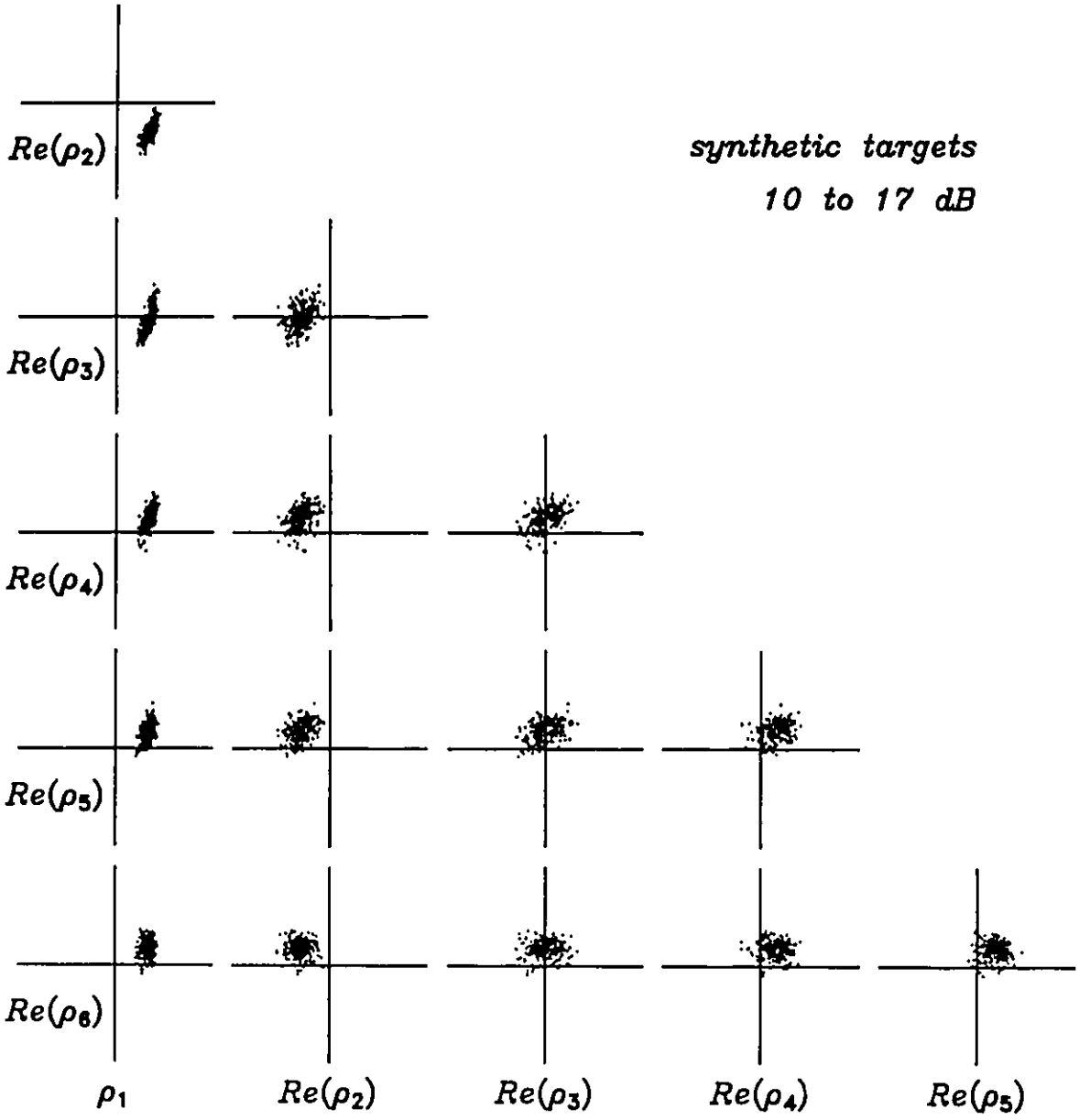


Figure 6.12b Matrix of scatter plots for synthetic targets in SNR range C

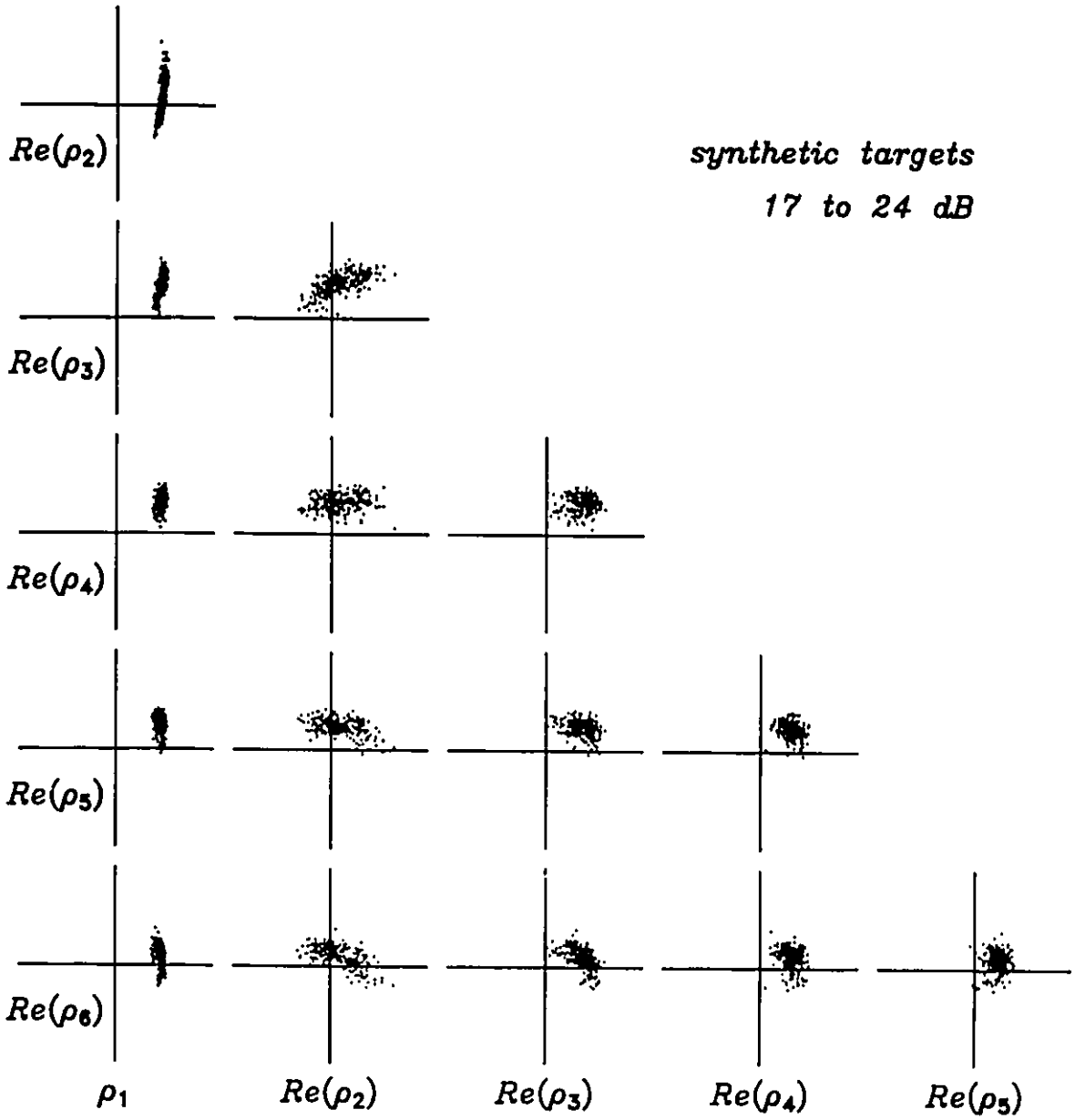


Figure 6.12c Matrix of scatter plots for synthetic targets in the SNR range D

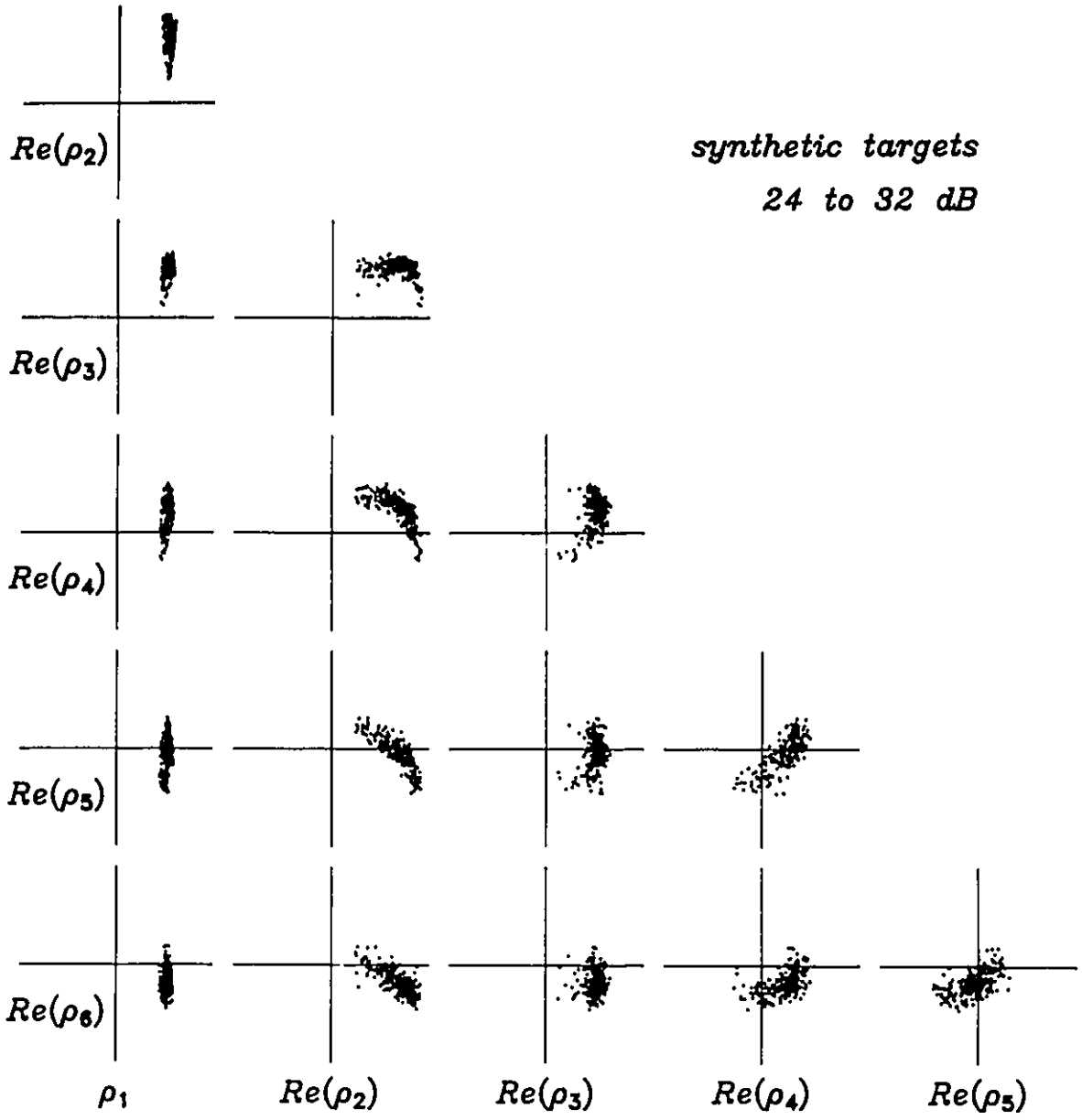


Figure 6.12d Matrix of scatter plots for synthetic targets in the SNR range E

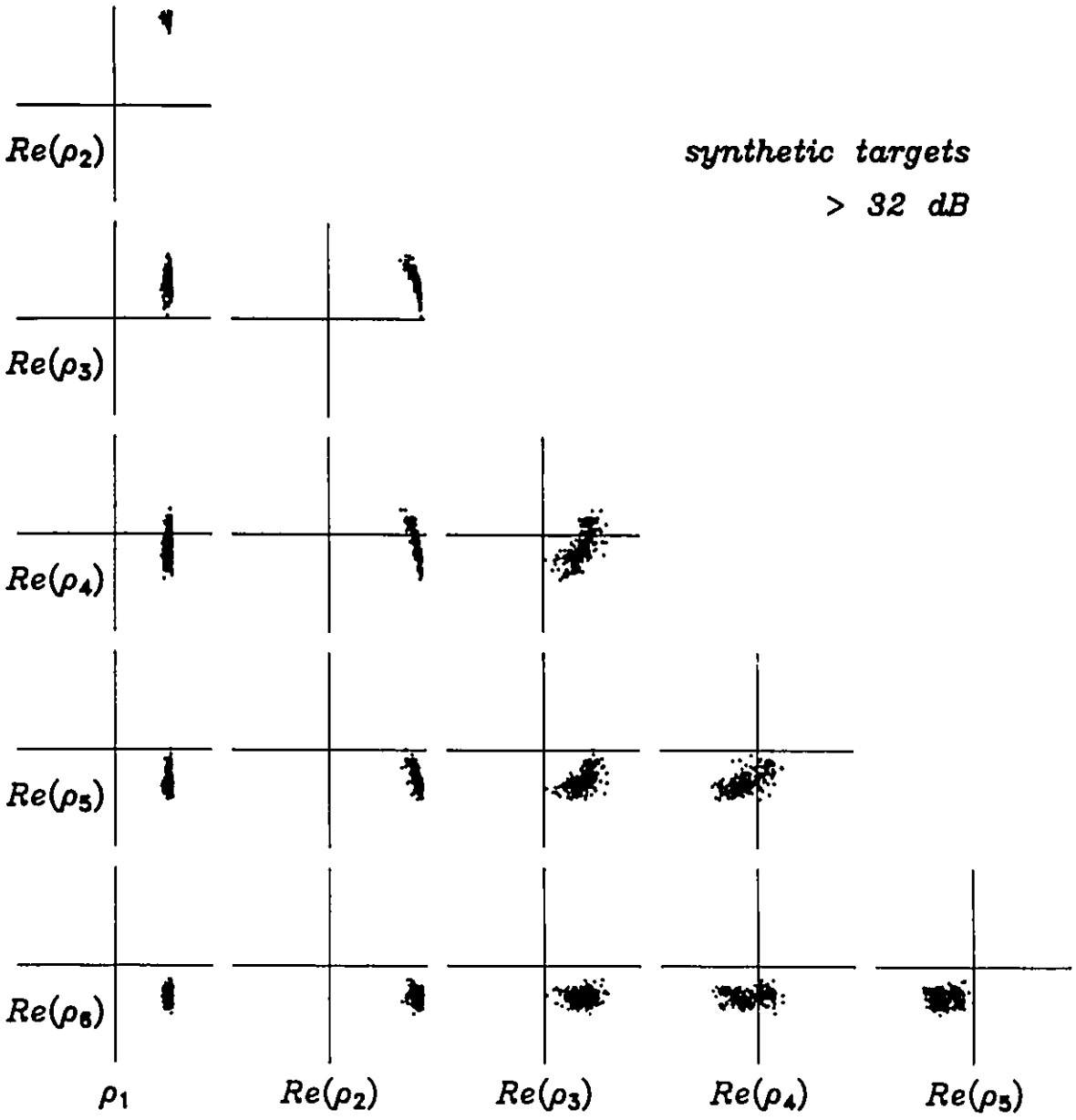


Figure 6.12e Matrix of scatter plots for synthetic targets in the SNR range F

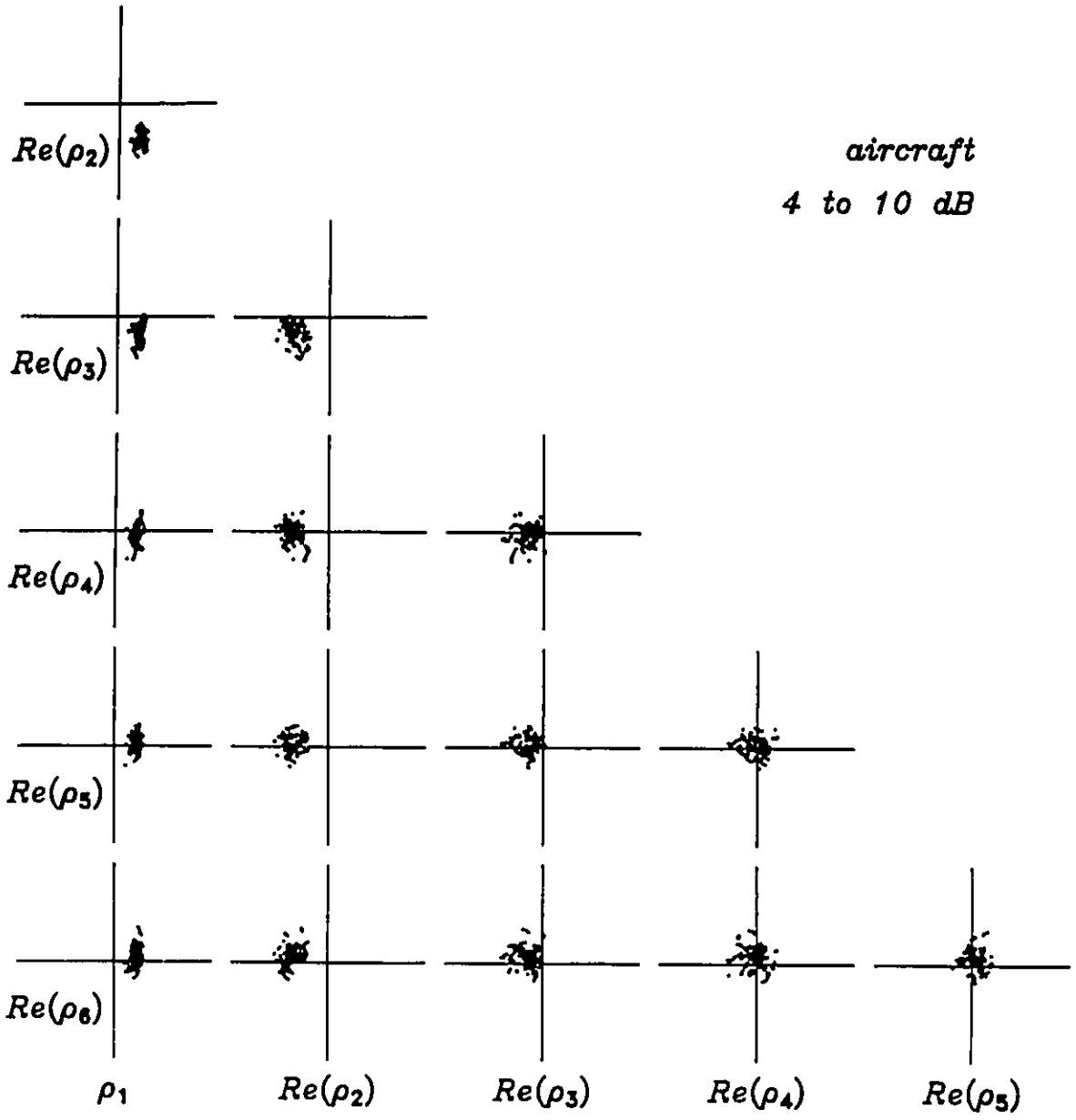


Figure 6.13a Matrix of scatter plots for aircraft targets in SNR range B

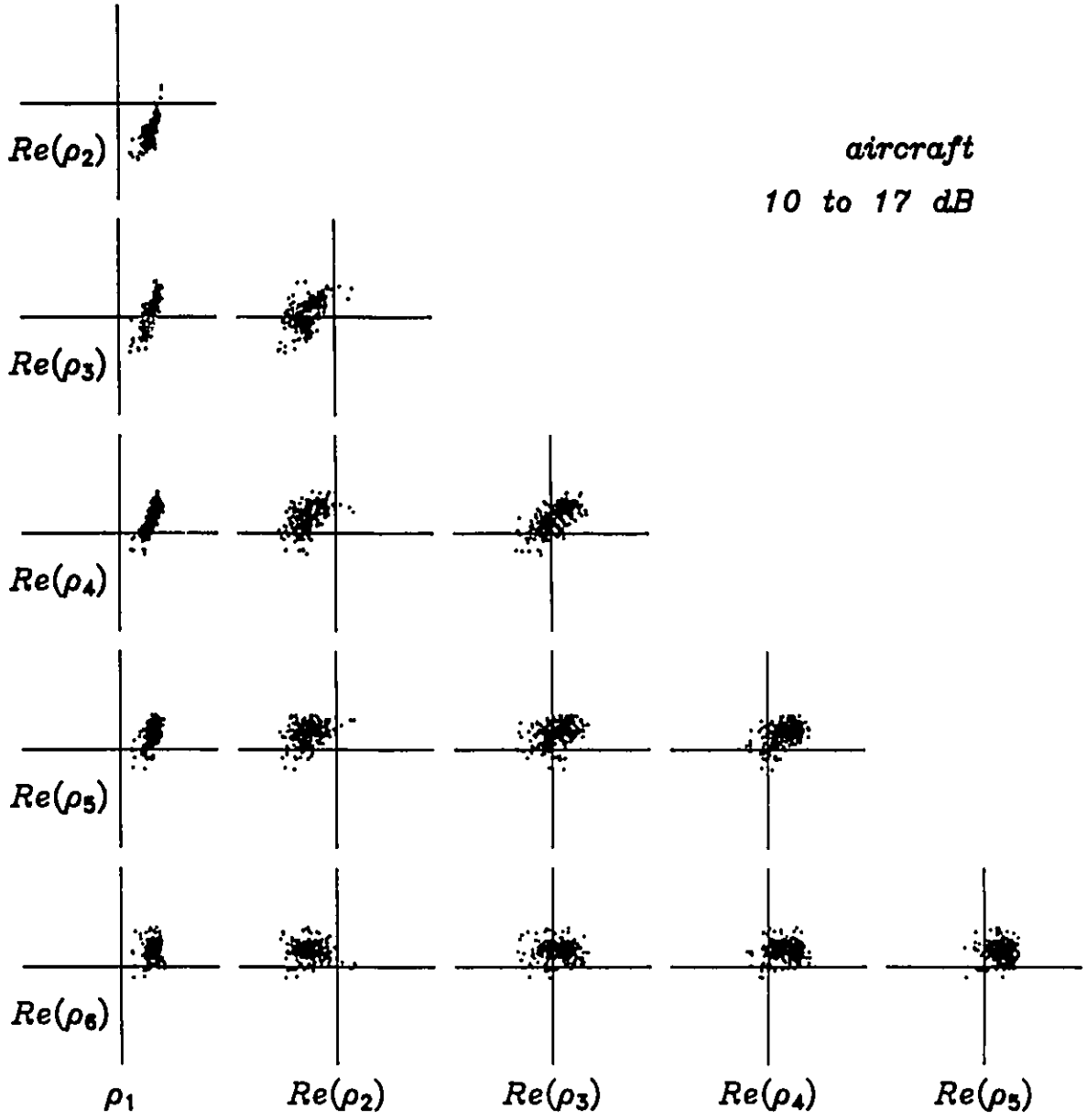


Figure 6.13b Matrix of scatter plots for aircraft targets in SNR range C

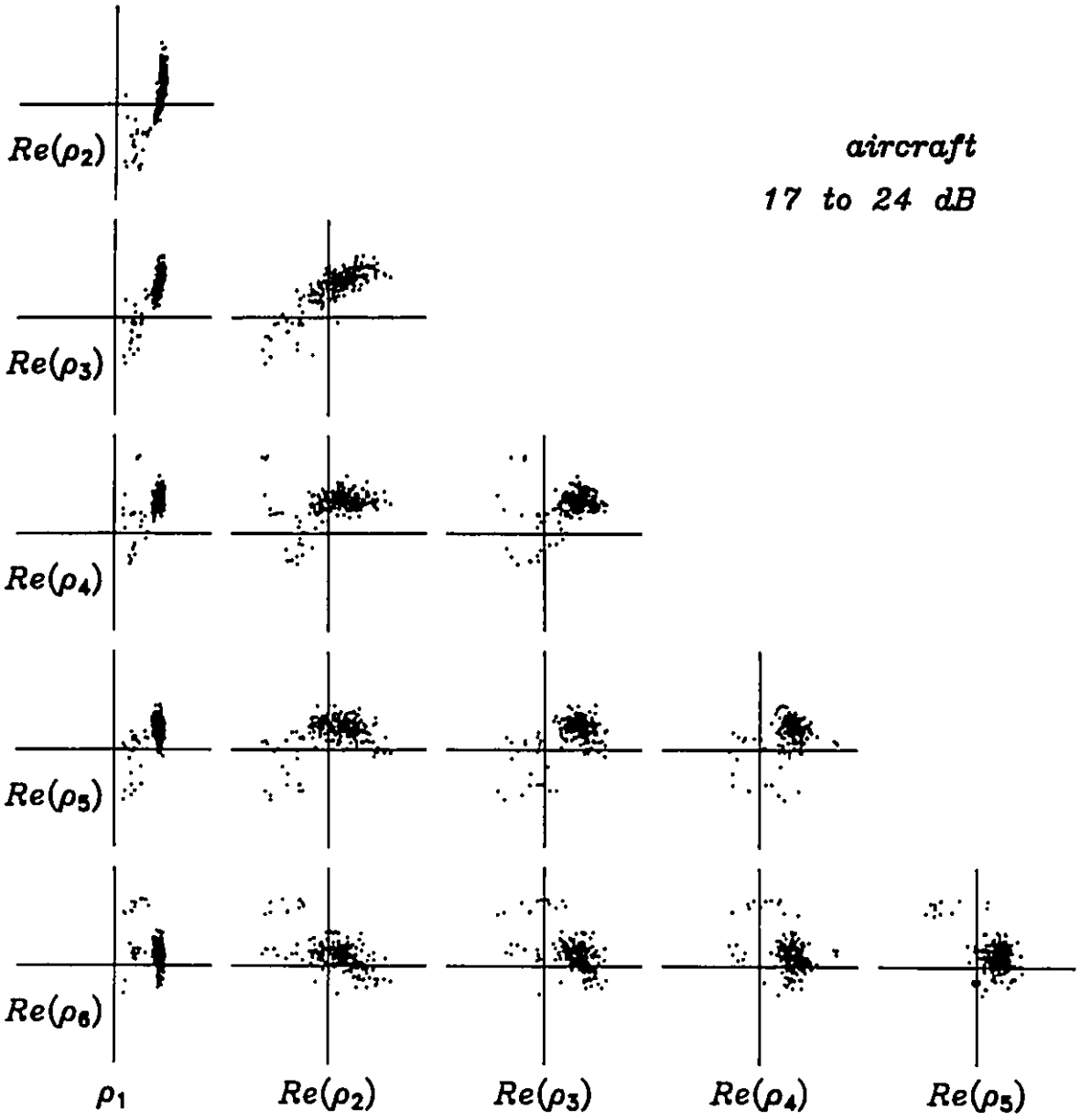


Figure 6.13c Matrix of scatter plots for aircraft targets in SNR range D

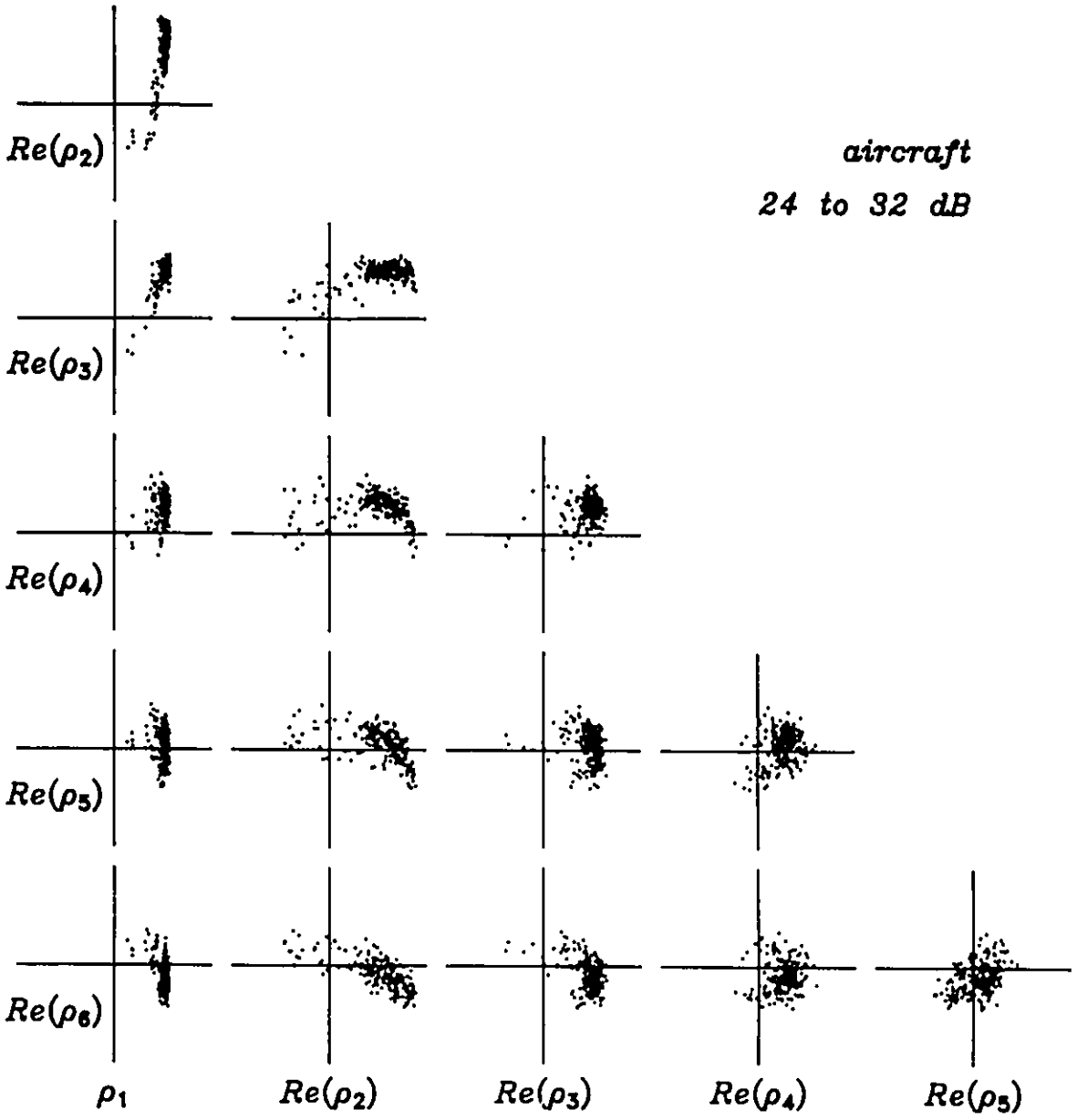


Figure 6.13d Matrix of scatter plots for aircraft targets in SNR range E

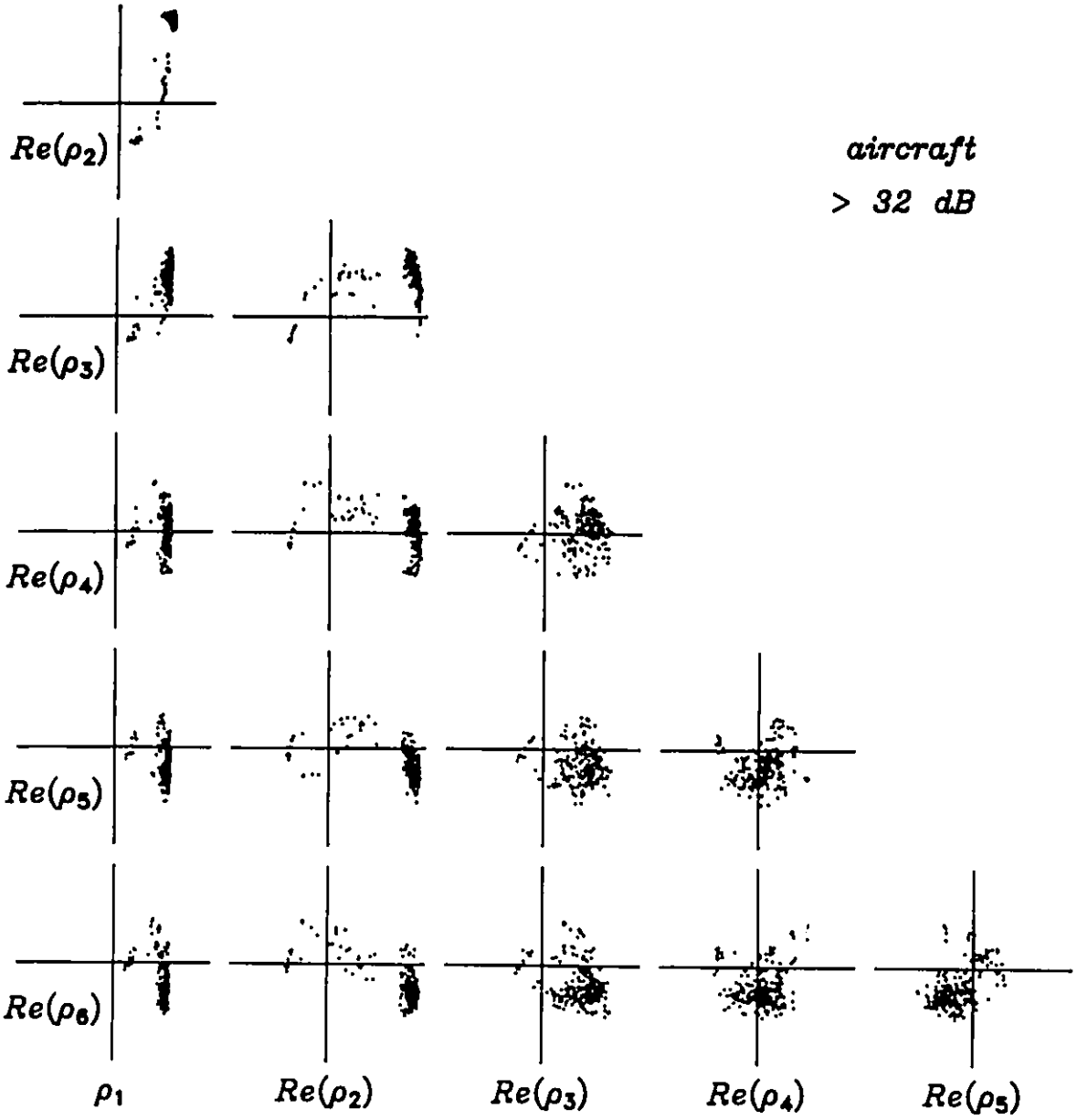


Figure 6.13e Matrix of scatter plots for aircraft targets in SNR range F

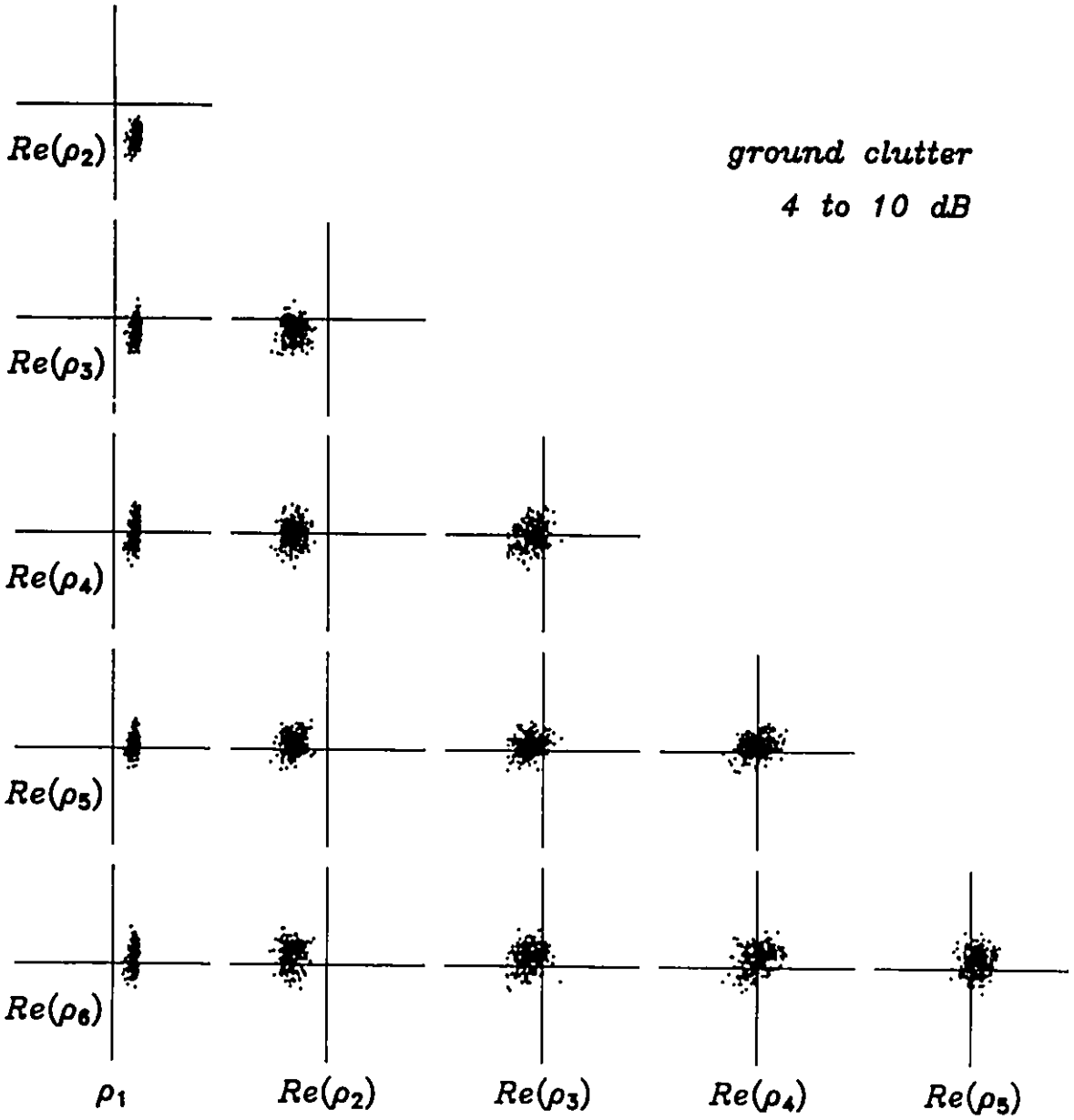


Figure 6.14a Matrix of scatter plots for ground clutter in SNR range B

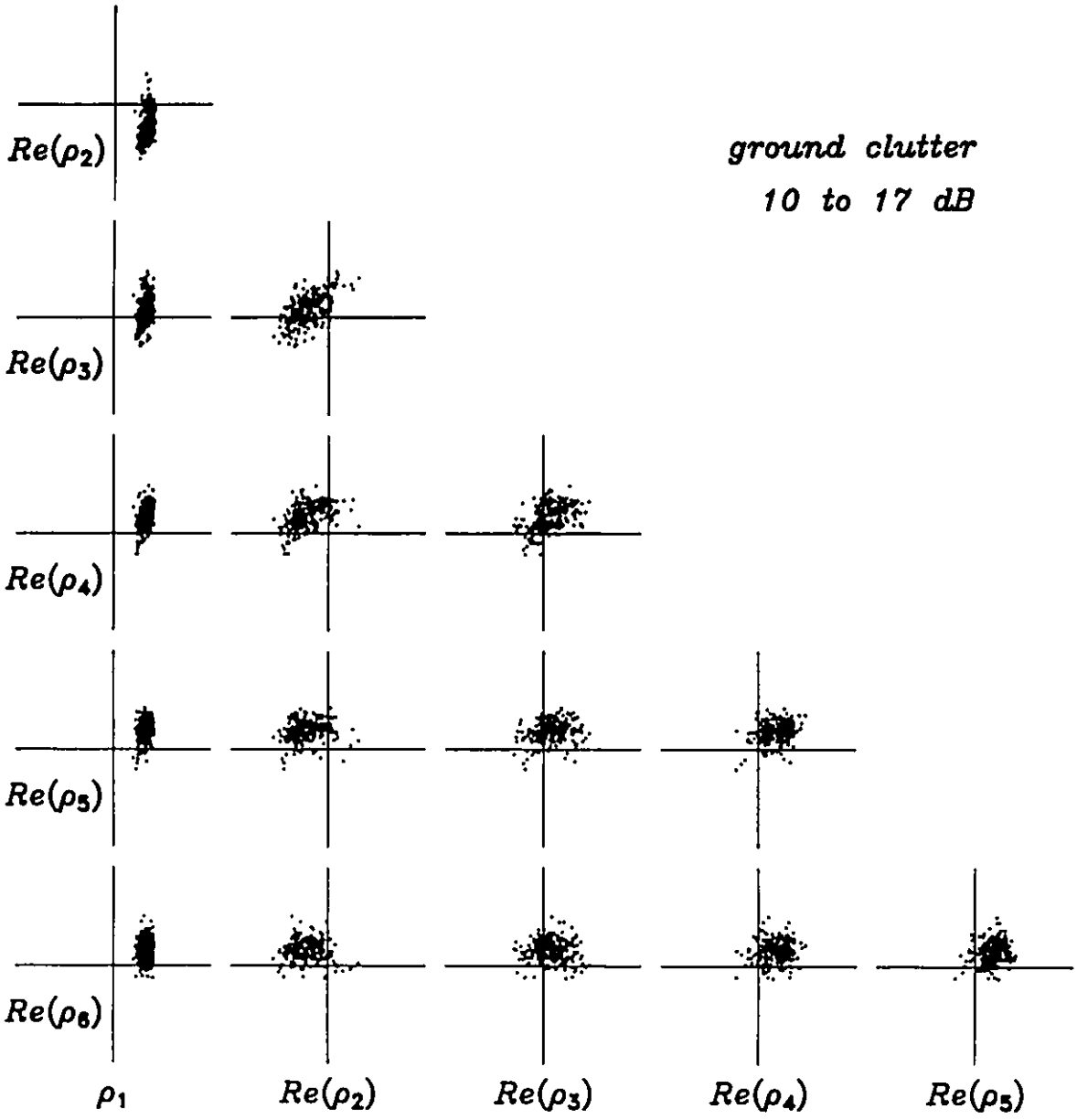


Figure 6.14b Matrix of scatter plots for ground clutter in SNR range C

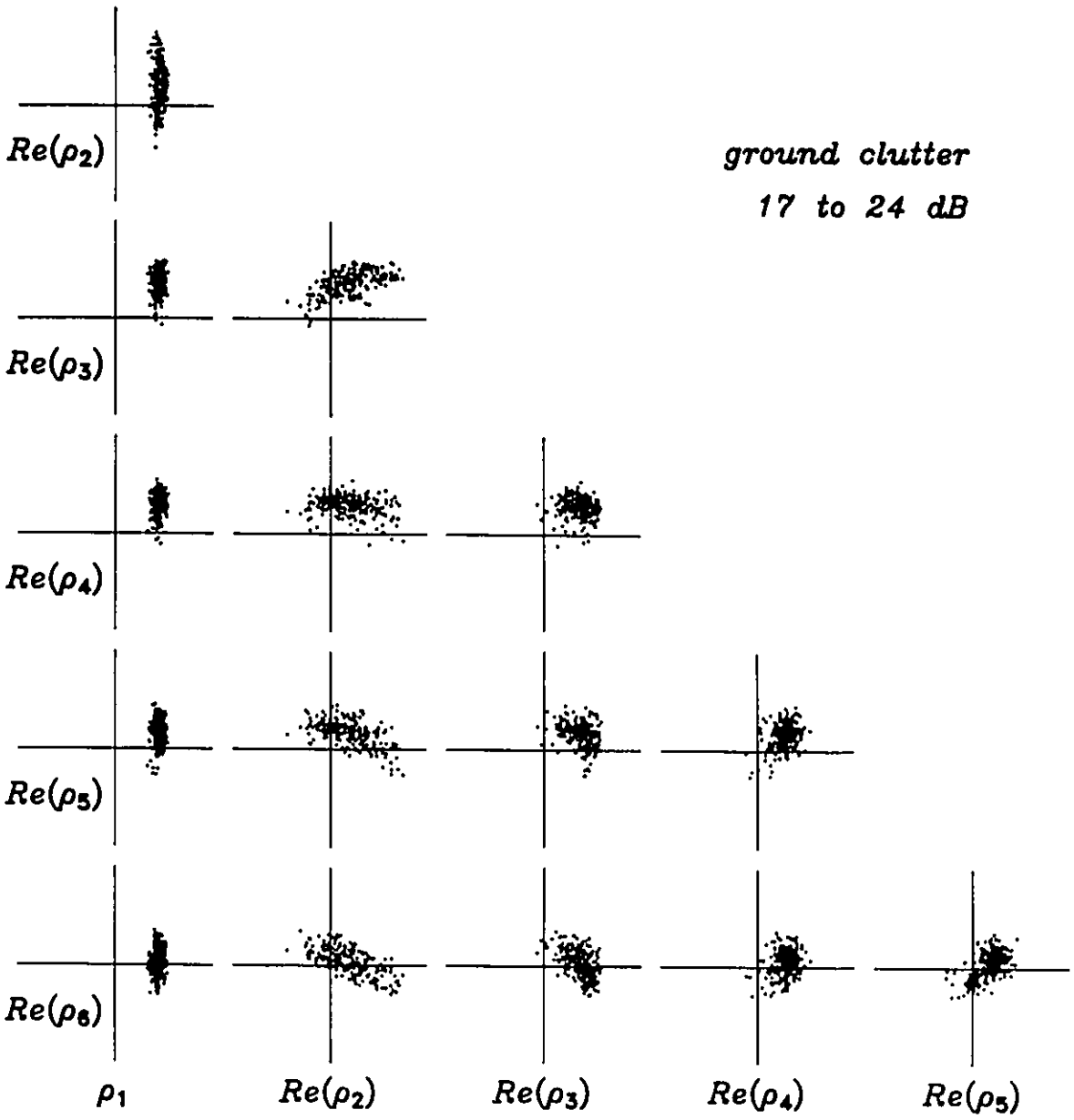


Figure 6.14c Matrix of scatter plots for ground clutter in SNR range D

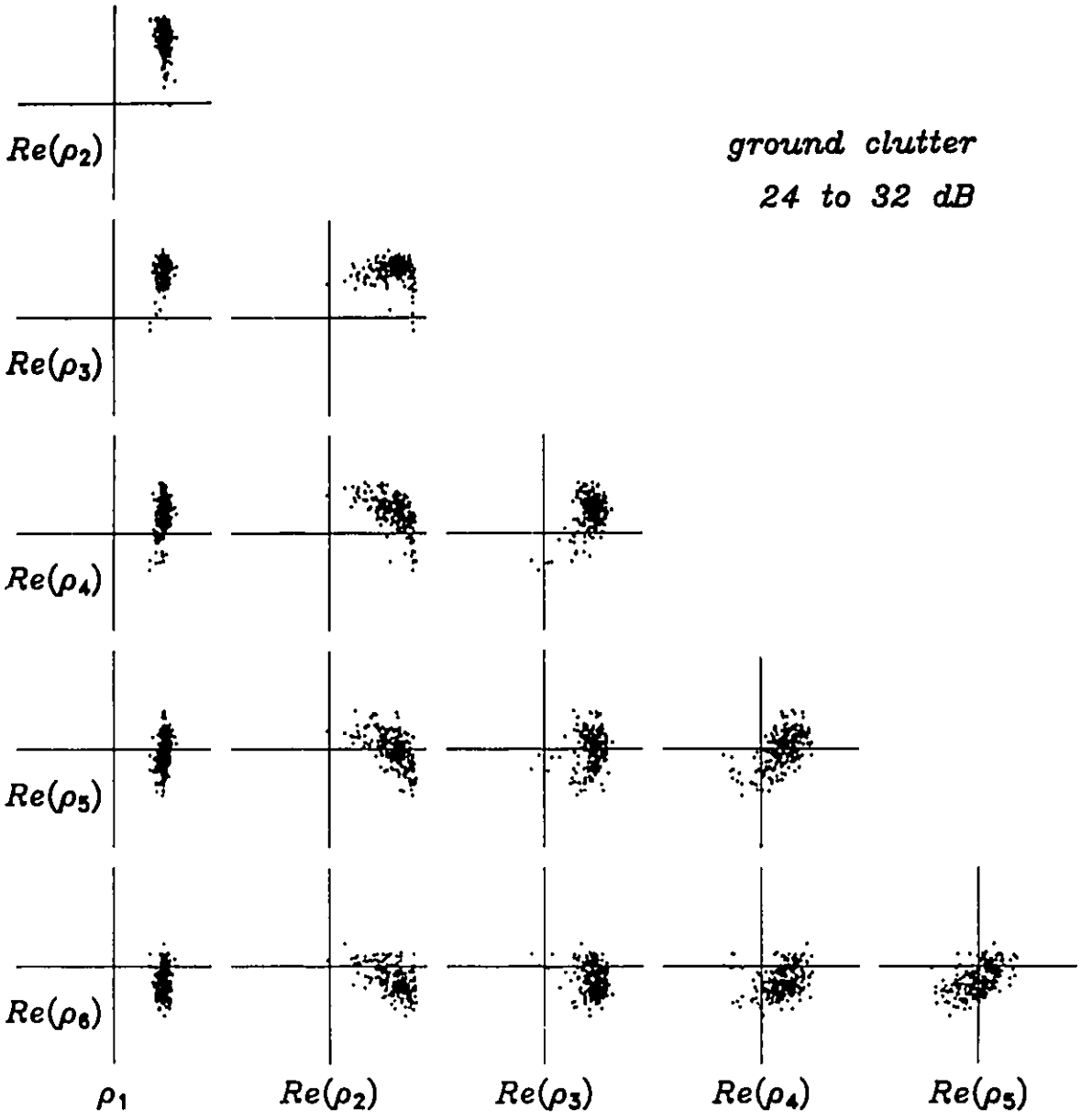


Figure 6.14d Matrix of scatter plots for ground clutter in SNR range E

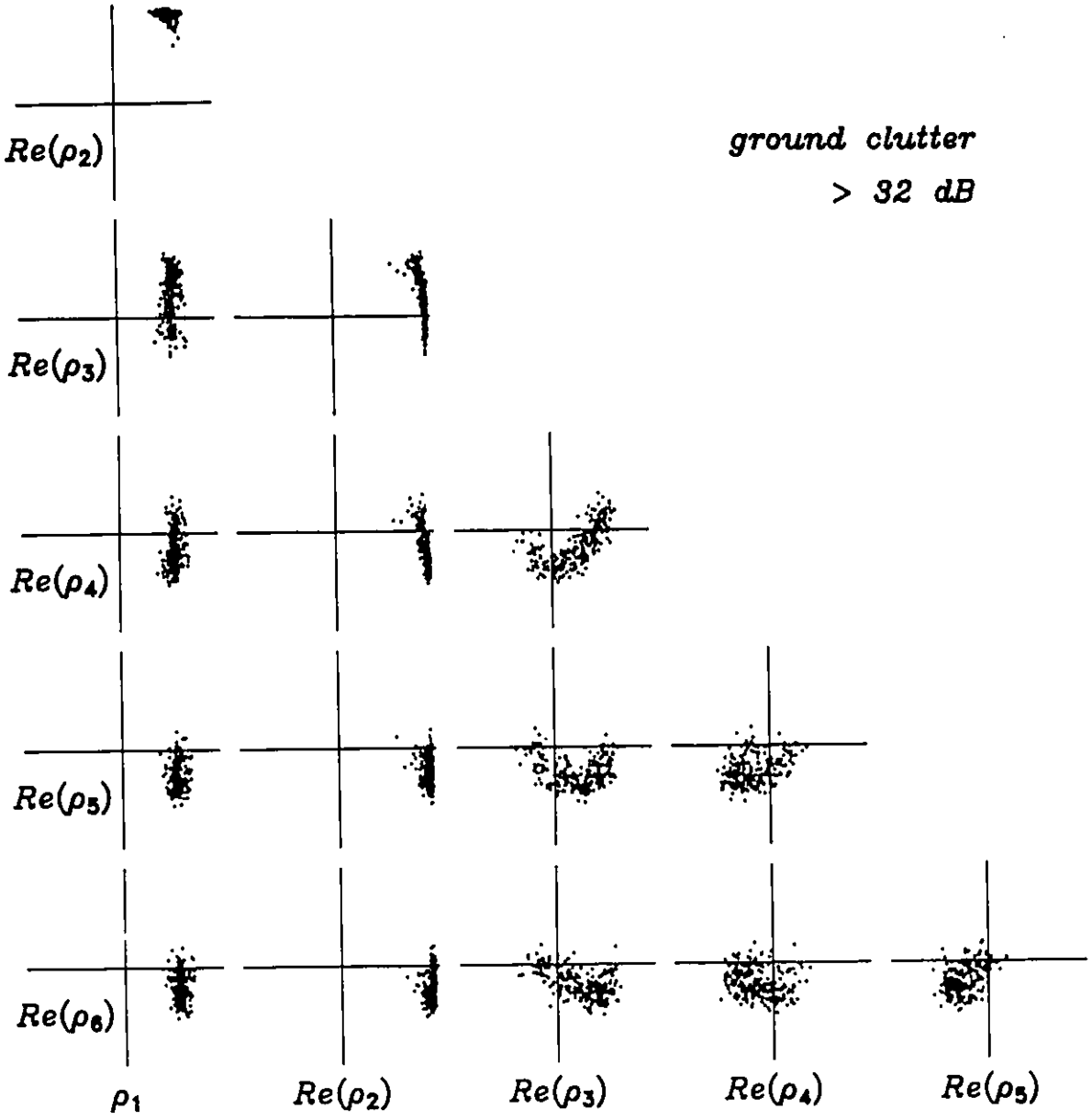


Figure 6.14e Matrix of scatter plots for ground clutter in SNR range F

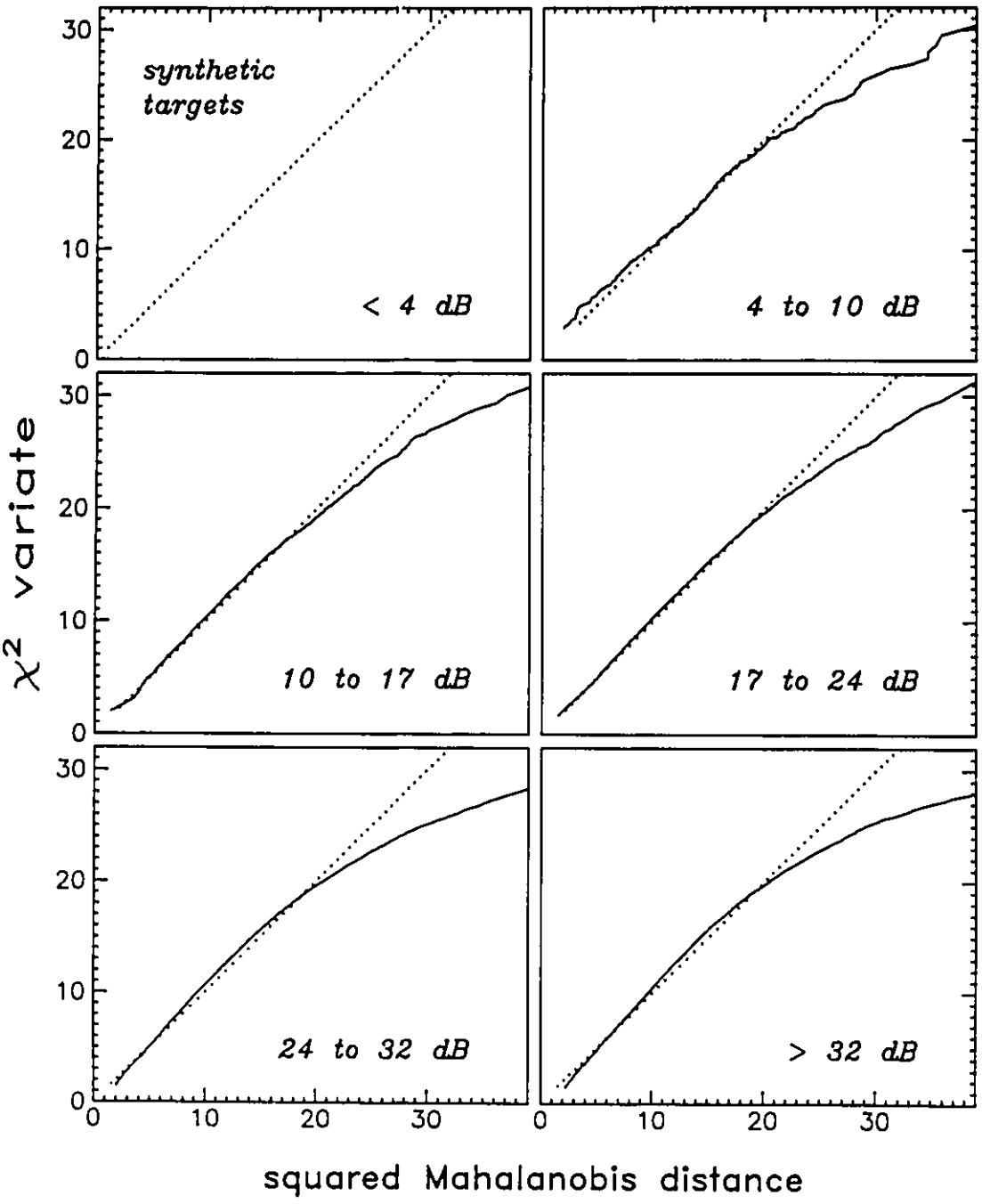


Figure 6.15a Quantile probability plots for D^2 of all synthetic target classes

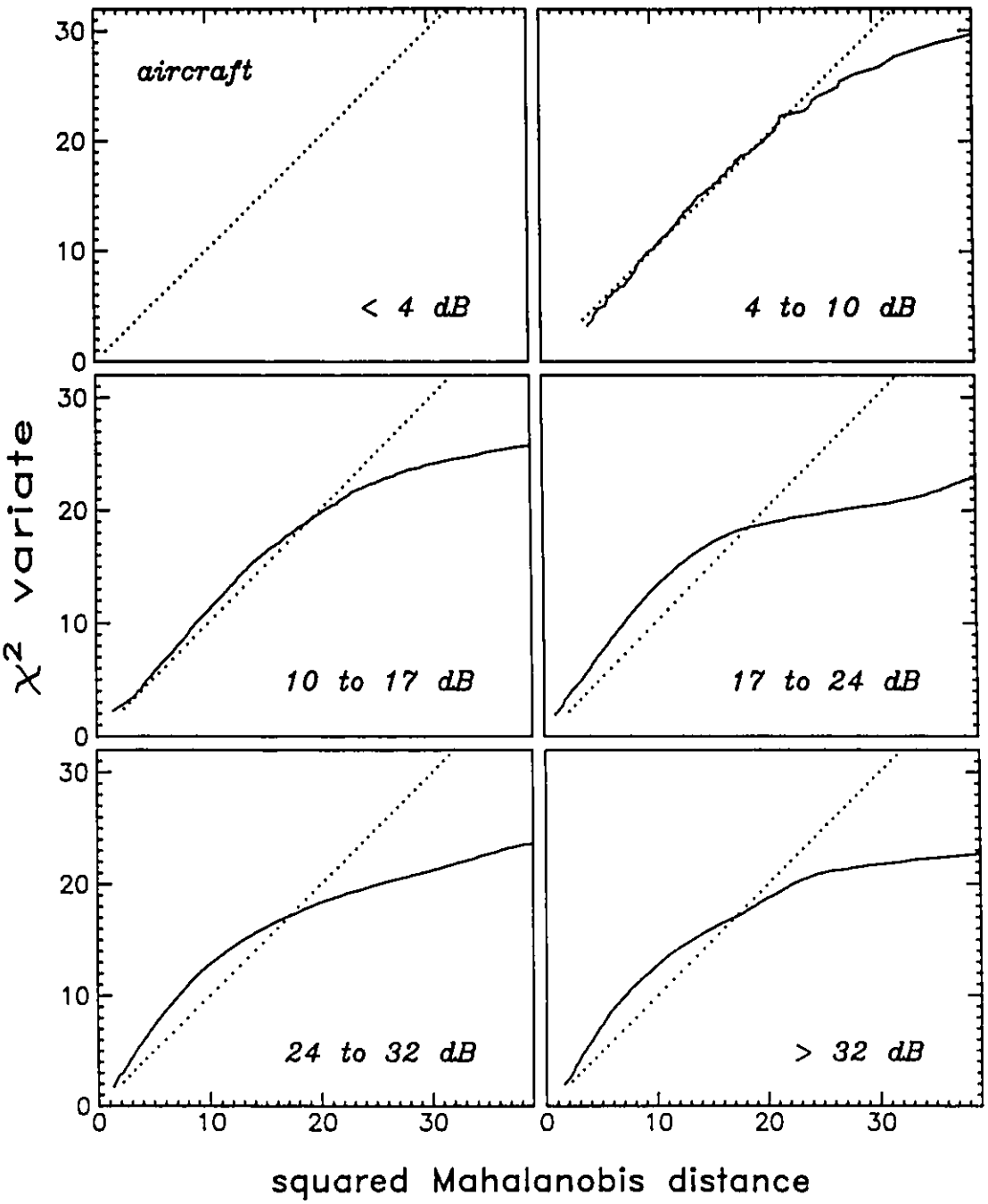


Figure 6.15b Quantile probability plots for D^2 of all aircraft classes

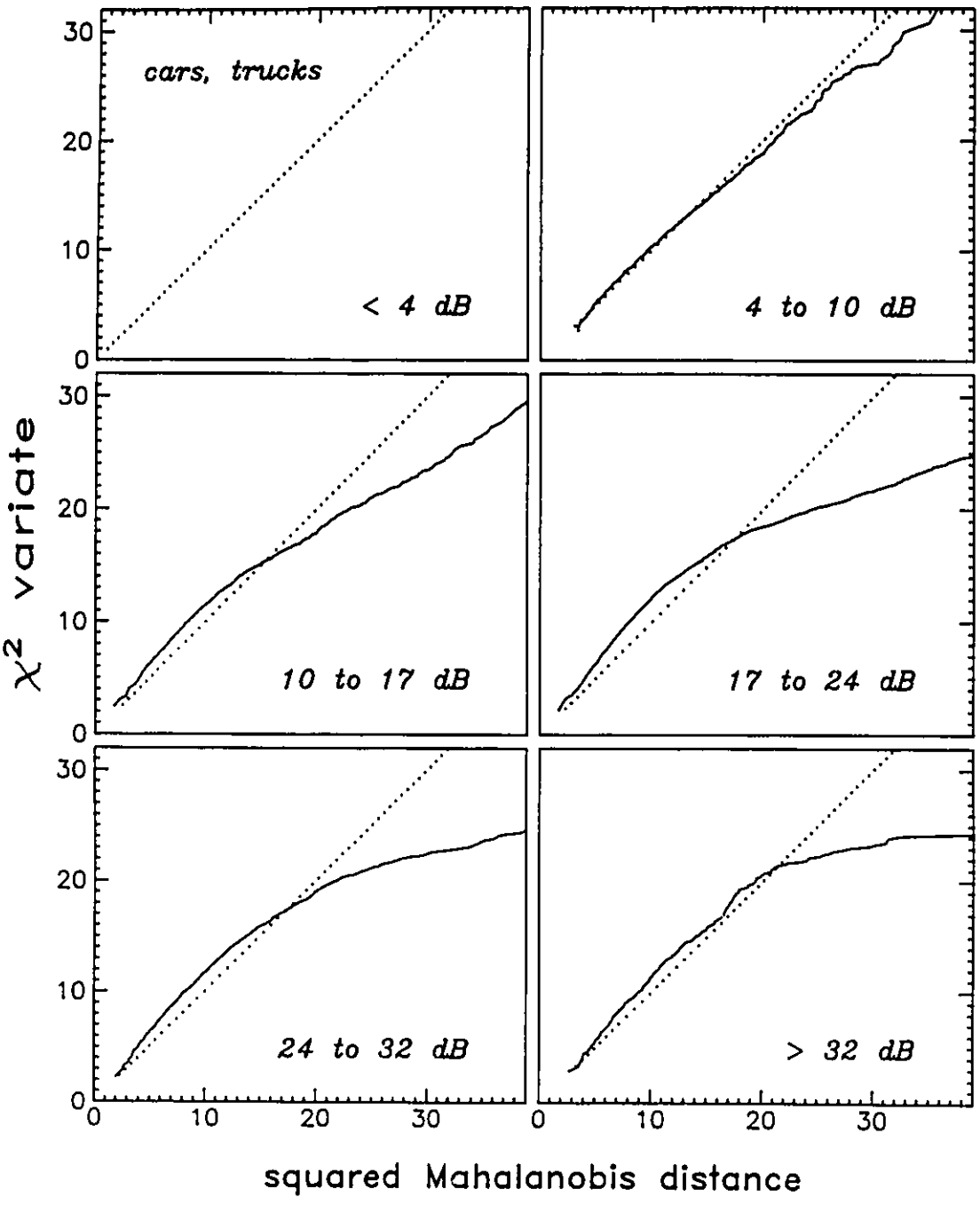


Figure 6.15c Quantile probability plots for D^2 of all cars and trucks classes

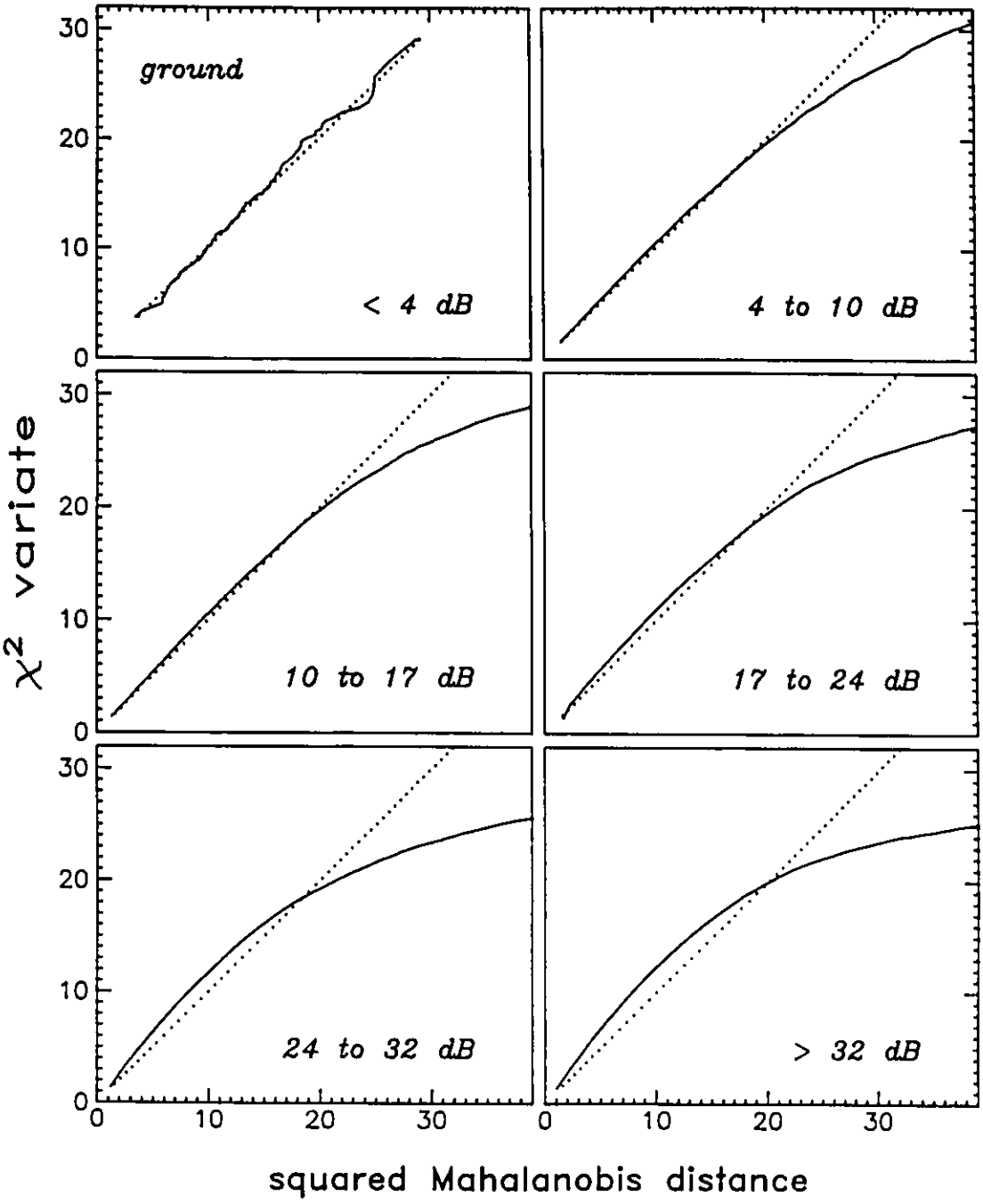


Figure 6.15d Quantile probability plots for D^2 of all ground clutter classes

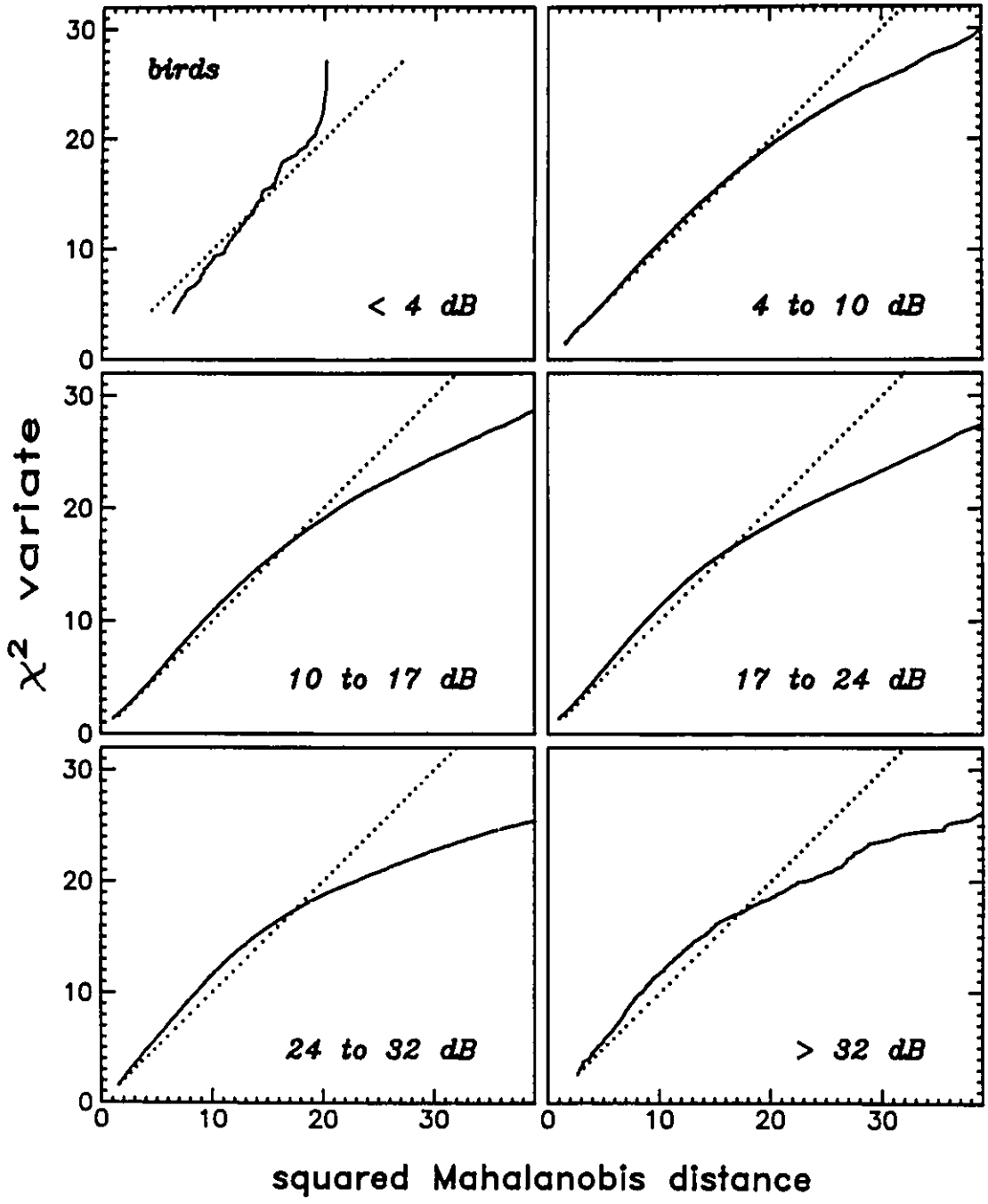


Figure 6.15e Quantile probability plots for D^2 of all bird clutter classes

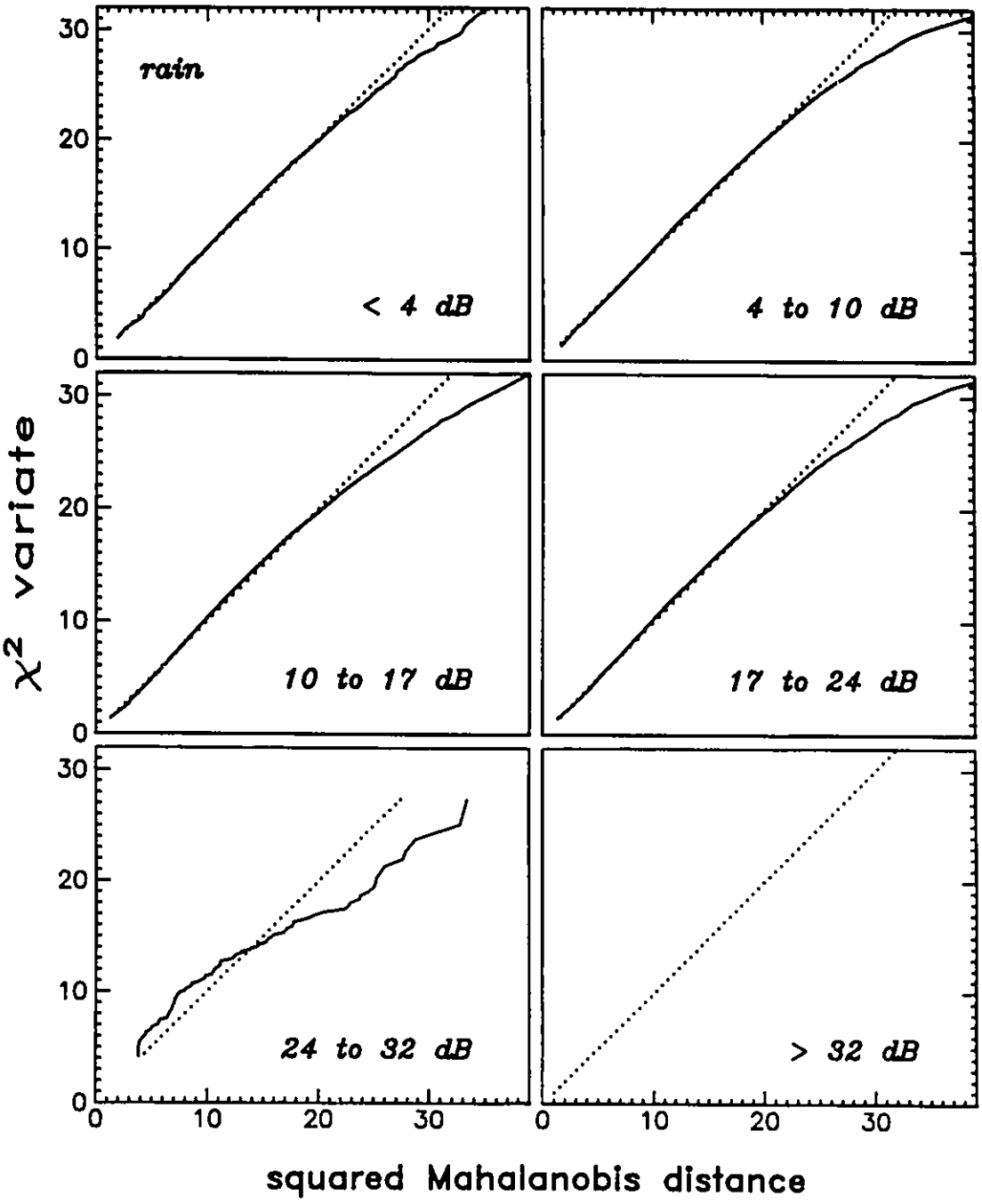


Figure 6.15f Quantile probability plots for D^2 of all rain clutter classes

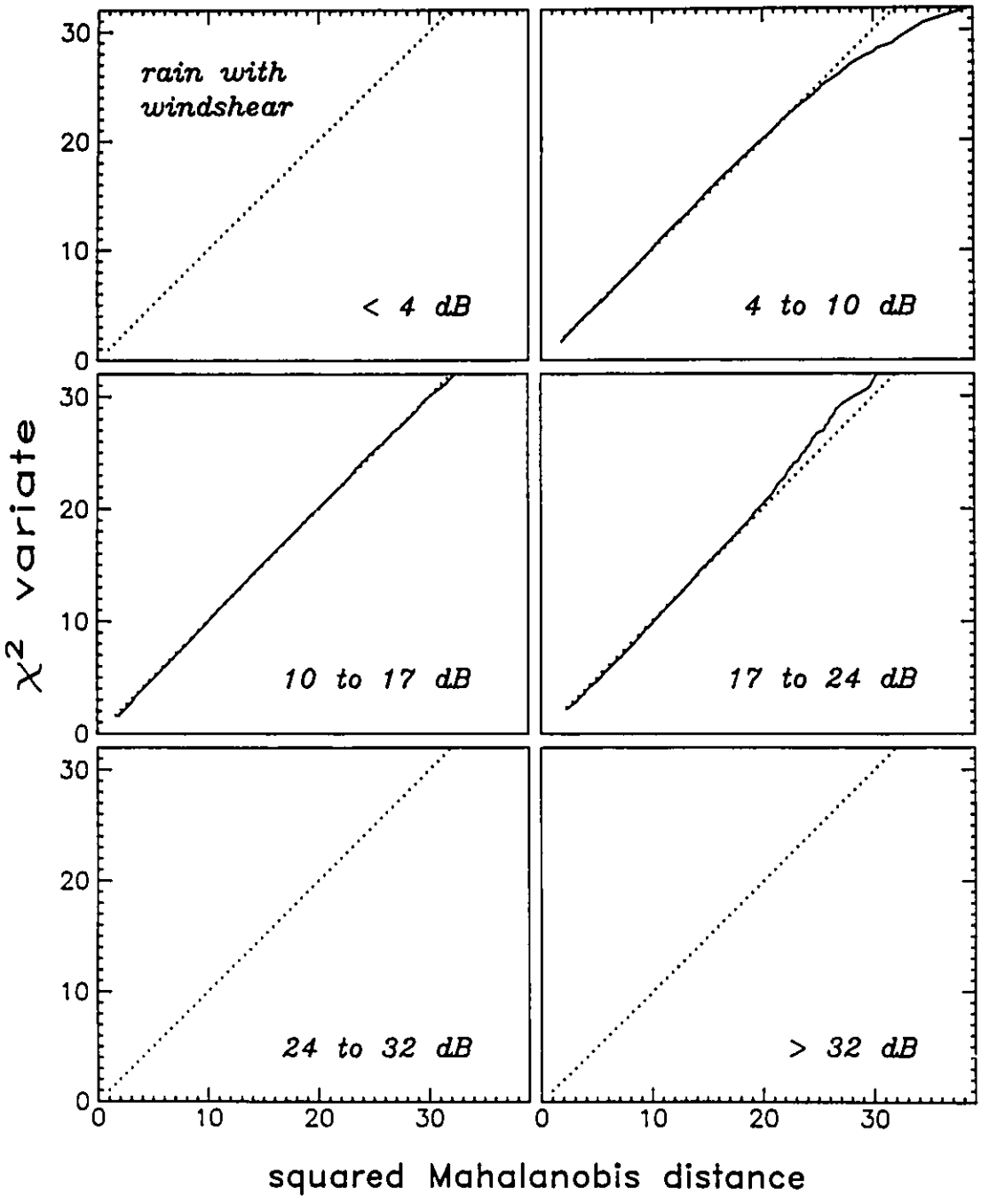


Figure 6.15g Quantile probability plots for D^2 of all rain with windshear classes

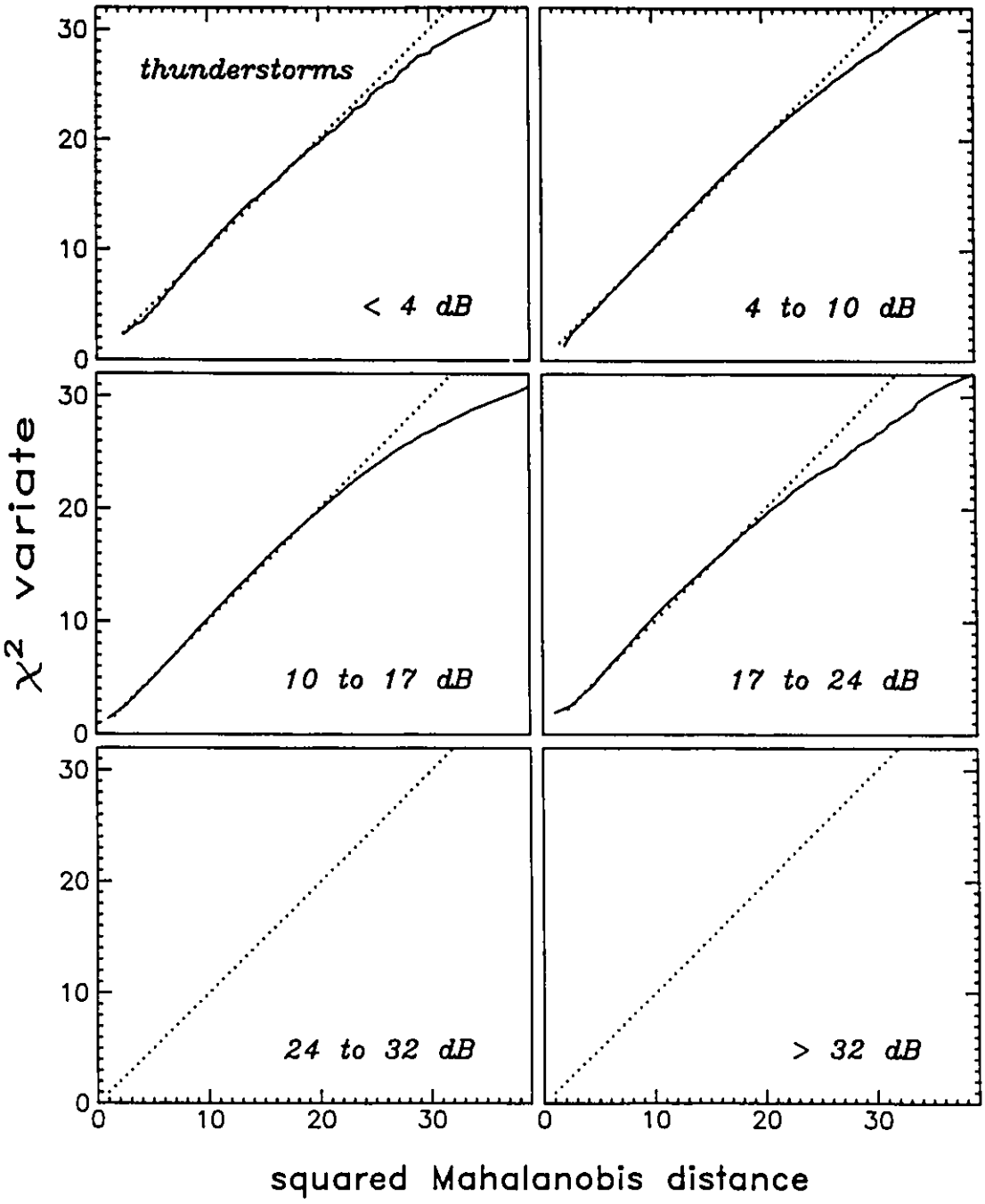


Figure 6.15h Quantile probability plots for D^2 of all thunderstorm classes

more likely that the residual non-normal behavior has been masked. This fact points to the caution which must be applied to the use of this type of plot as a test for multivariate normality.

6.2.4 Expected Classifier Performance

Several predictors of performance have been discussed in Chapter 4. It is convenient to concentrate on one of these for comparison and the analysis of trends. The Bhattacharyya distance is a useful measure, and its relationship with the upper and lower error bounds is given by Equation (4.42), and plotted in Figure 4.1. An enlarged section is shown for convenience in Figure 6.16.

One final unanswered question concerns the number of features to include in the class prototypes, keeping in mind the cost versus performance tradeoff. Figure 6.17 shows the increase in class separability of several selected class pairs as measured by the Bhattacharyya distance. While the distance keeps increasing as more reflection coefficients are added, there appears to be no clearly defined stopping point. It may be possible, however, to stop once a given performance is reached. A Bhattacharyya distance of 3 corresponds to an upper error bound of about 2.5%, and may serve as such a point. Unfortunately, some class pairs never reach this level of separability, which leaves the question unanswered. The performance shown in Figure 6.17a assumes that the features P_{dif} and P_{var} were not used. Figures 6.17b and 6.17c show the impact of these two features to be rather significant, especially when attempting to distinguish between "point target" classes such a aircraft, cars and trucks, and of course synthetic targets, and distributed-target classes including any type of weather and bird clutter.

Classification experiments, which will be discussed more completely in the

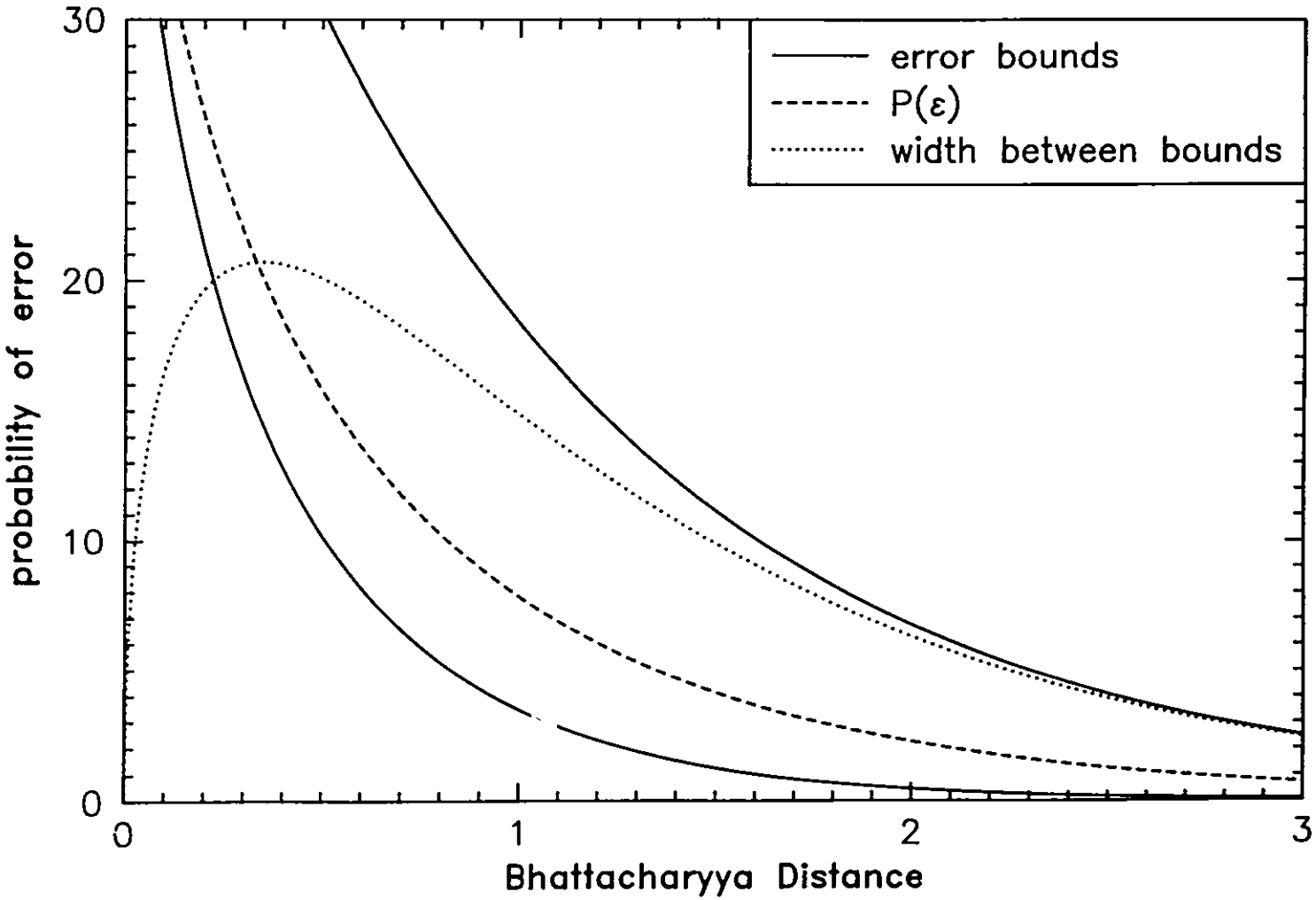


Figure 6.16 Bhattacharyya distances and error bounds

For an explanation see Figure 4.1. The region of interest has been magnified.

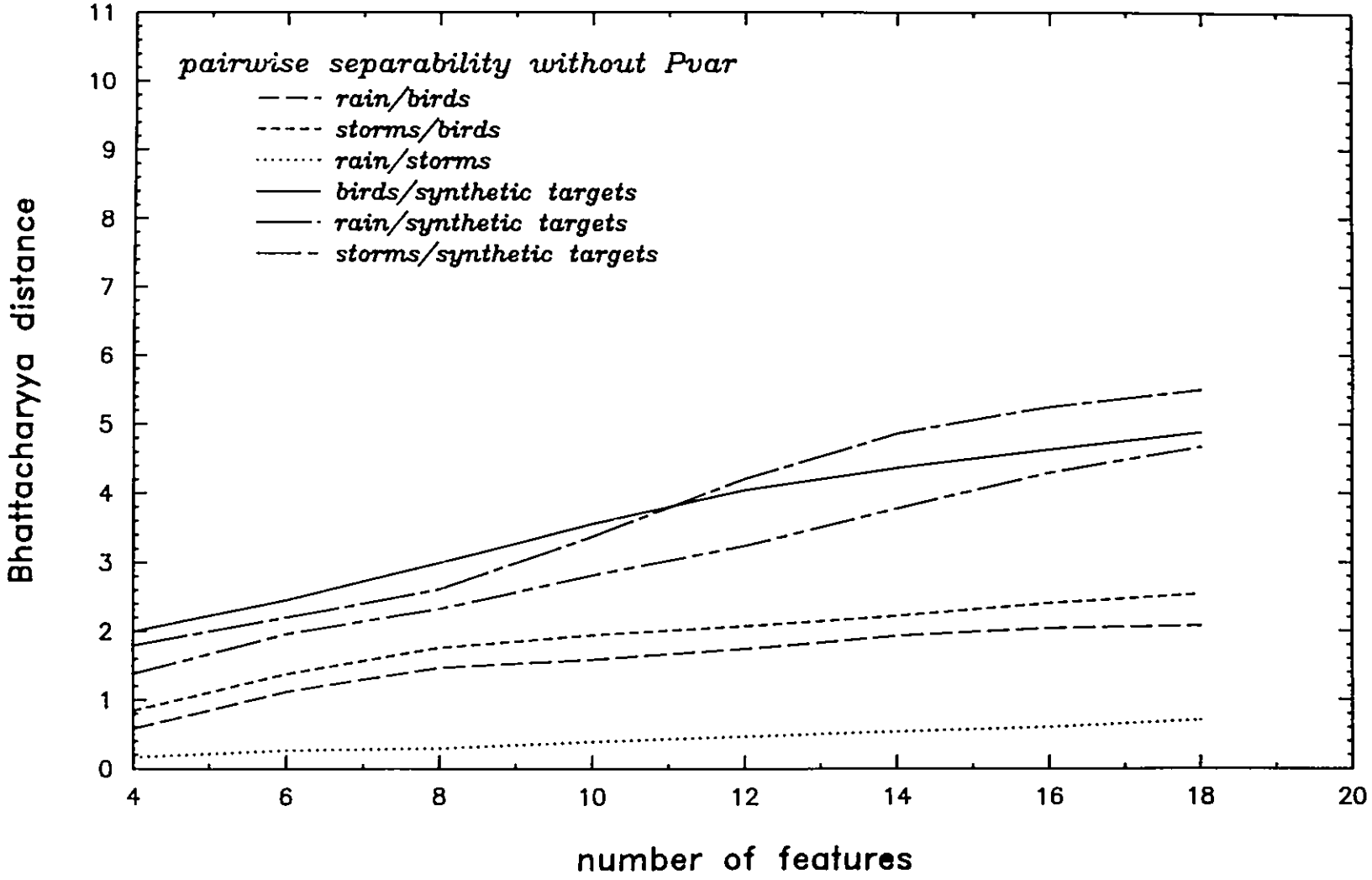


Figure 6.17a Pairwise class separabilities versus the number of features, not including *Pvar* and *Pdif*

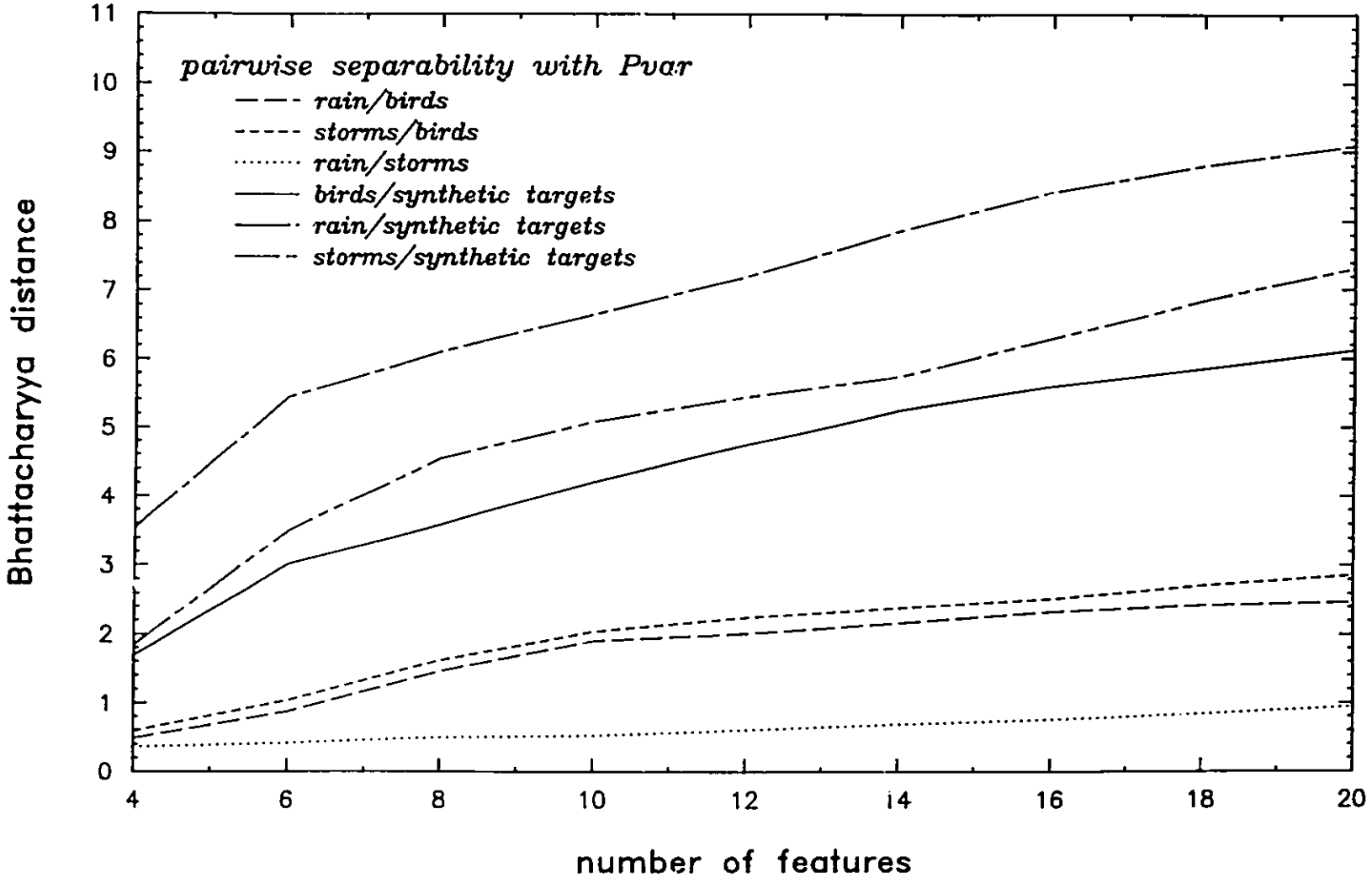


Figure 6.17b Pairwise class separabilities versus the number of features, including *Pvar* and *Pdif*

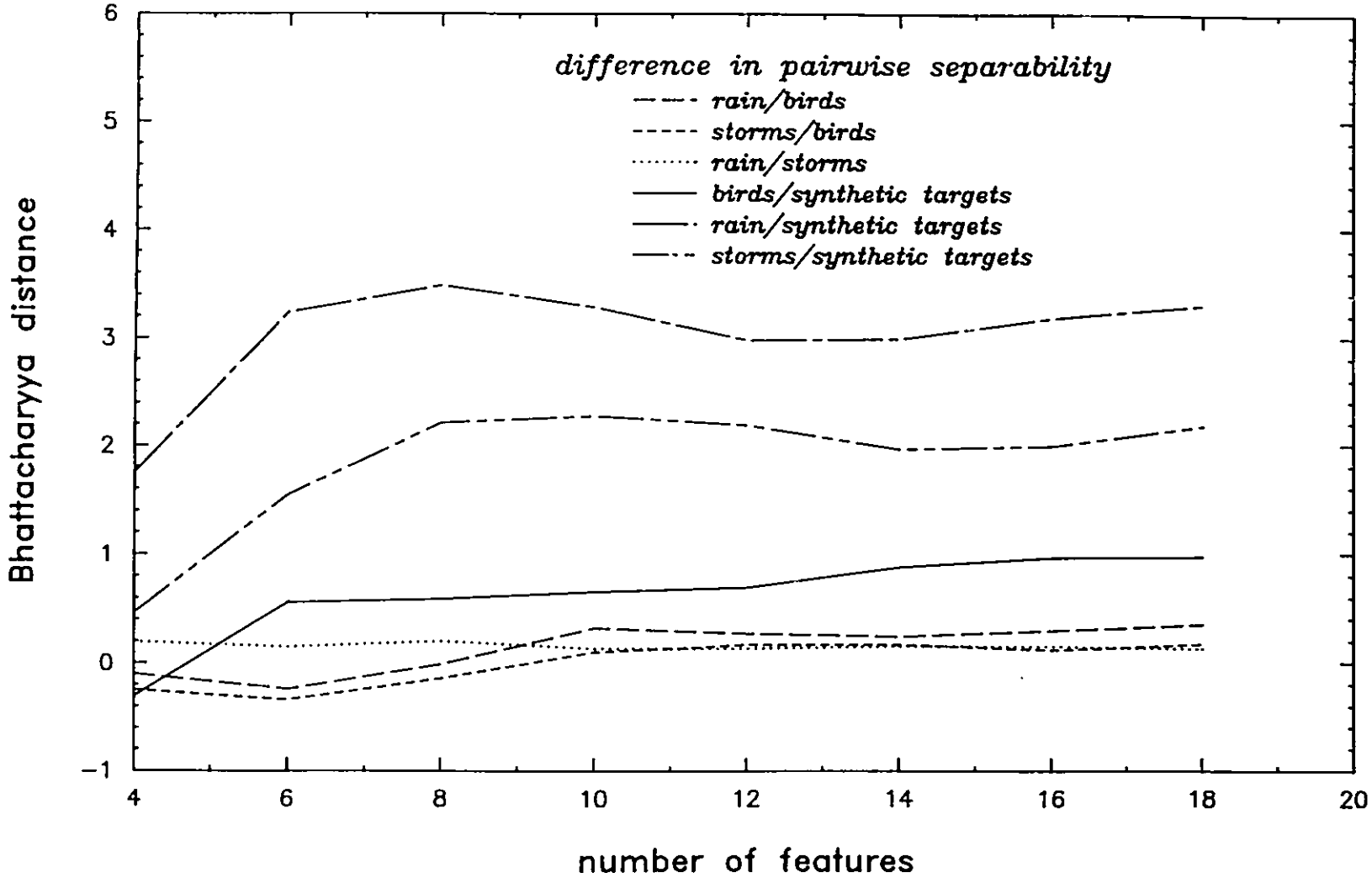


Figure 6.17c Increase in class separabilities due to the of features *Pvar* and *Pdif*

next section, tend to develop class error rates rather than pairwise rates. While there exist no theoretically satisfying class error rate predictors, an approximate one may be established using Equation (4.44). The individual pairwise error rates arising from the divergence calculation - Equations (4.24) and (4.26) - are summed to arrive at an approximate class error rate. As discussed in Chapter 4 this is, strictly speaking, not theoretically correct, but it does provide an indication of where the classifier is likely to perform. Figure 6.18 shows these rates for three different classes, and a comparison to the experimental rates of the bird clutter class. Again, the impact of using the features P_{dif} and P_{var} is clear. It seems particularly strong for lower orders and causes a leveling of the error rates once the number of features has reached 10. It should be noted here that, for the cases shown in Figures 6.17 and 6.18, class widths of 4 dB were used, centered around U_0 = 20 dB, and U_0 was retained as a feature. As a result, the separabilities were influenced by the difference in the mean values of U_0 for each class and cannot be considered the final results. For these, neither the Doppler frequency ϕ nor U_0 were used in the classifier; they were used only for pretesting, as outlined in Section 6.2.1. In view of Figures 6.17 and 6.18, the final features were u_0, ρ_2', ρ_3', ρ_4', ρ_5', P_{dif}, and P_{var}. Since the reflection coefficients are complex, the number of features is thus 11.

Pairwise performance expectations of this size of feature set, given the SNR ranges A to F, are plotted in Figure 6.19. The trend towards increasing separability with rising signal strength is clearly evident, although some clutter types remain difficult to distinguish from related types, even at large signal levels. This points to an inherent overlap in the Doppler characteristics which cannot likely be eliminated using the type of single scan processing contemplated in this thesis.

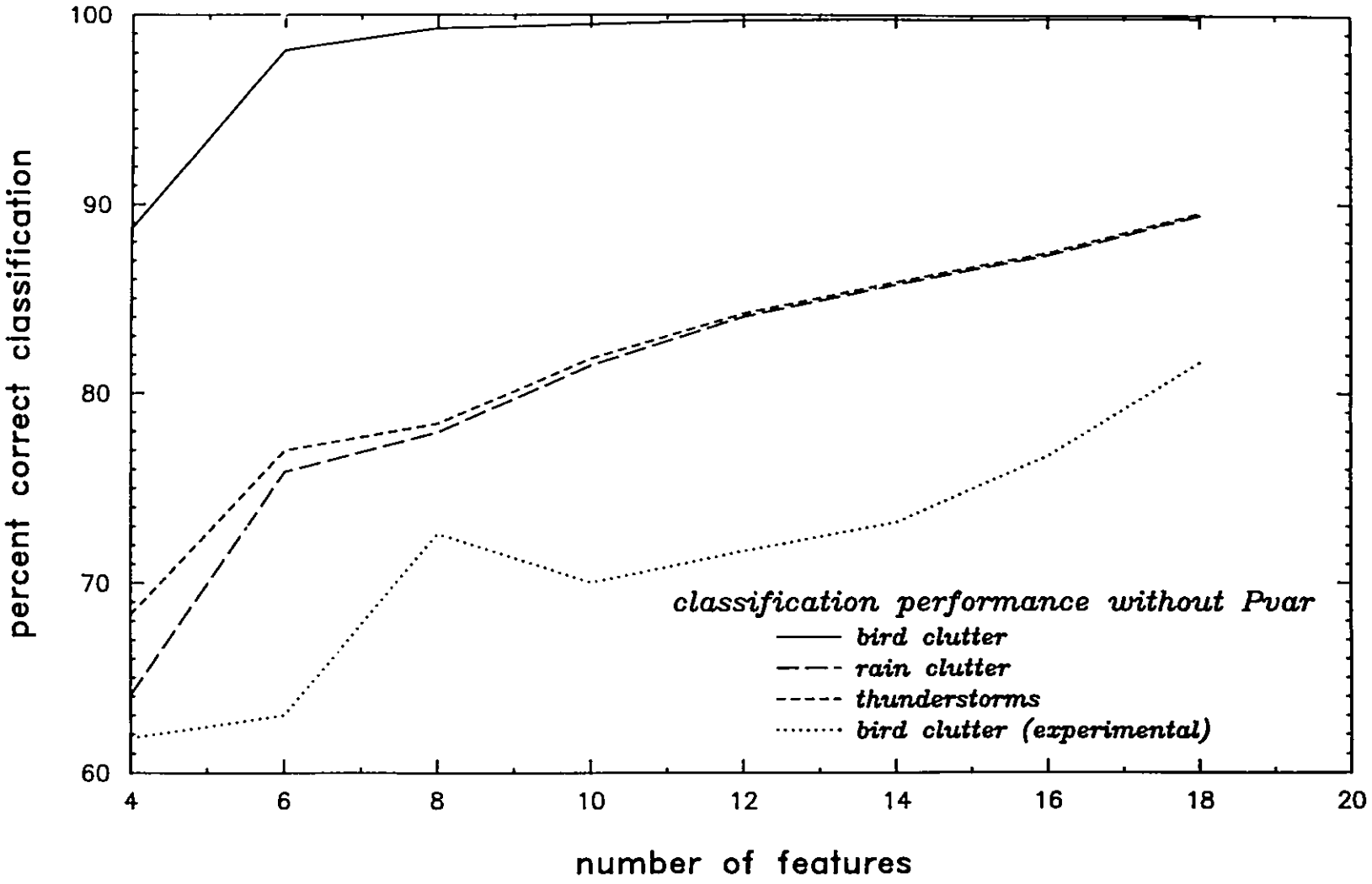


Figure 6.18a Predicted rates for correct classification, without using features *Pvar* and *Pdif*

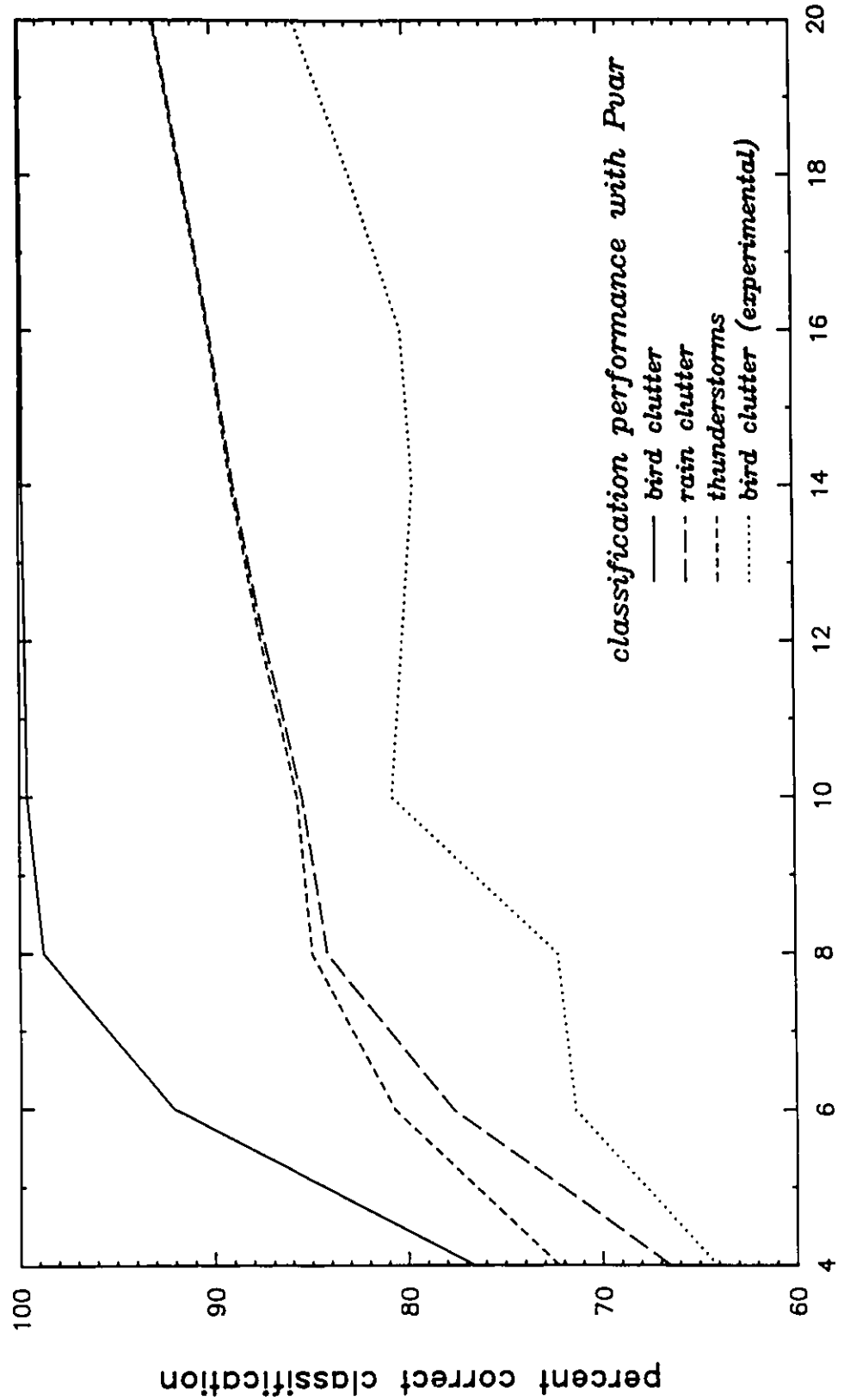


Figure 6.18b Predicted rates for correct classification, using features *Pvar* and *Pdif*

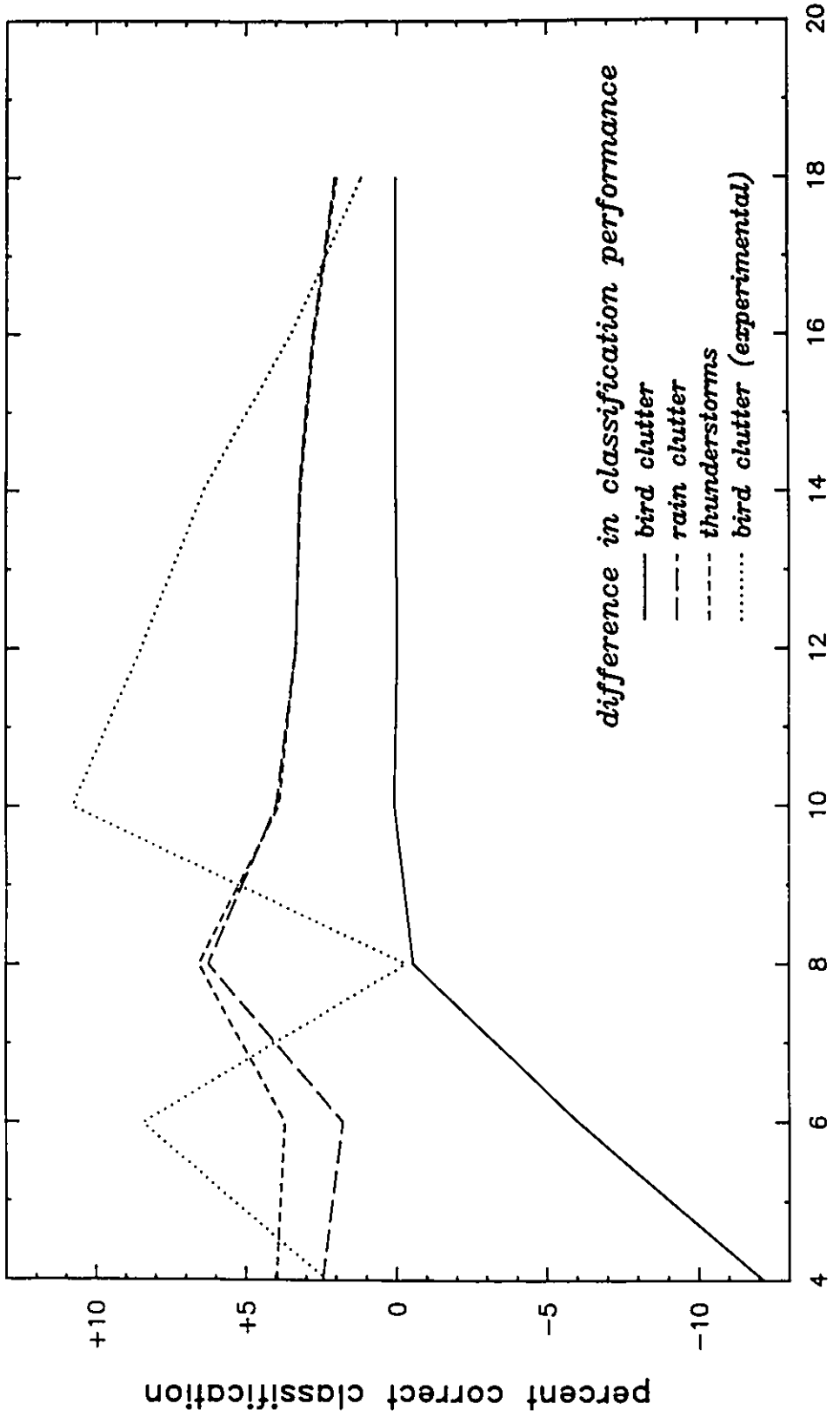


Figure 6.18c Increase in the rates for correct classification, due to the use of features *Pvar* and *Pdif*

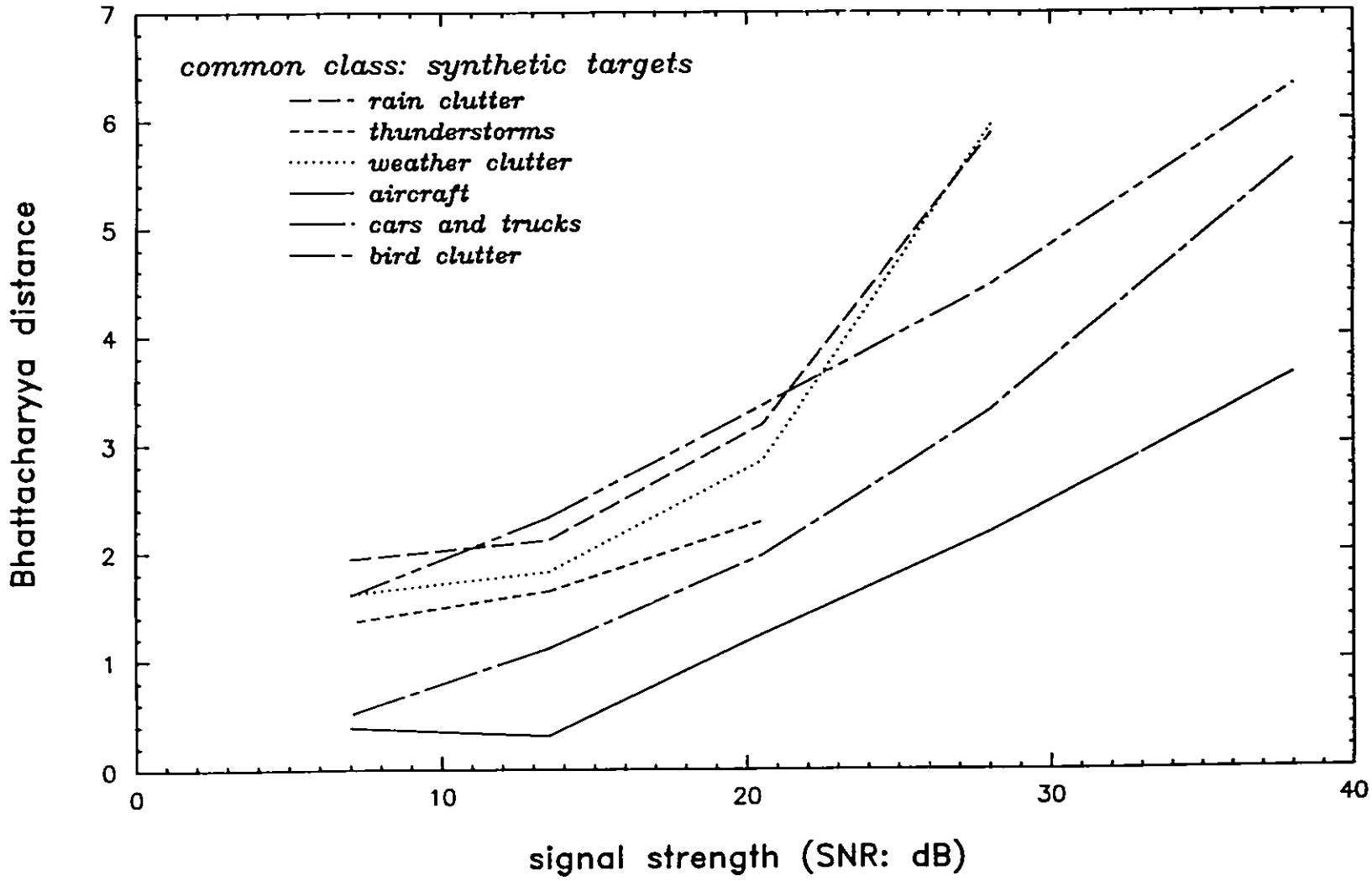


Figure 6.19a Class separabilities versus signal strength for synthetic targets

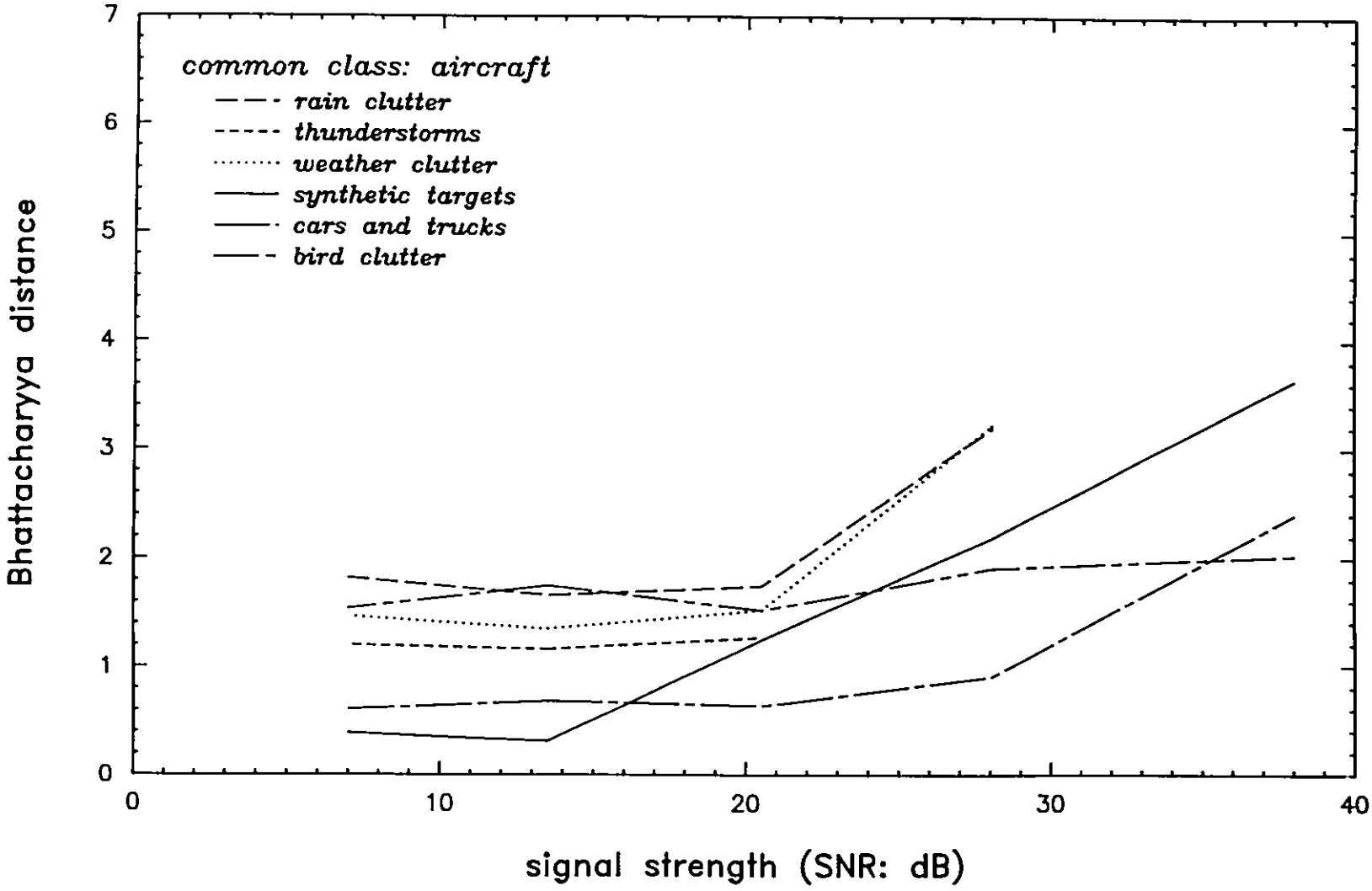


Figure 6.19b Class separabilities versus signal strength for aircraft targets

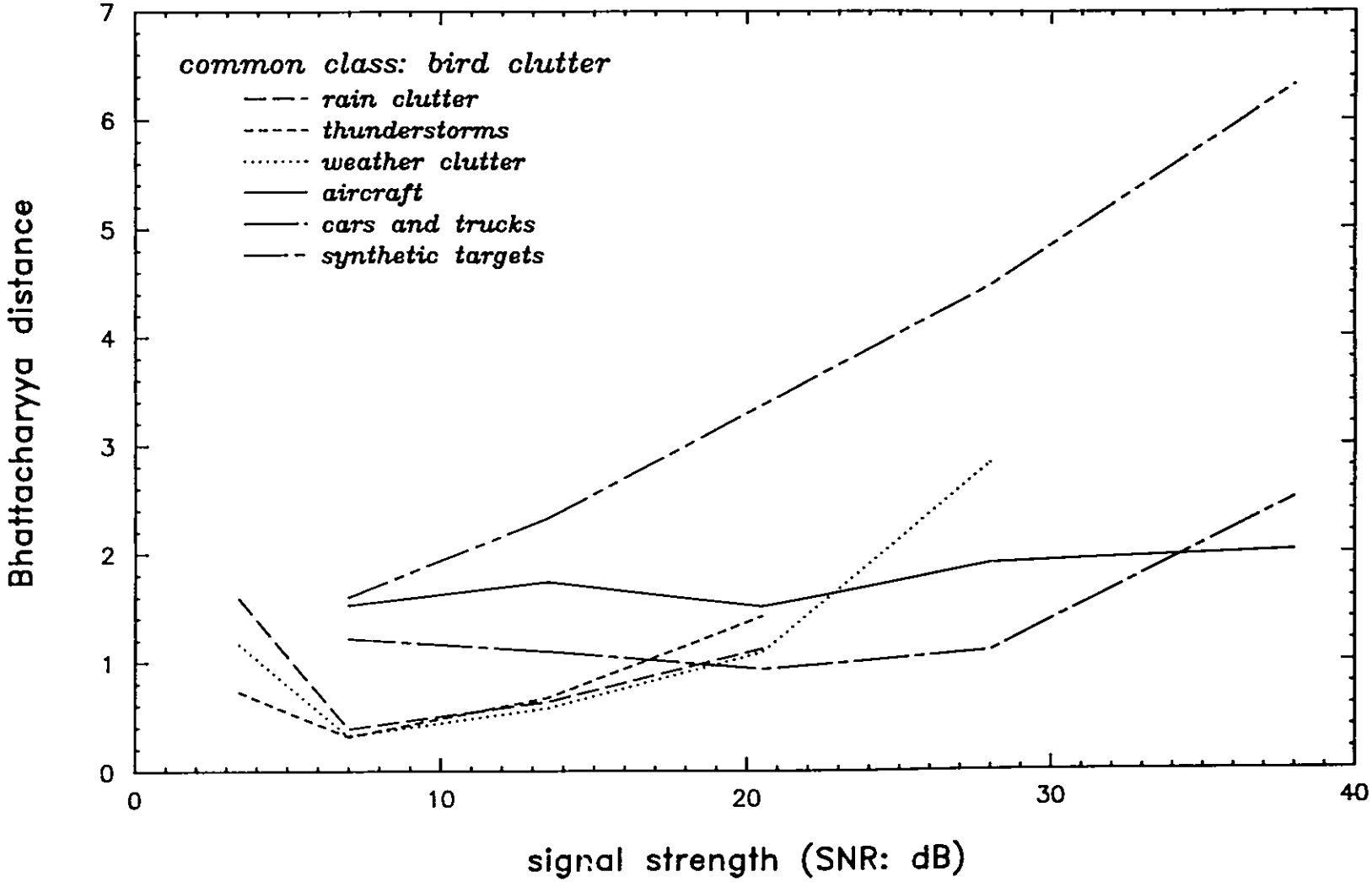


Figure 6.19c Class separabilities versus signal strength for bird clutter

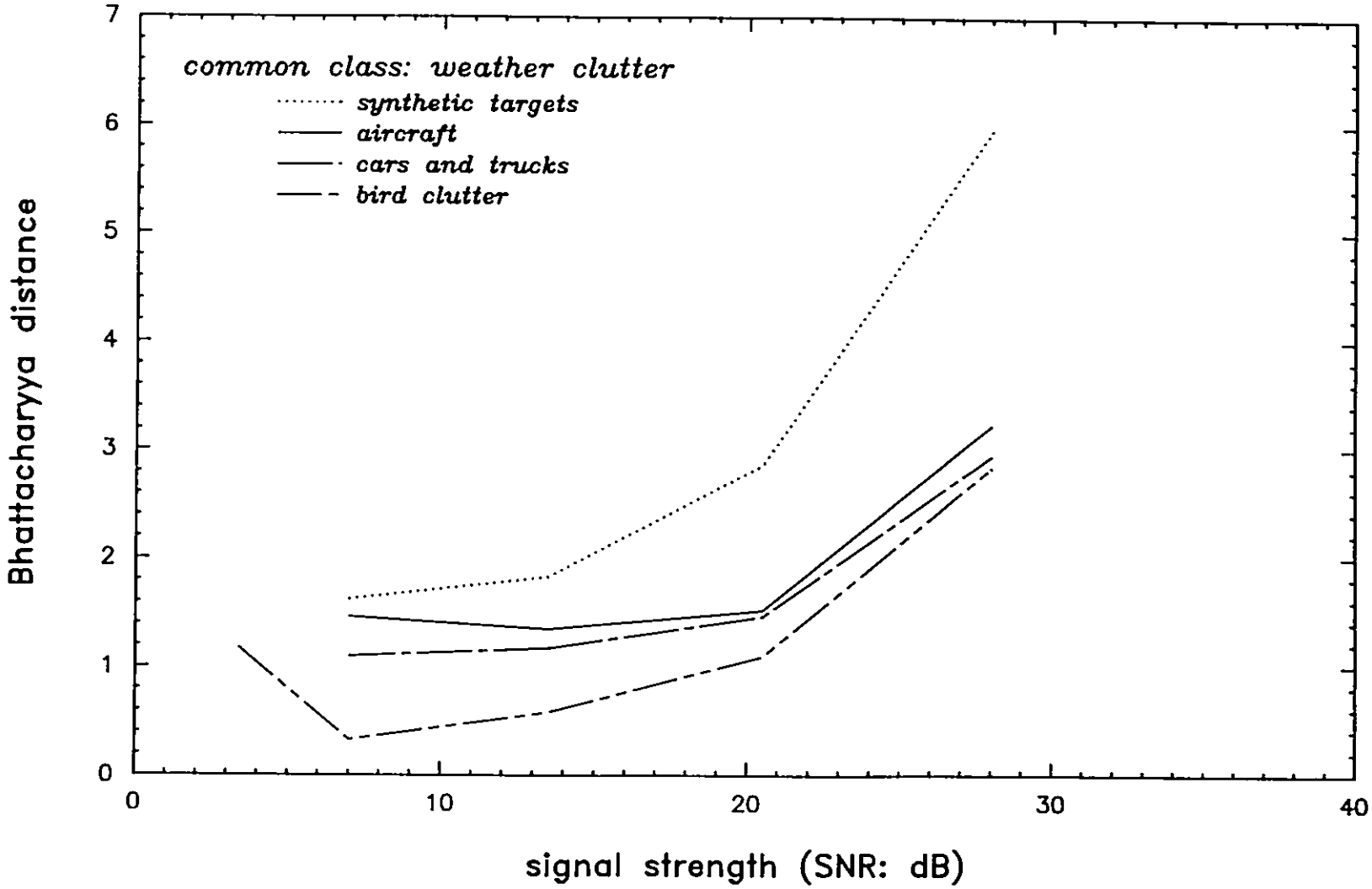


Figure 6.19d Class separabilities versus signal strength for weather clutter

6.3 Classification Experiments

The final test of any classifier is to test it against truly representative data with which it is expected to perform under operational conditions. This was done to the extent possible in a research environment. Quantitative results were obtained by testing against labelled samples, being careful to use data other than those used to train the classifier, whenever possible. This was done primarily by excluding samples from the class prototype estimates and then to test against them. In many cases, these came from different tapes or from different areas within one tape. In others, the prototype was formed from, say, every tenth sample in a labelled data set, but was tested against all of them. The results were not found to be substantially different, providing confidence that the prototypes were in fact representative of what they should be.

Qualitative, but very convincing results were obtained by the classification of complete and partial radar scans and the presentation of the classification decisions in the form of composite color overlays. Such color images can, unfortunately, not be reproduced in this thesis and individual black and white images are shown. These results demonstrate both the power and the limitations of the classifier for an operational environment. One of these limitations is the problem of underlying ground clutter, which must be treated separately to maintain control over the false alarm rate.

6.3.1 Classification Algorithm

The parametric Bayes classification algorithm was used throughout the classification experiments. Some simplifications were made, but its derivation follows the theory outlined in Section 4.1. Given the feature vector x, a discriminant function $g_i'(x)$ is computed for every defined class prototype μ_i, Σ_i. The feature vector x is assigned to the class c_j whenever

$$g'_j(x) < g'_i(x), \quad \text{all } i \neq j \tag{6.37}$$

The discriminant function used was a computationally simpler, but equivalent expression to the quadratic form of Equation (4.16):

$$g'_i(x) = (x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i) + ln|\Sigma_i| \tag{6.38}$$

Using the Cholesky decomposition, the inverse of the covariance matrix Σ_i can be written as the product of two lower triangular matrices,

$$\Sigma_i^{-1} = L_i^t L_i \tag{6.39}$$

and the discriminant function reduces to the evaluation of the inner product or squared norm of the transformed feature vector

$$g'_i(x) = \|L_i(x-\mu_i)\|^2 - 2\, ln|L_i| \tag{6.40}$$

Multiplication of this lower triangular matrix L_i with the zero-mean vector $x-\mu_i$ may be viewed as a Gram-Schmidt orthogonalization procedure [Wozencraft and Jacobs, 1965], and hence meets the requirements set out for equation (4.18). Note also that for a triangular matrix, the determinant of L_i is simply the product of its diagonal elements. During the classification experiments, of course, only the matrix multiplication and squared norm need to be evaluated: the matrix L_i and the quantity $2\, ln|L_i|$ arise from the training procedure.

6.3.2 Experimental Error Rates

The number of possible classification and testing scenarios is extremely large and cannot possibly be fully explored in this thesis. Hence, only a representative sample of results 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 amongst 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 provide 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 amongst more class prototypes will reduce the quality of each.

Table 6.1 presents classification results using samples from SNR range D. Six classes were defined and tested against over 100,000 labelled 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, 1972], and researchers in pattern recognition have long suggested the use of the so-called *leave-one-out* method of error rate estimation [see

labelled samples: / prototype name	birds (58,281)	thunderstorms (6,766)	rain (19,168)	aircraft (5,685)	cars, trucks (2,483)	targets (10,385)	total (102,768)
rejected	3.7	0.0	0.1	5.1	3.1	0.1	12.2
birds	68.3	2.9	4.6	0.5	5.4	0.1	81.7
thunderstorms	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	14.7	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

labelled samples: / prototype name	birds (58,281)	weather (25,934)	targets (8,168)	total (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

Table 6.1 Experimental 6-class classification results

The tables shows the percentage of labelled samples (columns) assigned to each class (rows). The sum of the row percentages is a measure of the bias toward that class at the expense of another. The individual class assignments are listed in the upper table. The results of combining these assignments (as outlined by the dashed lines, and 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 has been taken from SNR range D.

Kanal, 1974]. 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 [1978] 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 [1972] showed that both methods converge for large sample sizes, and for ratios of sample size to dimensionality of 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 21384 and 7447 feature vectors, respectively) has been partially compensated for by the inclusion of the synthetic target class (53999 feature vectors arising from approximately 2700 targets), and the assignment to erroneously classified samples to related classes enhances the confidence that the prototypes, as developed, are substantially correct. A perhaps 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 almost 82%, which should be considered quite satisfactory for a single scan, single resolution cell decision. Similar results for the B and C signal strength ranges yield averages of 72% and 77% respectively.

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 labelled 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, 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 for 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 6.1 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.

In selecting an optimum, or apparently optimum, classification scheme, several measures are computed. Two have already been mentioned: the average classification accuracy for the major classes, and the standard deviation of the biases. A third measure may be the percentage of samples rejected. A good set of class prototypes is one which correctly classifies most samples with few rejections and shows minimum bias toward any one cl Rejections arise from the test described in

Chapter 4, Equation (4.50). For the results presented in this chapter, the level of significance T'' was set at 1%, and the threshold T computed according to the specific combination of classes. Table 6.2 compares several possible combinations of prototypes in 3-, 4-, 5-, and 6-class scenarios. Note that the 6-class case is that shown in Table 6.1. The performance figures in the table show that the classifier is relatively insensitive to the number of classes, as long as certain combinations are used. For instance, using the rain and storm classes separately leads to worse results than keeping them combined, unless the point target classes have already been separated as in the 6-class scenario. It is somewhat surprising to see a worse performance when including the synthetic target class. There is, principally, no reason for this to be the case since all point targets are expected to have similar prototypes with only slightly larger variances than are found in the synthetic target class. The suspicion is therefore that the higher error rates are mainly due to the use of the same data for both prototype generation and testing, and that the number of samples for aircraft and cars and trucks was particularly low. As a result, the prototypes are likely to be biased toward the specific characteristics found in the samples, rather than being representative of the class as a whole.

6.3.3 The Problem with Ground Clutter

As is evident from the results presented in Section 6.3.1, there is still a relatively high number of misclassifications occurring. Radar operators are accustomed to operating with false alarm rates form 10^{-3} to 10^{-6}, depending on the type of radar used and its environment. These rates may be fixed in detection algorithms using the Neyman-Pearson test, and arbitrarily low error rates are then achieved at the expense of the probabilities of detection. As a result, the rate of not detecting the presence of a

class combination	B	C	D	E	σ_{bias}	% rejected
B TR AC	72.6	77.1	82.1	89.3	11.6	9.7
B TR ACS	72.0	76.9	81.0	87.0	7.1	12.9
B TR A	65.6	75.8	80.3	88.8	7.0	11.1
B TR S	65.5	72.9	76.9	81.9	20.6	14.6
B TR A C	72.6	77.1	82.1	89.4	14.9	8.4
B T R ACS	71.7	76.9	80.2		14.5	12.9
B T R AC	72.0	77.5	79.0		21.5	9.7
B T R S	64.6	72.1	75.5		26.8	14.5
B TR A C S	72.4	77.0	82.1	89.4	15.3	8.4
B T R A C	72.4	77.5	80.4		18.6	8.4
B T R A C S	72.2	77.4	81.8		16.6	8.4

Class prototype definitions:

B	bird clutter class only
T	thunderstorm class only
R	rain clutter class only
TR	combined rain and thunderstorms (weather class)
A	aircraft class only
C	cars and trucks only
S	synthetic targets only
AC	combined aircraft and cars and trucks (observed target class)
ACS	combined aircraft, cars and trucks, and synthetic targets (point target class)

Table 6.2 Experimental performance of various class combinations

Average classifier performance (in percent) is given according to SNR ranges. The performance measure is computed as explained in table 6.1. All samples (except synthetic targets) were always tested, including those for which no prototype was included in the class combination. The bias and rejection measures are those computed for the D range.

target (which is also an error rate) may be extremely high unless some type of clutter filtering is used. In the Bayes test, no error rate is fixed: all are jointly minimized. The consequence of this joint minimization is that no error rate is extremely high or low; all tend toward the same value, which itself is a function of the class separabilities. Typical rates (for misclassifications between major classes) range from 35 down to 2 percent: anything above that is not very usable, and rates below two percent require very separable classes. Since this rate is still high compared to radar operator expectations, the classifier cannot act as a primary detector as long as the zero-one loss function is used. Invoking real risks to reduce the error rate or rates which are considered false alarms (for example the rate at which a clutter class is misidentified as a target) necessarily biases the classifier and reduces its performance in the MAP sense.

While the widespread presence of any clutter type contributes false alarms, ground clutter is always present and presents a special problem. It is stronger than most other kinds of clutter, it is widespread, and it is similar to targets once frequency heterodyning is applied. The latter fact suggests that the Doppler frequency ϕ be added as a feature to increase the separability. Experiments using such a feature vector show that, indeed, ground can then be correctly identified with a probability of 96% to 99% (see Tables 6.3 and 6.4). As can be seen in Table 6.4, the target class will receive about 1.2% to 1.4% due to its similarity. This rate remains because the frequency for both ground and target classes must necessarily be centered at zero. The only difference is thus its variance, which is large for all classes except ground. Hence targets are not completely separable from ground clutter.

One solution is, of course, the application of risks as already discussed. Other approaches include the reduction of the amount of ground clutter entering the classifier, or to remove it completely. Reduction techniques using clutter maps such as

labelled samples: / prototype name	ground (45,298)	birds (58,281)	thunderstorms (6,766)	rain (19,168)	aircraft (5,685)	cars, trucks (2,483)	targets (10,385)
rejected	0.2	3.9	0.0	0.2	5.0	6.9	0.1
ground	96.4	1.5	9.8	0.4	5.1	0.0	0.6
birds	1.2	71.6	3.6	5.9	0.3	10.6	0.1
weather	1.0	13.6	76.5	90.1	4.3	7.4	4.9
synthetic targets	1.2	9.3	10.1	3.5	85.3	75.2	94.4
total	100.0	100.0	100.0	100.0	100 0	100.0	100.0

labelled samples: / prototype name	ground (45,298)	birds (58,231)	weather (25,934)	targets (8,168)	total (137,681)
rejected	0.2	3.9	0.1	5.6	9.8
ground	96.4	1.5	2.8	3.5	104.3
birds	1.2	71.6	5.3	3.4	81.5
weather	1.0	13.6	86.5	5.3	106.5
synthetic targets	1.2	9.3	5.2	82.2	97.9
total	100.0	100.0	100.0	100.0	300.0

Table 6.3 Experimental 5-class classification results including ground clutter

The tables shows the percentage of labelled samples (columns) assigned to each class (rows). The feature set for this case includes the doppler frequency ϕ, which has been randomized for all class prototypes except ground. The weather class prototype has been derived from a combination of both rain and thunderstorm clutter samples. The mean classification accuracy is then 84.2%, the standard deviation of the biases is 9.8%. All data has been taken from SNR range D.

given by Equations (6.34) to (6.36) have already been discussed in this chapter and are only marginally effective in this regard. They should still be used, however, to enhance classification of moving clutter on top of ground. Conventional MTI filters have been shown to be very effective in removing ground clutter; however, they also remove a considerable amount of slow moving clutter and any moving at its blind speeds. Narrowing the filter may be possible, although a minimum width is necessary due to the finite width of the clutter spectrum. Unfortunately, these filters change the spectrum of all clutter types and may therefore reduce classifier performance.

An alternate solution is to pretest the samples according to their Doppler frequency alone. This method has the benefit of being capable of removing virtually

	ground clutter samples classified as:					overall performance		
	correct	targets	birds	weather	rejected	average	σ_{bias}	% rejected
range A	98.6	1.4*	-	-	0.0	95.1	6.3	1.2
range B	97.1	1.4	0.8	0.7	0.1	77.9	17.2	2.2
range C	97.1	1.2	0.8	0.8	0.2	80.7	12.5	6.0
range D	96.4	1.2	1.2	1.0	0.2	84.2	9.8	9.8
range E	96.9	1.3	1.0	0.2	0.5	89.8	5.7	9.6
range F	97.4	1.4	0.3	-	0.9	88.6	8.6	9.2

* For Range A only one moving clutter prototype was defined including all available clutter and target samples.

Table 6.4 Experimental performance of ground clutter classification

Classifier performance (in percent) is given according to SNR ranges. The feature set includes the doppler frequency ϕ. Note the consistent percentage of samples assigned as synthetic targets. The performance measures in the right hand columns are computed as explained in table 6.1.

all ground clutter dominated samples while not affecting the classification of the other classes. It still introduces blind speeds, however, but since the test is only based on the center frequency and not the width of the clutter, the blind speed ranges may be made extremely narrow without sacrificing any of its ground clutter rejection capability. This frequency test is effectively a two-class classifier or detector, and either of the Bayes or Neyman-Pearson tests can be applied to it. For example, if a false alarm rate of 10^{-6} is desired, then the threshold must be placed at $\pm 4.76\ \sigma$ around the center frequency. The standard deviation of the ground clutter center frequency was determined to be $0.00678 f_s$; hence, the threshold must be placed at $0.0323 f_s$, or 21.2 Hz for f_s = 657.8 Hz. This corresponds to a velocity of about 4.7 knots (0.99 knots per standard deviation) and is much narrower than most MTI filters; only 6.5% of all aircraft echoes would be rejected if their Doppler frequency distribution is uniform. If a somewhat higher false alarm rate is acceptable, the rejection band can be made even tighter. Care must be taken, however, to consider the cause of apparent Doppler frequencies greater than those normally expected from ground clutter. Receiver saturation, rapid phase changes due to the changing clutter geometry as the beam scans by, and wind induced motion can all contribute to an increased false alarm rate, and some adjustment of the threshold according to operational experience is likely to be required.

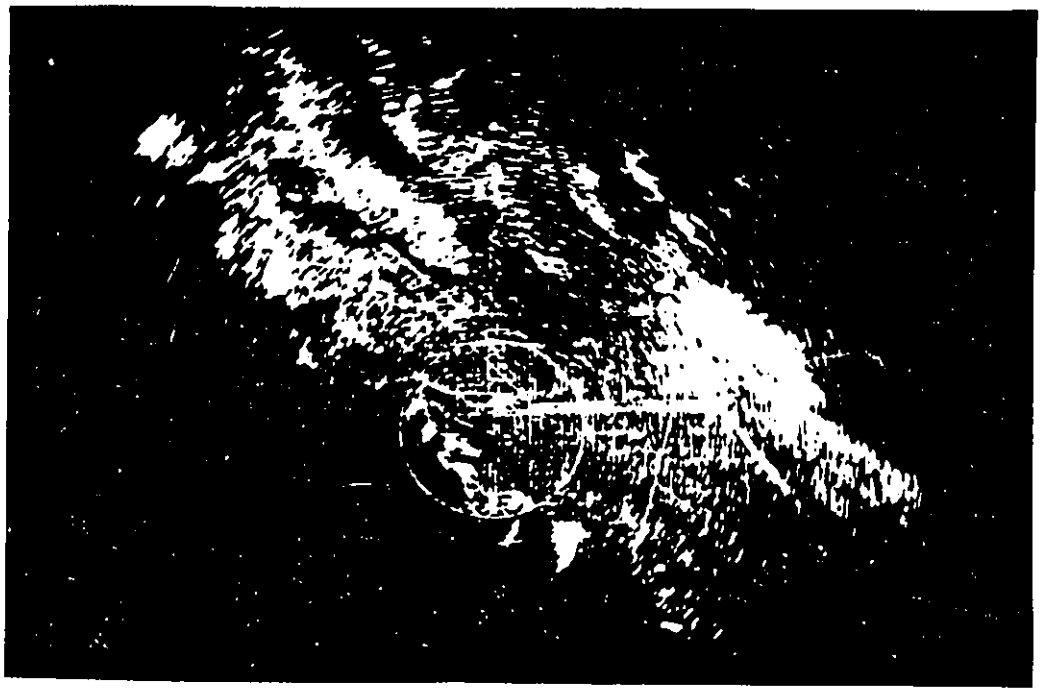
6.3.4 Classification Images

Using the same class prototypes as discussed in the previous section, complete scans of radar data were classified and simulated PPI images were color coded with the results. For the purposes of black and white presentation, individual classification results were displayed separately and photographed. Such presentation

necessarily limits the amount of information which could be displayed. For instance, the confidence limits discussed in Section 4.3 (Equation 4.47) were computed and decisions with a posteriori probabilities of less than 75% indicated as desaturated colors. Features rejected for classification and labelled "unknown" were completely desaturated, hence displayed in white. Such desaturation cannot, unfortunately, be displayed on black and white images, and was not included in the images described in this section. Nevertheless, the images do show vividly how the classification algorithm can be successfully applied.

All images except Figure 6.24 were taken from the data recorded at CFB Moose Jaw. Figure 6.20 shows the effects of classification in the presence of significant amounts of ground clutter, made visible by an atmospheric inversion which is found quite frequently in the Saskatchewan prairies. In addition to the terrain echoes and the line of hills south-west of Moose Jaw, a very prominent feature are electric transmission lines. The range rings superimposed on the image are spaced at 20 nautical miles. Other superimposed video map features include a network of major highways, outlines of the cities of Moose Jaw and Regina, three lakes, and several *permanent echoes* originating from tall communications towers. Rejection of noise and zero-Doppler ground returns results in the processed display in figure 6.20b. Note the patch of weather clutter 75 miles north-west of Moose Jaw (between 320° and 330°), which is barely visible on the unprocessed display. All other clutter returns are a combination of aircraft, birds and unknown clutter. Separation of the aircraft echoes results in the display shown in Figure 6.20c. Considering the amount of clutter present in the unprocessed data, this is an acceptable level of false alarms and scan-to-scan processing would have no difficulties isolating the aircraft amongst them.

The problem of not rejecting zero-Doppler clutter, and including ground



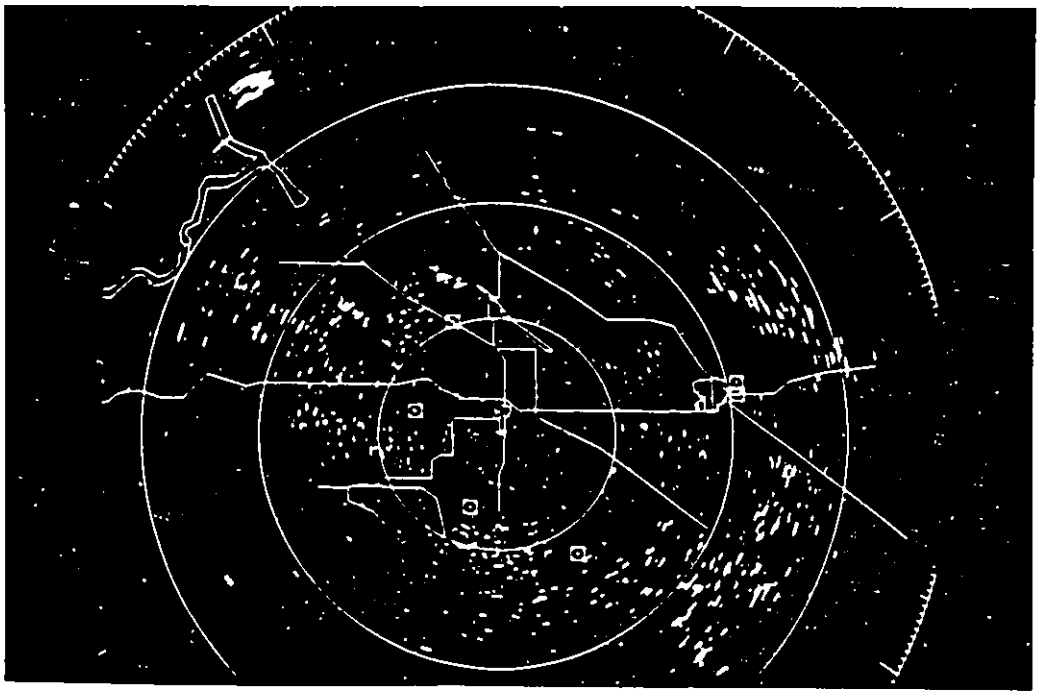


Figure 6.20a (top) Moose Jaw PPI display of unprocessed radar data

Figure 6.20b Processed display showing aircraft, bird and weather classes

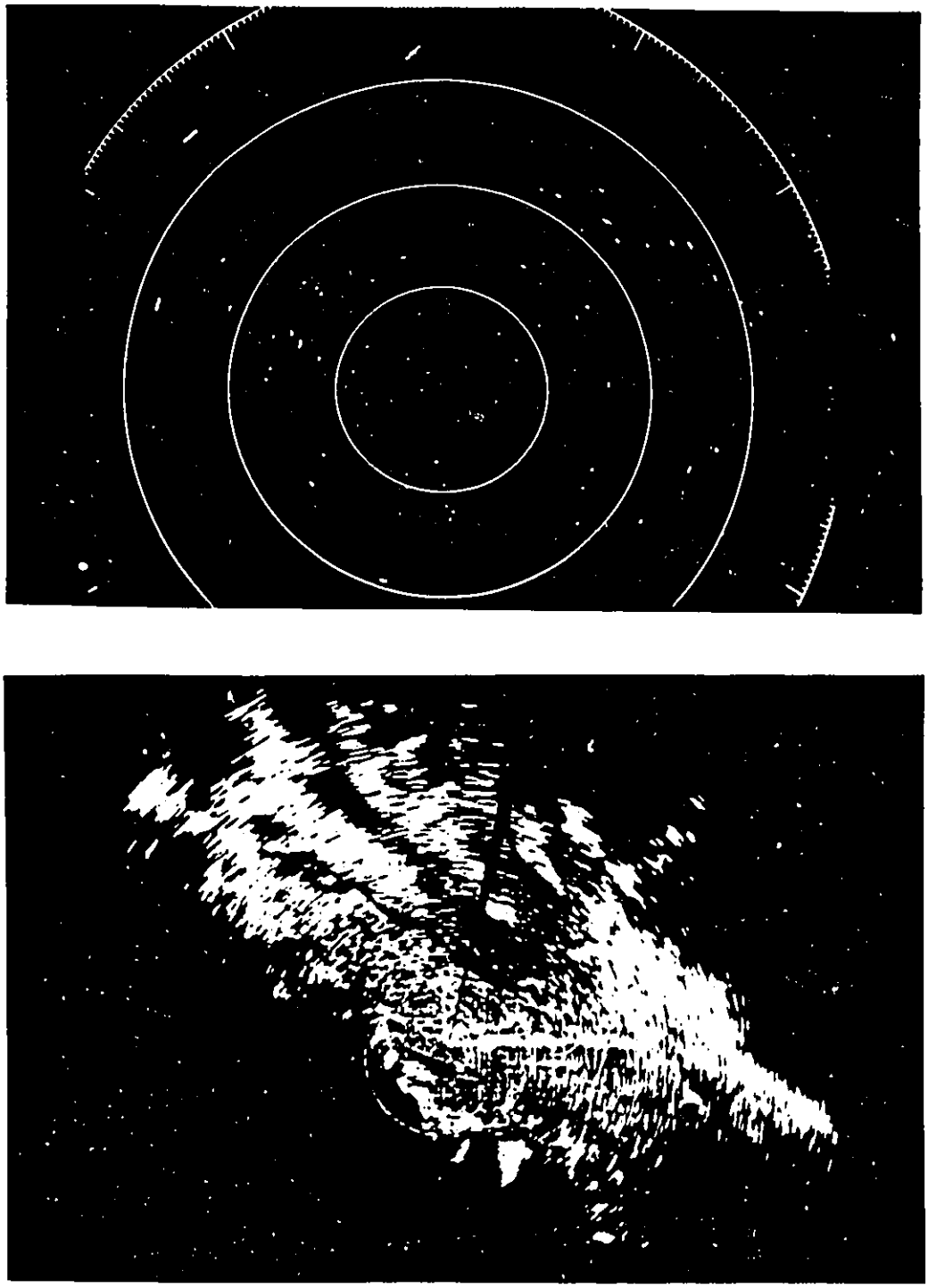
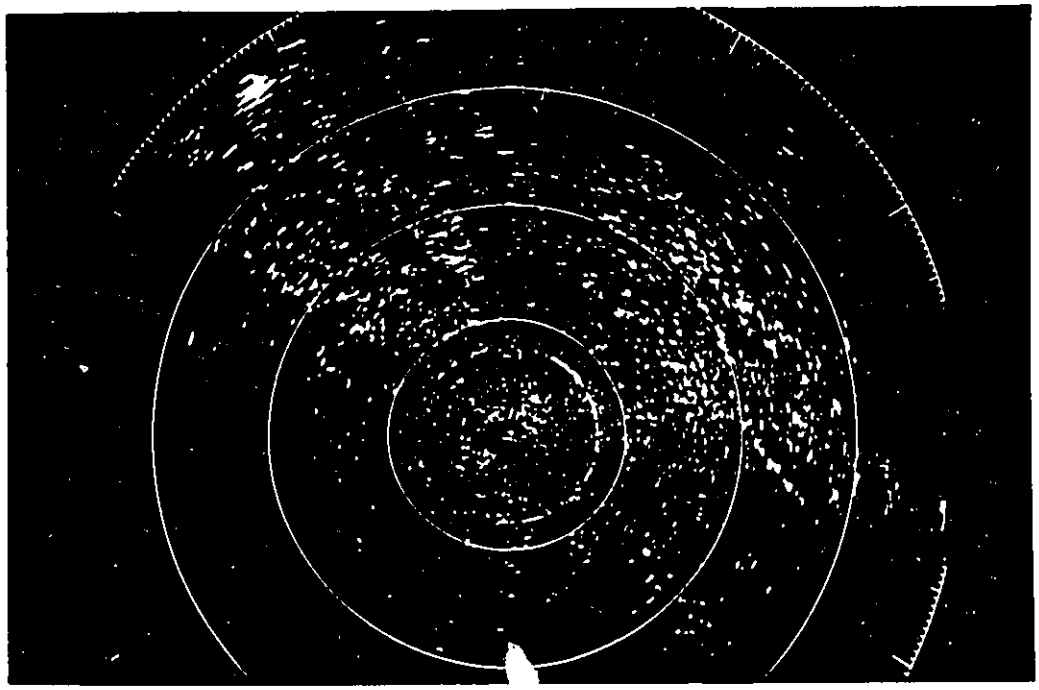


Figure 6.20c (top) Processed display showing aircraft targets only

Figure 6.20d Processed display showing ground clutter, when included as a class



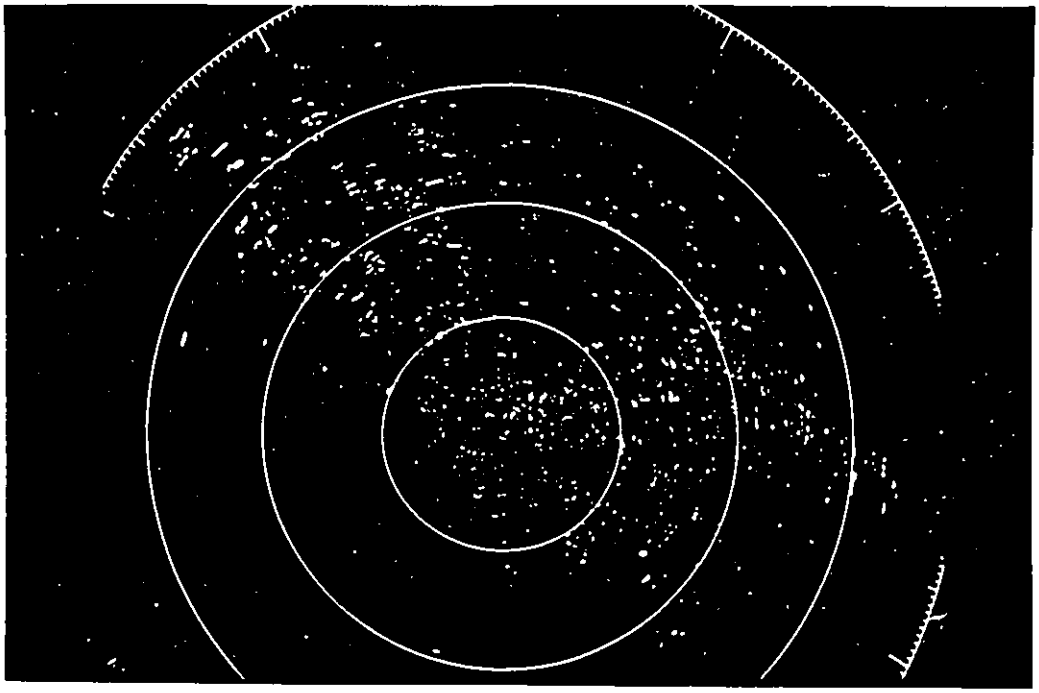


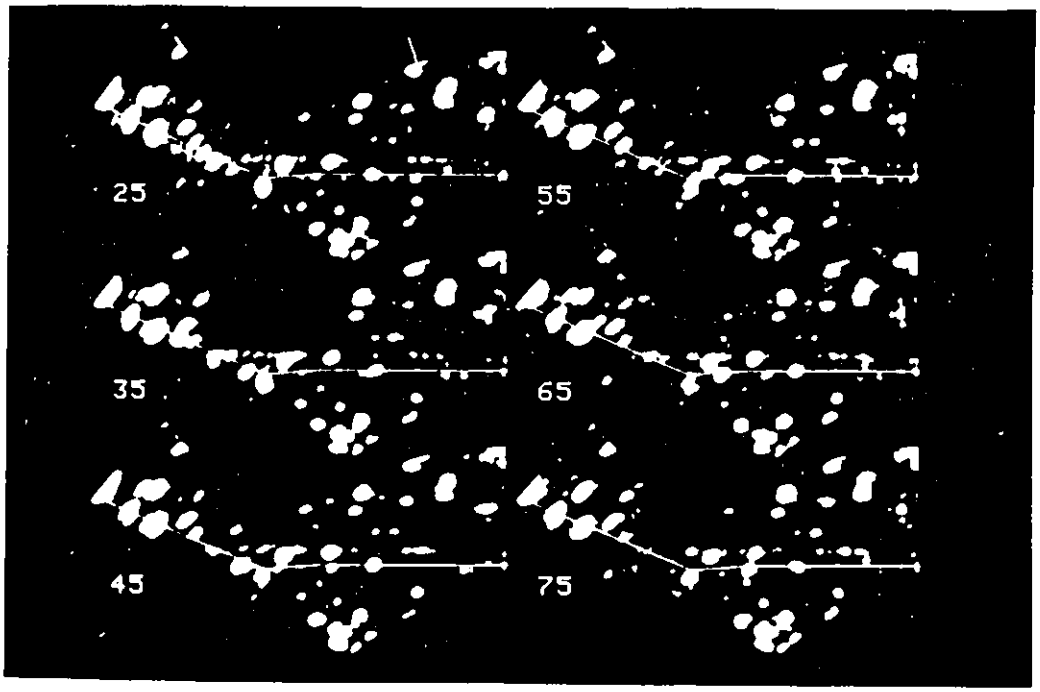
Figure 6.20e (top) Processed display of moving targets without zero-Doppler rejection

Figure 6.20f Processed display of aircraft only without zero-Doppler rejection

clutter as a class, is demonstrated in Figures 6.20d to 6.20f. Most of the ground clutter is correctly identified as can be seen in Figure 6.20d. The amount of clutter classified as "moving" is quite large (Figure 6.20e), and even after separating the aircraft targets (Figure 6.20f), the false alarm level is excessively high. For this reason, zero-Doppler rejection preceded all subsequent classification scenarios.

Figure 6.21a shows 6 scans of a 10-mile section of the Trans-Canada Highway west of Moose Jaw. The scans are separated in time by 50 seconds. The variability of the ground clutter is evident, as is the transmission line paralleling the highway. Applying the classifier clearly shows the presence of moving vehicles. While not obvious in the image, many of these echoes were in fact classified as synthetic targets with more confidence than cars and trucks. Crossover into clutter classes occurred less frequently, and only a few birds (beside the highway) were classified as targets. In fact, the data set from which this image was taken comes from a "heavy" bird clutter tape and contains a large number of bird echoes.

Bird clutter is dominant in Figures 6.22 and 6.23. Figure 6.22a shows a scan of data from another heavy bird clutter tape. This time a 25-mile range scale was employed and the video map of highways superimposed for reference. The range rings are spaced at 5 miles. The fact that bird clutter is present becomes evident only once the data is processed. The targets discovered in Figure 6.22b belong mostly to the "weak moving" class due to the small radar cross-section of birds, although 10 to 15% were in fact classified as birds. To strengthen the bird returns, STC was turned off for the image in Figure 6.23a. The presence of bird clutter is now clearly obvious, and saturation even occurs at short ranges. Nevertheless, many of these targets now fall into the bird clutter classes as shown in Figure 6.23b. The lack of bird targets in the center is mainly due to saturation and the lack of a good class prototype for bird clutter



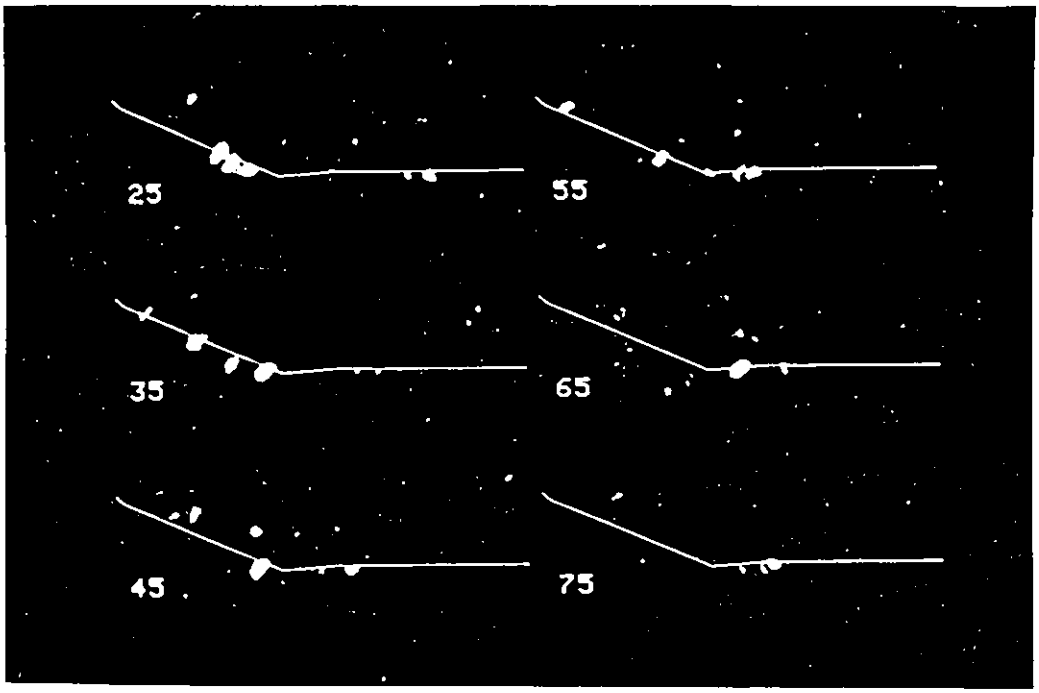
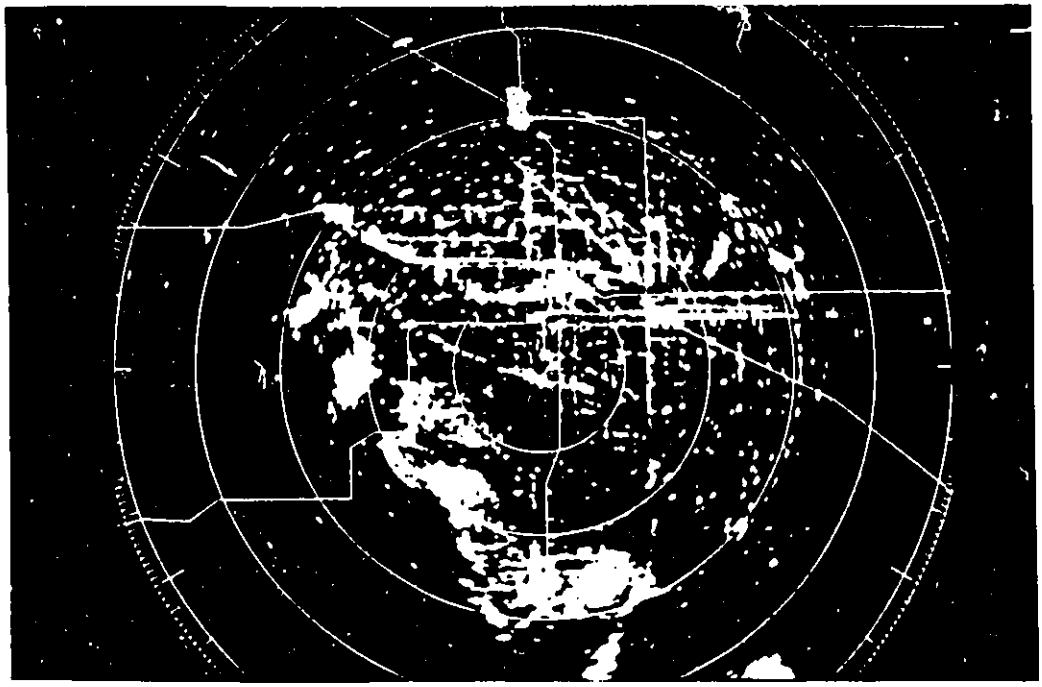


Figure 6.21a (top) Unprocessed displays of a 10-mile section of highway

Figure 6.21b Processed display showing vehicle traffic (cars and trucks)



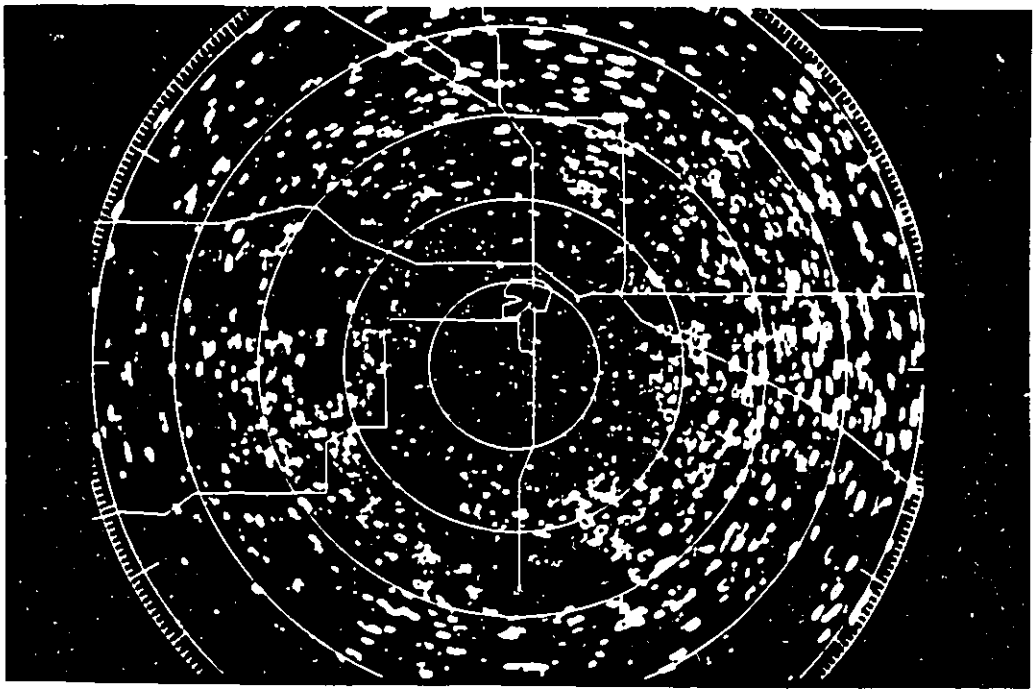
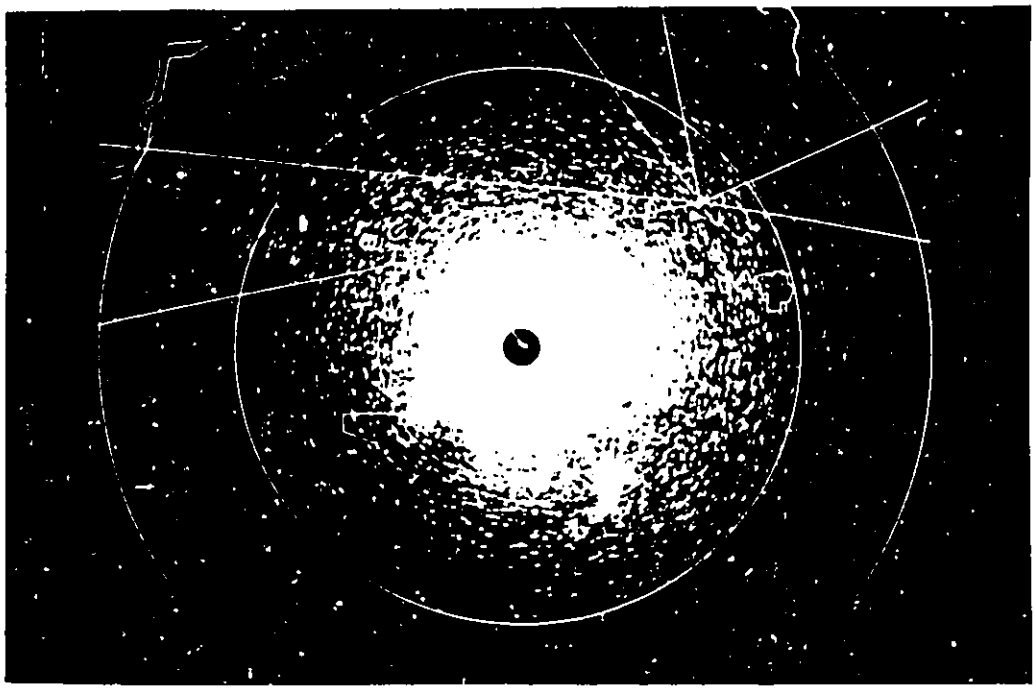


Figure 6.22a (top) Unprocessed display containing ground and bird clutter

Figure 6.22b Processed display showing bird and "weak moving" clutter



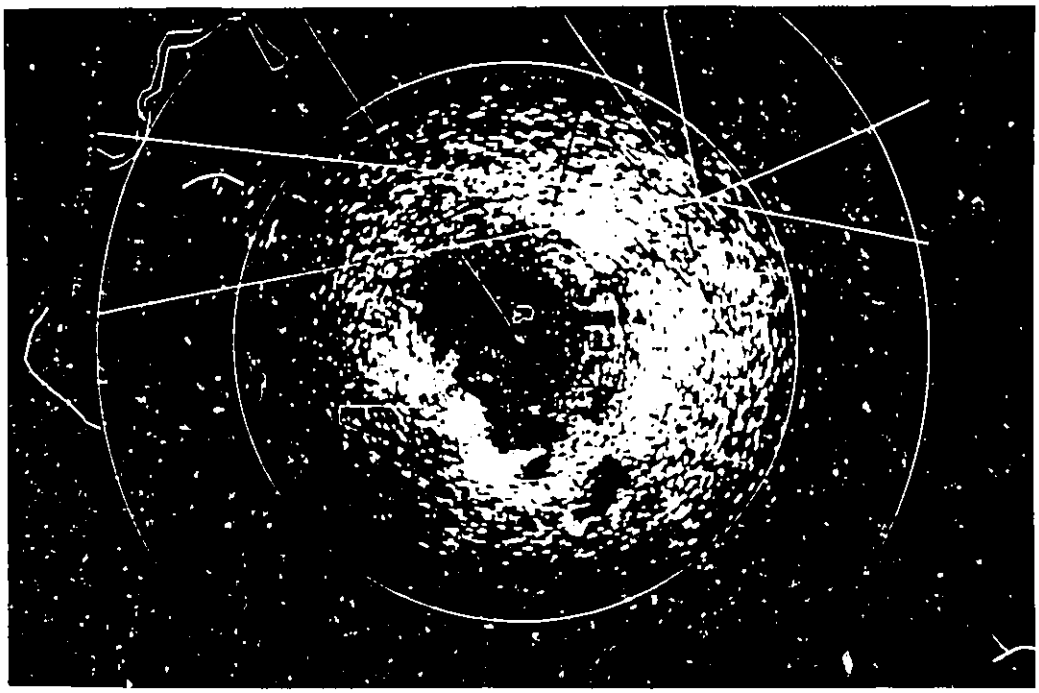
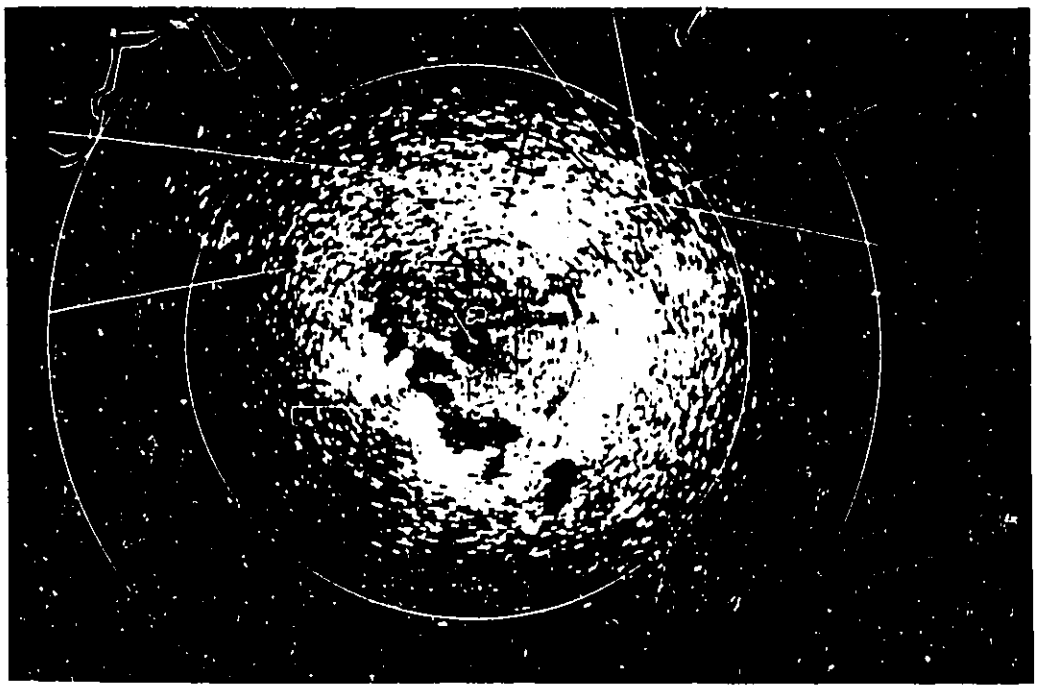


Figure 6.23a (top) Unprocessed display containing ground and bird clutter (no STC)

Figure 6.23b Processed display showing bird clutter only



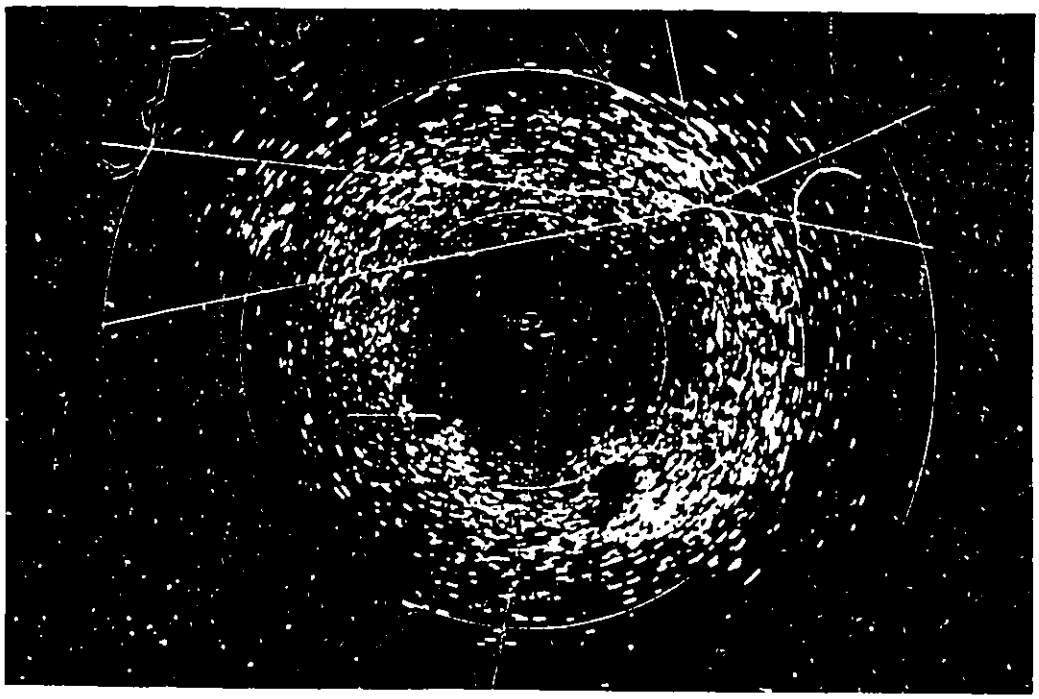


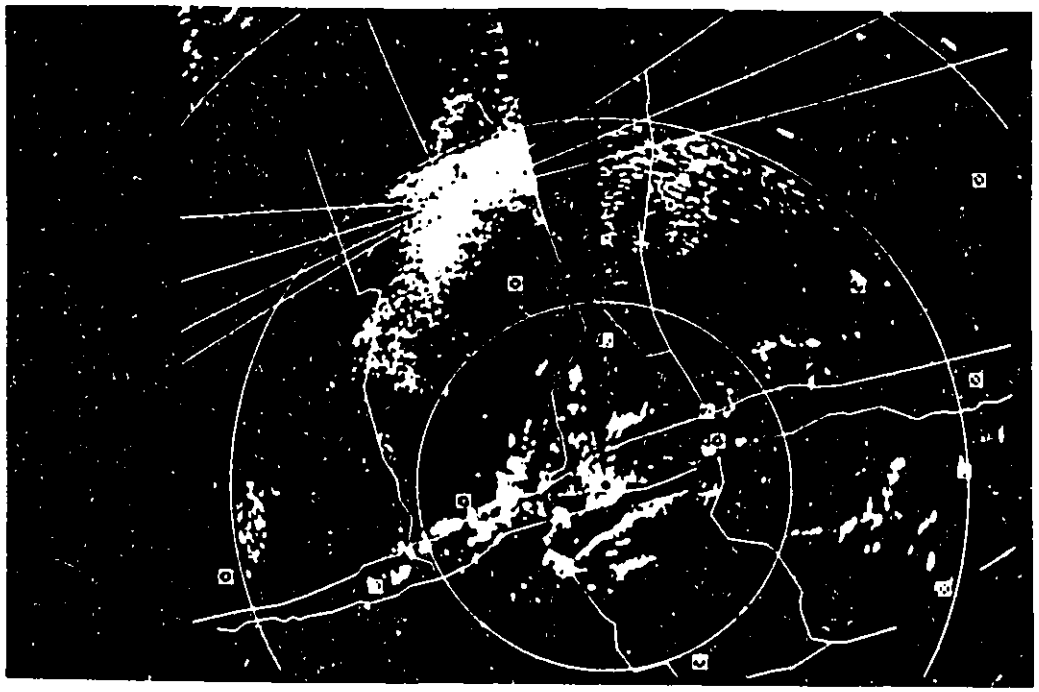
Figure 6.23c (top) Processed display showing bird and "unknown" clutter

Figure 6.23d Processed display showing misclassifications as rain and storms

in SNR-range F. Most of this short range clutter was in fact classified as "unknown", except when rejected due to zero Doppler frequency. The combined bird and "unknown" clutter classes are displayed in Figure 6.23c. Note the absence of clutter where ground clutter dominates. Such a large amount of clutter is also expected to result in a great number of misclassifications, even if the percentage is small. Clutter identified as rain and thunderstorms is shown in Figure 6.23d, although the density is much lower than that of Figure 6.23b. Note the presence of "MTI wedges" in all of the processed images. These occur when the birds move tangentially to the radar, and the direction of movement of the migrating birds can thus be established as north-westerly. The range rings in these images are spaced at 20 miles.

Areas of widespread rain clutter, recorded at CFB Trenton, are shown in Figure 6.24a. The range rings are spaced at 10 miles, and the break in clutter intensity just inside the 20-mile ring is due to beam switching from high to low. Operational restrictions prevented the use of a single beam only. The video map shows airways north of Trenton, highways, and ground based permanent echoes. High-intensity areas in Figure 6.24b are those classified as weather clutter, and low intensity areas are "weak moving" clutter which could not be identified further. Note, once again. the break in the clutter north-west of Trenton which is due to the zero-Doppler rejection filter. Note also the absence of the two targets in the north-east, which are clearly evident in Figure 6.24a. Although not shown, both were classified correctly as aircraft.

The final sequence of images shows the classification of area (as opposed to frontal) thunderstorms, recorded at CFB Moose Jaw in the late afternoon. The range limits of these two scan sections are 30 to 70 miles, with 10 mile range rings. The scan numbers shown at the bottom indicate that the difference in time is just over 3 minutes (38 scans or 188 seconds). The high-intensity areas in Figure 6.25b were correctly



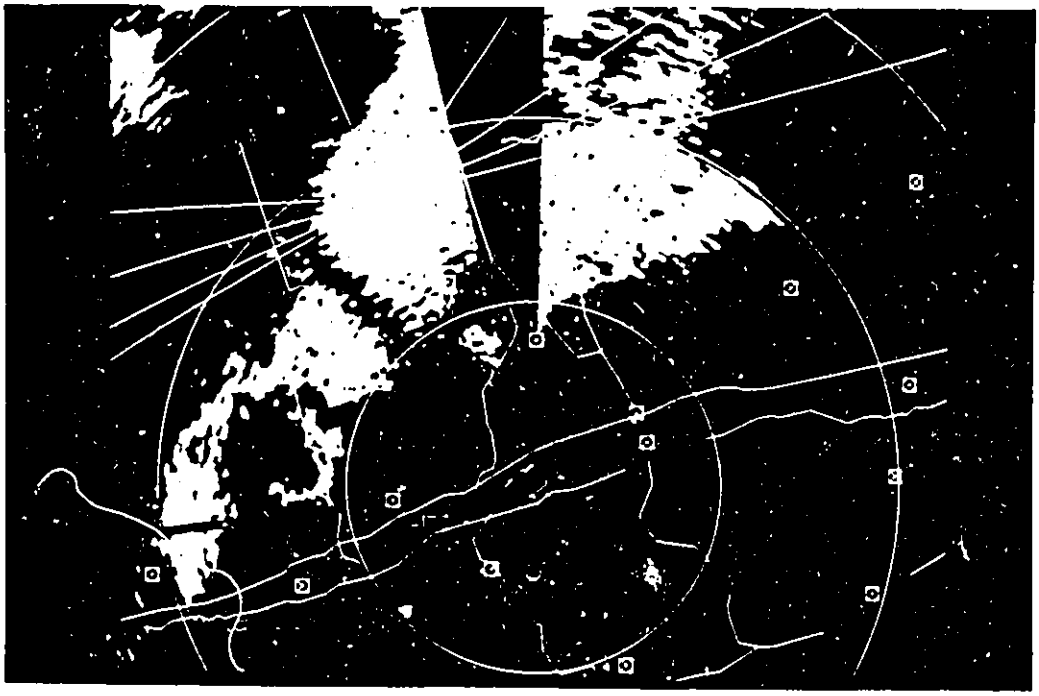
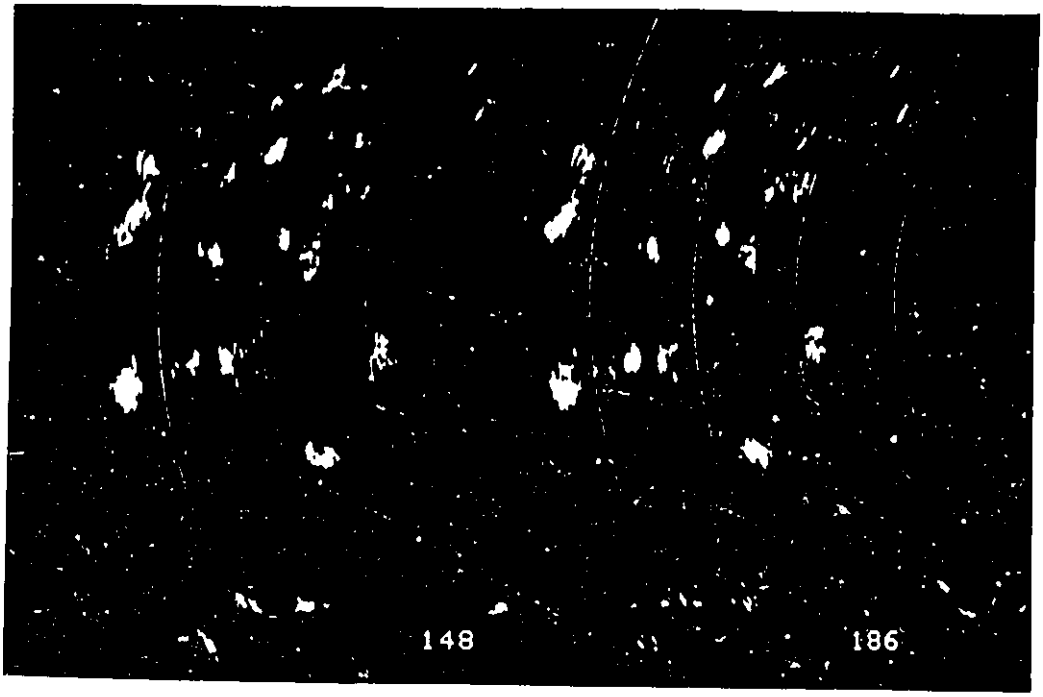


Figure 6.24a (top) Unprocessed display containing rain clutter (Trenton)

Figure 6.24b Processed display showing weather and "weak moving" clutter



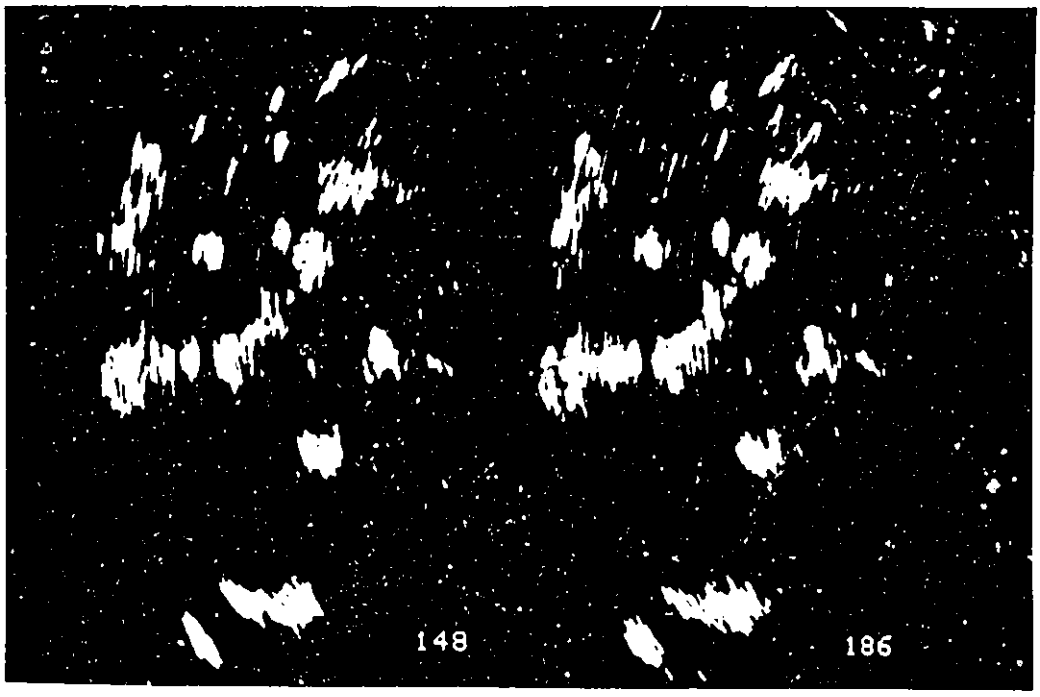


Figure 6.25a (top) Two unprocessed scans of area thunderstorms

Figure 6.25b Processed scans showing areas classified as thunderstorms and rain

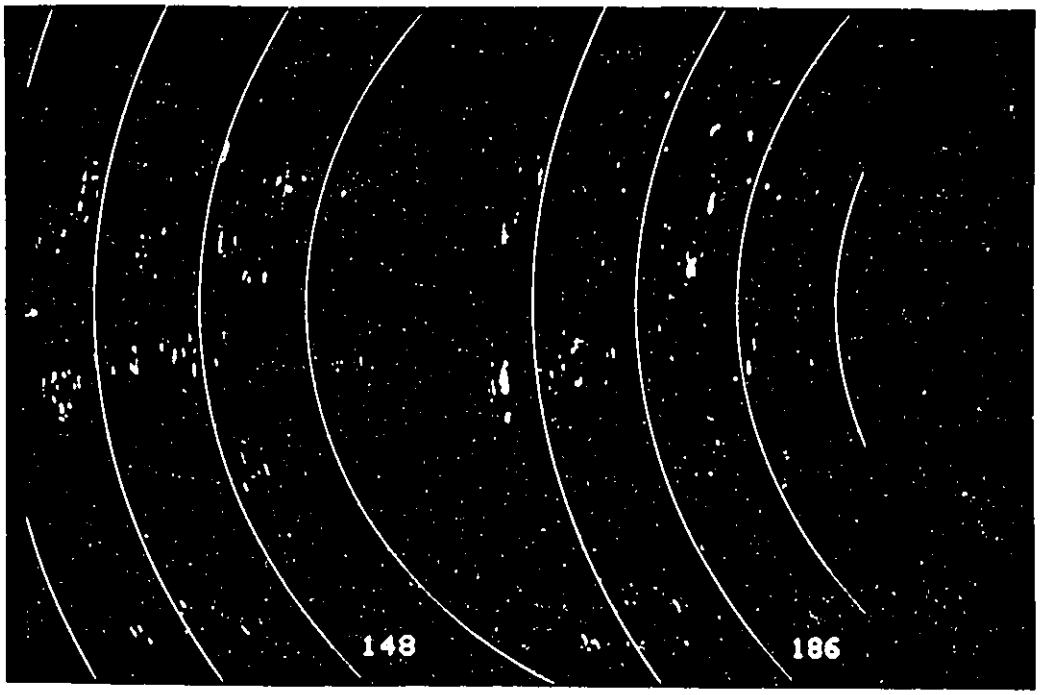


Figure 6.25c Processed scans showing areas misclassified as bird clutter

classified as thunderstorms, whereas the lower-intensity areas indicate rain clutter. Areas misclassified as bird clutter are shown in Figure 6.25c.

6.4 Summary

In this chapter, the feasibility of classifying clutter based on a feature set derived from parameters of the Doppler spectrum was examined experimentally. Radar data were specifically recorded for this purpose at two different airports using L-band air traffic control radar systems. This allowed a rather exhaustive study of the feature statistics, both marginal and multivariate, of a variety of target and clutter types. Differences between them became clear and class prototypes were constructed and tested. Several classification strategies were evaluated. Statistical experiments showed that correct classification of 70 to 90% is possible, with a mean classification accuracy of 82%. Simulated PPI images of classified radar data provided visual verification of the utility of the classifier.

The classification strategy that appears to have the most promise for real-time implementation is the following:

1. The first reflection coefficient is computed using to Equations (6.4) to (6.7) for every time series. The Doppler frequency is computed and tested for proximity to zero. Rejection occurs according to the considerations discussed in Section 6.3.3.
2. The multisegment reflection coefficients for all remaining time series are calculated using Equations (6.29) to (6.32). Equation (6.33) is used to set a rejection threshold to prevent the feature statistics from becoming too different from those used in the tightly controlled training samples. It also

serves to reject noise dominated time series which would otherwise have to be classified.

3. Compute the feature set U_0, u_0, ρ_2', ρ_3', ρ_4', ρ_5', P_{dif}, and P_{var} using Equations (6.21) to (6.28).

4. Use the SNR based feature U_0 to select a set of classifiers within one of the ranges A to F.

5. Compute the discriminant functions for each class using the remaining features as vector x and test for a minimum according to Equations (6.37) to (6.40).

6. Assign the feature vector to the "unknown clutter" class if the smallest discriminant exceeds the threshold T as per Equation (4.50)

7. Finally, evaluate the confidence in the decision by calculating the a posteriori probability of x originating from the selected class using Equation (4.47).

Chapter 7

TOWARDS REAL–TIME IMPLEMENTATION

The results presented in Chapter 6 show that radar clutter classification based on data from a single resolution cell is indeed possible. Refinements to the algorithm will undoubtedly improve the quality of the decisions both in terms of decision confidence and the experimental probabilities of correct classification. The major benefit of such a classifier – the provision of immediate identification of clutter sources to radar operators – can, however, only be obtained through its real-time implementation and use alongside existing ATC radars.

Use of the classifier places several operational constraints on the radar. Fixed PRF is absolutely required (see Chapter 3), as is minimum STC and linear polarization. Unfortunately, such a radar configuration tends to adversely affect existing detection algorithms due their to limited dynamic range and CFAR capabilities, and the design of the MTI itself. As a result it may be necessary to run the classifier for only one out of, say, ten scans. MTDs are somewhat less susceptible to this type of configuration (in fact, they require a fixed PRF themselves) and may be able to coexist better with the classifier. In addition, the detection properties of the classifier can also be exploited to some advantage.

In this chapter, the computational requirements for implementation of the classifier are discussed. No recommendations are made regarding the preferred architecture since the continuing and rapid improvements in technology, which will drive the selection of the architectures, are difficult to foresee. Instead, the discussion

focuses, in Section 7.1, on the complexities of the algorithms and its computational load. The data processing rate requirements are discussed in Section 7.2. The last section shows that floating-point arithmetic is not required except for the prototype estimators, which need not be computed in real time. The required word lengths are calculated based on the dynamic range of the data.

7.1 Computational Requirements

The number of operations executed before arriving at a classification decision is determined not only by the number of calculations required according to the formulas provided in Chapter 6, but also on the efficiency of the hardware and software used to implement them. The amount of overhead resulting from compiler inefficiencies, memory speed and availability, and the inherent limitations of the instruction set, can easily exceed 75% of the total execution time. Overhead is defined as those operations not directly required for the calculations, for example loop control, memory access, and management and storage of intermediate results. With special-purpose architectures and efficient software it is possible, however, to carry out most of the overhead functions in parallel with the main operations and thus add very little time to that required by the calculations themselves. The next two sections thus deal only with the basic calculations.

7.1.1 Feature Extractor

The main component of the feature extractor is the multisegment Burg formula together with the lattice PEF as defined by Equations (6.16), (6.17), and (6.18). The number of multiplications and additions, both complex and real, are shown in Tables 7.1 and 7.2. The expressions shown are for the general case. The values for K,

Burg order	complex multiply	complex add	real multiply	real add	real divide
1	$K(N-1)$	$K(N-1)-1$	$4K(N-1)+2$	$4K(N-1)-1$	1
2	$K(N-2)$	$K(N-2)-1$	$4K(N-2)+2$	$4K(N-2)-1$	1
3	$K(N-3)$	$K(N-3)-1$	$4K(N-3)+2$	$4K(N-3)-1$	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
total	$\frac{KM}{2}(2N-M-1)$	$\frac{KM}{2}(2N-M-1)-M$	$2KM(2N-M-1)+2M$	$2KM(2N-M-1)-M$	M
TRACS	2700	2695	10810	10795	5

Table 7.1 Calculations required for the complex multisegment Burg formula

The number of segments K, the segment (time series) length N, and the number of lattice stages M for the TRACS-ASR are 45, 15, and 5, respectively.

lattice stage	complex multiply	complex add	real multiply	real add	real divide
1	-	-	-	-	-
2	$2K(N-1)$	$2K(N-1)$	-	-	-
3	$2K(N-2)$	$2K(N-2)$	-	-	-
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
total	$K(M-1)(2N-M)$	$K(M-1)(2N-M)$	-	-	-
TRACS	4500	4500	-	-	-

Table 7.2 Calculations required for K complex M^{th}-order lattice filters

The TRACS parameters are the same as in table 7.1.

N, and M used for the results in Chapter 6 are 45, 15, and 5 respectively, and yield the number of operations listed for the TRACS. No provision has been made for the case of rejected data as discussed in Section 6.2.1 (Equations 6.29 and 6.30). Clearly, the application of such tests will increase the amount of computational overhead somewhat but, because radar data usually include large quantities of thermal noise samples, this increase is usually more than offset by the substantial reduction in the number of remaining basic calculations.

The reflection coefficients must be transformed into the feature set required by the classifier. Table 7.3 shows the number of operations required for this process. The feature U_0 is defined by Equations (6.20) and (6.25), u_0 by (6.24), ρ'_m by (6.21), and P_{dif} and P_{var} by (6.26) and (6.27). Beside the basic operations, several functions are also required for these transformations, whose complexities depend somewhat on the implementation. Frequently a combination of convergent series and look-up tables are used which execute quite rapidly to a given level of precision. The number of such operations required is small, however, and the total computational load can be seen to be dominated by the Burg formula and the lattice filter equations.

The results presented in Chapter 6 were obtained using a Floating Point Systems AP-120B array processor. This 6 MHz computer has one floating-point multiplier and one adder, both of which can run in parallel with a combined maximum throughput rate of 12 Mflops (million floating-point operations per second). The time required to compute one feature vector is thus determined by the number of multiplications or additions, whichever is greater. A 5^{th}-order lattice based feature extractor with the number of operations shown in Table 7.4 (41,086 real multiplications and 41,088 real additions) thus requires 41,088/6,000,000 = 6.85 ms to execute. A complete and full resolution radar image extending 80 nmi in range with 3340 sweeps

feature	complex multiply	real multiply	real add	real divide	square root	log	arc-tangent	sign change
ϕ	-	-	-	-	-	-	1	-
U_0	-	$2KN+1$	$2KN$	1	-	1	-	-
u_0	-	2	2	1	1	1	-	-
ρ'_m	$2(M-1)$	-	-	2	-	-	-	1
P_{dif}	-	$2K$	$2K$	2	-	-	-	-
P_{var}	-	1	K	2	K	-	-	-
total	$2(M-1)$	$2K(N+1)+4$	$2K(N+1)+K+2$	8	$K+1$	2	1	1
TRACS	8	1444	1487	8	46	2	1	1

Table 7.3 Calculations required for the reflection coefficient transformations

The feature set transformations are defined in chapter 6. The TRACS parameters are the same as in table 7.1.

	multiplications	TRACS	additions	TRACS
Burg formula	$4KM(2N-M-1)+2M$	21610	$4KM(2N-M-1)-3M$	21585
lattice filter	$4K(M-1)(2N-M)$	18000	$4K(M-1)(2N-M)$	18000
feature set	$8(M-1)+2K(N+1)+4$	1476	$2(M-1)+2K(N+1)+K+2$	1503
total	$16KMN-8KM^2-6KN+2K+10M-4$	41086	$16KMN-8KM^2-6KN+3K+M-4$	41088

Table 7.4 Total basic calculations expressed in terms of real valued operations

The equivalent number of real valued operations for the special functions (divide, square root, logarithm, arc-tangent, and sign change) have been excluded. The TRACS parameters are the same as in table 7.1.

(1280 samples each) thus requires a minimum of 8 hours and 8 minutes to be converted into feature vectors. The additional operations required by the special functions and overhead can easily increase this time to over 10 hours. Execution times of this magnitude have been observed on clutter filled radar scans where data rejection was low.

7.1.2 Classifier

The discriminant function computations given in Equation (6.40) involve no more than the multiplication of a triangular matrix with the difference between two vectors, and the evaluation of the squared norm of the resultant vector. The number of multiplications is thus equal to the number of additions of which $d(d+3)/2$ are required, where d is the number of dimensions in the feature vector. When P_0 is excluded as a feature – it is used to preselect the bank of classifiers – but the Doppler frequency ϕ is included, then $d = 2(M+1)$ and the total number of multiplications and additions becomes $(M+1)(2M+5)$. The discriminant function has to be evaluated for every class and followed by a test for the minimum. Even for ten classes, which is high if preselection is used, this number is still very small compared to the feature extractor and thus does not significantly increase the total execution time. For the full scan example given in Section 7.1.1, discriminant computation would add only 64 seconds for every class tested.

7.2 Real–Time Processing Speed Requirements

Operating the classifier in real time is essential if it is to be useful for ATC applications. The image resolution at which classification is to proceed is a major determinant toward the rate requirements, and it may well be necessary to implement

low-resolution systems in order to achieve permissible throughput rates. As may be seen in this section, high-resolution processing will mostly likely require extensively parallel and pipelined architectures.

Air traffic control radars generally use L or S-band frequencies with pulse widths ranging form 6 μs down to 0.5 μs [Brookner, 1977]. Data sampling rates are determined by the pulse width alone and thus can range from 4 to 64 samples per nautical mile. The TRACS-ASR, which was used exclusively in this research, is rather typical with its 16 samples per mile. While there is no reason to limit the classifier to these pulse widths and data rates, it is unlikely that shorter pulses will be used in air traffic surveillance radars in the near future. The following calculations will therefore be representative.

Using the TRACS-ASR again as an example, the maximum data rate is the 1.2946 MHz sampling rate. Approximately two thirds of these samples are taken inside the instrumented range of 80 nmi. The other third, taken from beyond this range, may be used to compute thermal noise levels and zero voltage offsets, but they contribute nothing to the classifier and need not enter the computation. Assuming that all 80 miles are to be classified, then the remaining average data rate is approximately 863 kHz or 842 kHz, depending on the PRF in use. The computational throughput rate must clearly equal this, unless the resolution is reduced. Based on the values in Table 7.4, the combined rate for the Burg formula and the lattice filter is then 68.35 Gflops, considerably above the processing speeds achieved to date. Introducing parallel structures, such as one computer for each of the K segments (K = 45) reduces the rate to 1.519 Gflops for each. Lowering the resolution to the 3 dB points in both range and azimuth, that is computing one feature vector for every second range sample and every seventh sweep, reduces the requirements for each computer even further,

down to 108.5 Mflops. It is important to remember, however, that these rates are maximums assuming that every cell is processed and no data is rejected. A reduction factor of two to ten can be expected under normal circumstances since clutter is rarely present in 100% of the data.

A further reduction of the required throughput rates may be realized by recognizing that instantaneous updates may not be necessary. Since clutter changes relatively slowly, it may be possible to only process, say, one scan per minute. For the TRACS-ASR this would contribute an additional processing speed reduction factor of 12. This then leaves a 9.041 Mflops rate for each of the 45 lattice filters and Burg processors. It should now be clear that, depending on the available computing power, various scenarios exist (including ones not mentioned here) which allow real-time implementation of the feature extractor.

Similar rates are needed by the classifier itself, although the number of parallel processors is much smaller. A classifier for the high resolution case requires 77.68 Mflops per class to operate in real-time, 5.548 Mflops per class for the low resolution case, and 462.4 kflops per class when only one scan in 12 is classified. These rates must of course be adjusted by the efficiency with which the discriminant function tests are executed and, as before, by data rejection rates and the required overhead. It is obvious, however, that these requirements are rather light compared to those of the feature extractor.

7.3 Finite Precision and Fixed–Point Arithmetic

Special-purpose high-speed signal processing applications are frequently implemented using integer and fixed-point arithmetic. Floating-point multipliers and adders are considerably more complex and require more computing resources. It is

therefore logical to ask what the required minimum precision is before performance will deteriorate excessively. If the use of fixed-point arithmetic is possible then considerably higher processing speeds will generally become available.

Word length requirements are determined by two factors: dynamic range of the data to represented, and maximum allowable numerical or quantization error. Both are addressed in this section.

7.3.1 Feature Extractor

The Doppler information present in the data is extracted by the lattice filter on a stage-by-stage basis until the residuals are effectively white. At this point no more extraction is possible and all higher order stages will simply pass the residuals unchanged. Assuming that sufficient data is available to ensure accurate reflection coefficient estimates, the final residuals are then representative samples of the measurement noise with power W. This implies that the dynamic range of the each filter stage (Equations 6.17 and 6.18) need not exceed the dynamic range of the data itself, except for the additional range required for numerically accurate representation of the reflection coefficients themselves. An upper bound on the magnitude of the reflection coefficients can be determined as follows. For exactly determined reflection coefficients

$$P_0 \prod_{m=1}^{\infty} (1 - |\rho_m|^2) = W \tag{7.1}$$

The coefficient magnitudes cannot exceed one, therefore

$$P_0 \prod_{m=1}^{M} (1 - |\rho_m|^2) \geq W \tag{7.2}$$

where equality holds only for a perfect M^{th}-order autoregressive fit to the data. Since $P_0 = (n_0 + 1)W$ (n_0 is the signal-to-noise ratio), a composite bound may be written as

$$\prod_{m=1}^{M} (1 - |\rho_m|^2) \geq \frac{1}{n_0 + 1} \tag{7.3}$$

If only one coefficient ρ_i takes on a nonzero value, then

$$|\rho_i| \leq \left[\frac{n_0}{n_0 + 1}\right]^{1/2} \tag{7.4}$$

Clearly this bound is conservative since, in most cases, more than one coefficient will be nonzero and Equation (7.3), if evaluated for specific values of ρ_m, will provide a tighter bound for the remaining values. In Appendix D, an upper bound for $|\rho_1|$ is given as

$$|\rho_1| \leq \frac{n_0}{n_0 + 1} \tag{7.5}$$

with equality occurring only in the single-complex-sinusoid-in-noise case. Nevertheless, for the purposes of determining dynamic range requirements, Equation (7.4) is useful and increases the word length requirements by only one bit over that of Equation (7.5).

The magnitude bound of Equation (7.4) can then be used to establish the minimum word length requirements based on the largest signal-to-noise ratio expected from the data. The TRACS-ASR word length of 9 bits plus sign leads to a dynamic range of 46 to 49 dB (see Appendix B). To fully represent a value for n_0 of 65535

(48 dB) requires a minimum of 17 bits for $|\rho_i|$ and 16 bits for $|\rho_1|$ as shown in Table 7.5. This minimum word length does not necessarily represent the signal strength with sufficient fidelity, however, to permit accurate feature extraction. The SNRs in Table 7.5 are computed from Equation 7.4 as follows

$$n_0 = \frac{|\rho_i|^2}{1 - |\rho_i|^2}, \qquad |\rho_i| = \frac{2^q - 1}{2^q} \tag{7.6}$$

where q is the word length. To determine the quantization step sizes, the SNR for a specific bit representation of $|\rho_i|$ can be written as

$$n_0(n) = \frac{|\rho_i|_n^2}{1 - |\rho_i|_n^2}, \qquad |\rho_i|_n = \frac{2^q - n}{2^q} \tag{7.7}$$

Step sizes expressed in decibels correspond to ratios of $n_0(n)$ for the desired values of n which, when q is large, may be found to approximate

$$\frac{n_0(n_1)}{n_0(n_2)} \approx \frac{n_2}{n_1} \tag{7.8}$$

This equation shows that the first quantization step size ($n_1 = 1$, $n_2 = n_1 + 1$) is 3 dB, too large to maintain fidelity. To achieve a maximum step size of 0.1 dB for an SNR of 48 dB, n_1 must equal 43 which requires another 6 bits in addition to the basic requirements of 17. The resulting 23 bit mantissa corresponds to the standard single-precision floating-point word length of most 32-bit computing architectures.

word size q bits	2^q-1 (all ones)	maximum n_0 (dB) $\lvert\rho_i\rvert$	$\lvert\rho_1\rvert$
8	255	21.05	24.07
9	511	24.07	27.08
10	1,023	27.09	30.10
11	2,047	30.10	33.11
12	4,095	33.11	36.12
13	8,191	36.12	39.13
14	16,383	39.13	42.14
15	32,767	42.14	45.15
16	65,535	45.15	48.16
17	131,071	48.16	51.18
18	262,143	51.18	54.19
19	524,287	54.19	57.20
20	1,048,575	57.20	60.21
21	2,097,151	60.21	63.22
22	4,194,303	63.22	66.23
23	8,388,607	66.23	69.24
24	16,777,215	69.24	72.25
25	33,554,431	72.25	75.26
26	67,108,863	75.26	78.27
27	134,217,727	78.27	81.28
28	268,435,455	81.28	84.29
29	536,870,911	84.29	87.30
30	1,073,741,823	87.30	90.31
31	2,147,483,647	90.31	93.32
32	4,294,967,295	93.32	96.33

Table 7.5 Maximum signal–to–noise ratios for given word lengths

The signal-to-noise ratios result from assuming equality in equations (7.4) and (7.5) and using the largest reflection coefficient representable (all ones) by the word length shown.

Maximum word-length requirements may be found by examination of the Burg formula itself. From Equation (6.17) the number of magnitude accumulations can be seen to be $K(N-1)$ in both the numerator and the denominator for the first reflection coefficient. Setting K and N equal to 45 and 15, respectively, allows for a maximum growth of 10 bits over the word length of the squared magnitudes of the data values. Since squaring doubles the length of the mantissa, a 9-bit data word length thus leads to a maximum of 28 bits. In the strictest sense, the word length of subsequent stages would grow exponentially if the residuals are to be computed to the fullest possible precision. The size of the estimation error makes such growth rather pointless, however, and the 28-bit mantissa length can safely be maintained for the high-order coefficients. In fact, the AP-120B array processor used in this research has a 27-bit mantissa plus sign which lead to quite acceptable results.

The tradeoff in choosing between the minimum and maximum word lengths is simply cost versus the preservation of significance in the data. Here, the term "significance" takes on a meaning slightly different from that conventionally used in mathematics. What is most important is the performance of the classifier, and the actual computed values are only of consequence in how they affect the probability of correct classification; they need not be accurate per se. Thus the statistical distribution of the values is of greater consequence and, ideally, all values will occur with equal frequency in order not to degrade the powers of discrimination in the feature. Maintaining significance is therefore synonymous with maintaining accurate classification, which in turns requires small quantization step sizes in order to limit the number of occurrences of each possible value. Without transformation, the first reflection coefficient has exactly the wrong characteristics: it tends to be clustered close to magnitude 1 where, in terms of the equivalent signal strength, quantization step size

increases rapidly as unity is approached. The normality transformations discussed in Chapter 5 help in this regard, but it is just as important not to loose the significance *before* it is applied. Markel and Gray [1976] examine several possible transformations to equalize the distributions, and of these some are very similar to the one presented in this thesis. Their aim, however, is the coding and transmission of the predictor coefficients, not the calculation itself which may proceed at greater levels of precision.

So far only the mantissa requirements have been addressed. Considerations for exponent size are somewhat different and deal more with varying scaling requirements. Floating-point representation normalizes the mantissa so that all bits are always available to preserve significance. Hence if such normalization is not required, or if the significance is not present to begin with, then little is gained through its use. Two observations may be made in reference to the Burg formula and the lattice PEF. First, significant reflection coefficient magnitudes are those close to unity, and normalization does not take place unless values fall below 0.5. Any loss of significance occurring as the coefficient approaches unity thus cannot be avoided by conventional normalization (that is the elimination of leading zeros). Second, the filter operates by subtracting data values of similar magnitude whose significance lies only in those bits already available before normalization. Normalizing the data values thus only serves to minimize numerical errors, but these are generally well below the measurement errors and estimation uncertainties, even without normalization. It may therefore be concluded that floating-point representation is not essential, although it does, undeniably, improve the numerical performance of the feature extractor somewhat.

Using fixed-point arithmetic does, however, place stringent scaling requirements on the hardware designer. Overflow and underflow are rare in floating-

point arithmetic with long exponent words, but they are more likely to occur when the exponent is absent. Knowledge of the largest and smallest meaningful values is imperative and must be included in the word length considerations.

The loss of precision associated with increasing signal strength, as the reflection coefficients approach unity, may be avoided in the case of the feature u_0 if $|\rho_1|$ is not calculated directly. Recall that

$$u_0 = 10\,log\left[\frac{r}{1-r}\right], \qquad r = |\rho_1| = \frac{|R(1)|}{R(0)} \tag{7.9}$$

thus

$$u_0 = 10\,log\left[\frac{|R(1)|}{R(0) - |R(1)|}\right]$$

$$= 10\,log|R(1)| - 10\,log[\,R(0) - |R(1)|\,] \tag{7.10}$$

Precision is now limited to that with which $R(0)$ and $|R(1)|$ can be determined. The quantity $[\,R(0) - |R(1)|\,]$ is always greater than or equal to the measurement noise power W, hence the subtraction operation does not introduce an additional loss of significance. Note that this does not avoid the computation of ρ_1 completely, since it is still required for the computation of the high-order coefficients and frequency heterodyning (the features ρ'_m).

Some experimental work regarding the minimum number of bits required to represent the reflection coefficients may be found in [Markel and Gray, 1976]. Although the authors deal only with real valued speech signals, some insight may be gained from studying their approach. They do not relate their results to SNR since they work purely with autoregressive models, but they do examine the sensitivity of the

information in the coefficients to the number of bits used to represent them.

7.3.2 Classifier

The precision requirements for the classifier are somewhat easier to define. The multiplication of the feature vector by a predetermined triangular transformation matrix leads to another vector whose squared norm (sum of squares) is then computed. The result is the squared Mahalanobis distance which is approximately χ^2_d distributed, with mean d and standard deviation $\sqrt{2d}$. For $d = 12$, over 99% of the values will be found between 0 and 27 which implies that the dynamic range of the result is quite small and independent of the data. Added to this squared distance is the logarithm of the determinant of the autocovariance matrix. This value is a function of the underlying distribution of the feature set and is therefore also independent of the dynamic range of the data. It can, however, take on values within a somewhat greater range than those of the Mahalanobis distance. Word length requirements are therefore strictly a function of the quantization limits imposed by the inherent discrimination powers of the feature vector itself.

The matrix multiplication is essentially a feature vector transformation to zero mean and unit variance. The bounds of unity on the reflection coefficient magnitudes are thus removed and minimum word lengths should increase by at least one or two bits to accommodate this transformation and requantization. The next step is squaring and this does, strictly speaking, double the word length requirements. But since this is the last operation other than the summing of the squares, and no square root is subsequently taken, such doubling is not necessary. The important considerations now are the statistical distribution of the data and the final quantization step sizes which must be small enough to maintain the powers of discrimination. For

example, the maximum density of the χ^2_d density with $d = 12$ is approximately 0.082, which implies that if only one out of, say, 10,000 data samples is allowed to yield the same squared Mahalanobis distance, then the quantization step size must not exceed 1/820. A 10 bit fraction will meet this requirement. Another 7 bits plus sign are necessary to represent the integer portion, defining a combined word length of 17 bits plus sign. Again, as for the feature extractor, standard single-precision floating-point arithmetic should work acceptably. Fixed-point arithmetic should also work well since the significant dynamic range of the result is not large. It may be necessary, however, to lengthen the word sizes for the matrix multiplication, since the individual products before summation may in fact exceed the range of the result considerably.

Fixed-point arithmetic will not likely work well for the prototype estimator and the matrix inversion algorithm. In fact, since the formation of the autocovariance matrix requires squaring, and the subsequent Cholesky decomposition computes the square root, double-precision floating-point arithmetic is indicated, especially if large numbers of labeled samples are to be included in the estimate. Fortunately, the estimation need not proceed in real time, and the use of double precision should not be a problem. The final triangular matrix which represents the prototype is then adjusted to the precision used by the classifier.

7.4 Summary

The calculations carried out in this chapter clearly show that, given sufficient computing resources, the feature extractor and classifier can be implemented for real-time operation. Several tradeoffs may be made to limit required throughput rates to specific values. Using the TRACS-ASR with its 850 kHz average data rate, a net rate (not including computational overhead) of 70 Gflops is required for full

resolution classification for every scan. Significant rate reductions may be achieved using a combination of lowering of resolution and the computation of only one scan per minute. Reductions of two orders of magnitude were discussed in this Chapter, but further reductions are possible using parallel structures.

Fixed-point arithmetic can usually be executed faster than floating point, and is possible for both feature extractor and classifier if certain minimum word length requirements are met. These requirements depend largely on the dynamic range of the data. In general, floating-point mantissa lengths of 23 bits (standard single precision) should work acceptably, although fixed-point word lengths of 16 bits will be barely sufficient for the TRACS dynamic range of almost 50 dB. 32-bit fixed-point arithmetic should, however, work well with proper binary point placement.

Chapter 8

CONCLUSIONS AND CONTRIBUTIONS

The research described in this thesis clearly shows that single-scan classification of radar clutter into broad classes such as birds, weather, and targets is indeed possible, and can be achieved with a mean classification accuracy of approximately 80% using features derived from the Doppler spectrum. Classification into subclasses such as rain and thunderstorms, or aircraft and ground-based vehicles is also possible, albeit with a lower success rate. It is important to note here, that these rates are likely fairly conservative, and improvements using additional features (not necessarily arising from the Doppler spectrum) should be possible. This thesis describes many of the considerations and tests which must be applied to all such features before they can be accepted as discriminants.

After an introduction to the classification problem in Chapters 1 and 2, the physical characteristics of the clutter and how they affect the Doppler spectrum were described. The theory of how statistical classification must proceed is summarized in Chapter 4, and in Chapter 5 the considerations behind feature selection and extraction were discussed. The selection of separable features is the key to successful classification, as is the knowledge of the statistics of these features. Improper matching of these statistics to those assumed in the derivation of the classifier can lead to poor and even misleading classification decisions.

Many classifiers have been proposed, but few have found wide spread application in practice. Without extensive testing on real data, this work would have

been little more than an academic exercise. An important mandate for this research was to develop a comprehensive radar data base, analyze the data, and with it evaluate the classifier. The results of this work were detailed in Chapter 6. In the final chapter, the proposed algorithm was analyzed in terms of its computational and numerical requirements.

There are several contributions offered by this thesis. First, the data base collected is, perhaps, one of the most extensive and comprehensive of its type. The recordings are totally digital and thus offer the full radar dynamic range to the researcher, which was not possible previously using analog video recorders. Additional benefits of the digital recordings are the retention of full coherence, noise levels, scan indexing, timing and precise position information (range and bearing). As a result, the confidence in ground truth and clutter labelling can be very high indeed. The volume of data offered and its variety of clutter types allowed unprecedented levels of classifier testing.

The second contribution is undoubtedly the in-depth research into the characteristics and statistics, both marginal and multivariate, of the clutter features. A significant portion of this thesis is devoted to this analysis, along with attempts to match the actually encountered, highly non-normal and nonlinear statistics to the requirements of the parametric Bayes classifier. Theory from several disciplines was brought together for this purpose. Even if the solution is not optimal, and no claim is made to that effect, it appears practical and certainly points the way for future endeavors in this field. Again, due to the large volume of data analyzed, confidence in the results is high.

Finally, the classification experiments carried out were extensive in their depth and breadth and, again due to the large volume of data available, confidence in

the results is high. Considerable care was taken to ensure the correctness of the class labels. Approximately one hundred color-coded images were generated to verify the success of the classifier with real data and in terms of a potential form of presentation to the operator. Unfortunately, only a small number of black and white sample images could be reproduced in this thesis.

This thesis also offers several minor contributions. These are the results of analyses performed in support of the research itself. They are discussed in the appendices and include the expected value of the complex reflection coefficients and their sensitivity to quadrature errors. A method for finding correction factors to compensate for these errors has also been developed. This method turned out to be essential for data preprocessing, before proceeding to feature extraction and classification. Another contribution is the relatively detailed study of the range performance of the TRACS-ASR. This knowledge was important to establish the expected clutter levels and to assist with the identification of the clutter sources themselves.

Appendix A

TRACS–ASR SPECIFICATIONS

The radar data used throughout this research was 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 Department of National Defence (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 secondary surveillance radar (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 nautical miles, 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	22.306 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 and accuracy	1/16 nmi (116 m)	
azimuth sampling interval	0.088°	
azimuth accuracy	within 0.264°	

* see Section B.2

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

TRACS–ASR DETECTION PERFORMANCE

B.1 Range Performance

The detection performance of a radar can be computed by the application of the radar equation

$$D_r = \frac{P_t \tau G_t G_r \sigma \lambda^2 F_t^2 F_r^2}{(4\pi)^3 k T_s L R^4 D_0} \qquad \text{(B.1)}$$

where D_r is the margin of the received signal above what is required for detection, expressed as a signal-to-noise ratio. The required SNR for detection is given by D_0. For the TRACS-ASR, the parameters used in (B.1) are as follows (see also Appendix A)

P_t	pulse power to antenna	1.5 MW
τ	pulse width	2 μs
G_t	antenna transmit gain	33 dB
G_r	antenna receive gain (low beam)	33 dB
σ	radar cross-section	as required
λ	wavelength (Moose Jaw)	22.306 cm
F_t	pattern and propagation factor, transmit	as required
F_r	pattern and propagation factor, receive	as required

k	Boltzmann's constant	1.38×10^{-23} J/K
T_s	system noise temperature	707 K
L	transmit feedline and atmospheric losses	1.1 dB + 0.02 dB/nmi
R	target range (slant range)	as required

The required detection margin D_0 varies with the number of hits per beamwidth, the detection requirements, and the type of detection scheme used by the radar signal processor. For instance, to detect a non-fluctuating target using a single pulse with a probability (P_D) of 0.5 and a false alarm probability (P_{FA}) of 10^{-6}, a signal-to-noise ratio of 11.2 dB is required. The TRACS radar uses two coherent digital MTI processors (one each for the I and Q channels), then recombines the two channels and forms the logarithm. A CFAR processor (8 range sample mean estimator) equalizes the signal levels before pulse-to-pulse integration. The integrator uses a 7/8 feedback constant, which is close to optimum for the 14 hits per beamwidth expected. Such a non-coherent integration log-CFAR detection scheme achieves a processing gain of just under 6 dB, and thus only 5.3 dB are required for the same detection parameters. The recorded data was, however, not restricted to using this suboptimal detector and fully coherent processing could be applied. A fast fourier transform (FFT) with data weighting matched to the beamshape is in fact an optimum coherent integrator and meets the above detection requirements on 1.7 dB SNR targets (integration gain of 9.5 dB). Increasing the allowable false alarm probability to 10^{-4} reduces the required SNR to just under 0 dB. A thorough discussion on radar detection performance may be found in [Blake, 1980].

In light of the above discussion, D_0 is set to 0 dB in the plot of detection margin D_r versus range R (Figure B.1). In this plot the propagation factors are also set

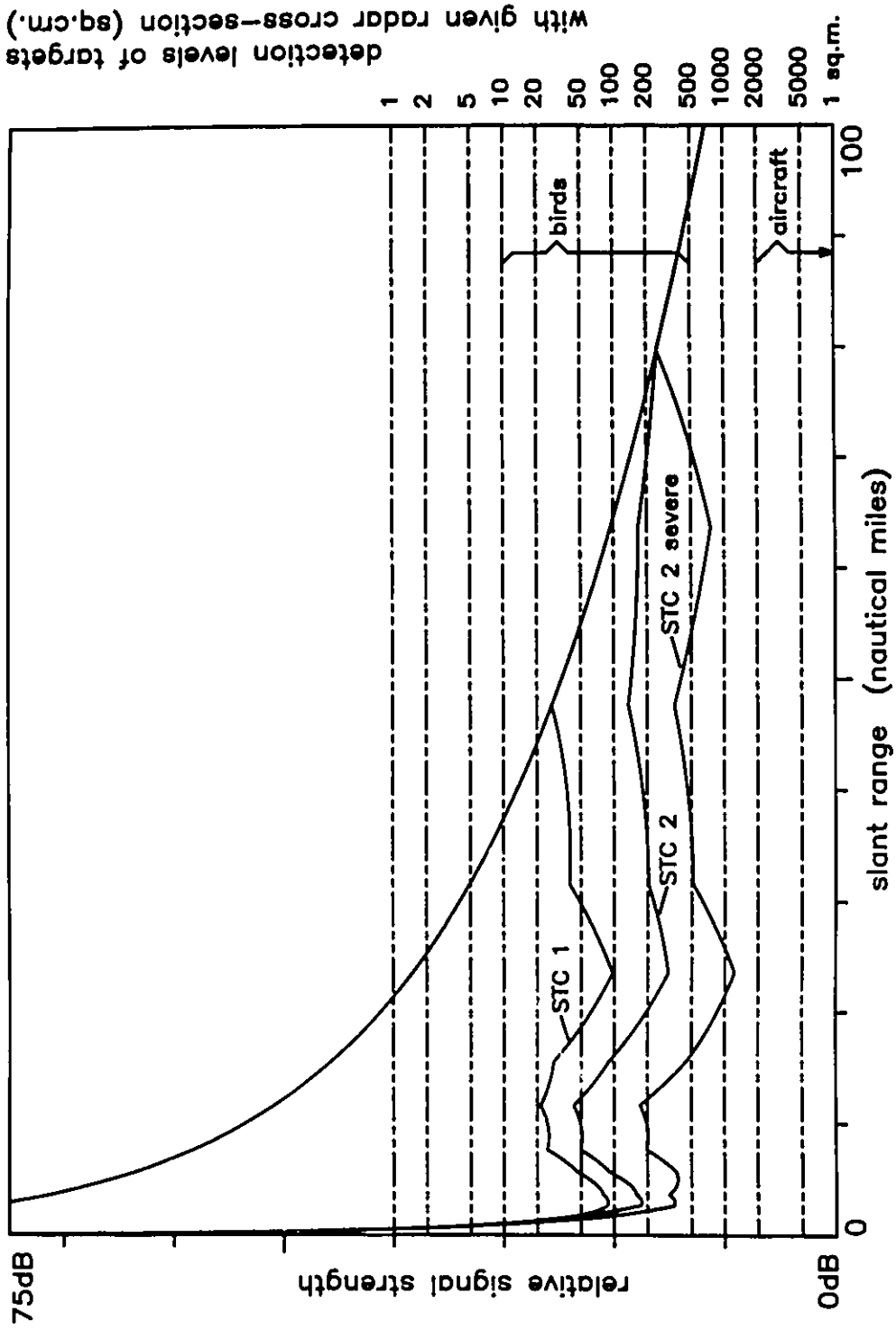


Figure B.1 TRACS–ASR detection performance at the beam center

The curves show the detection margin for targets with radar cross-sections of 1 m^2 or less.

to 1, which corresponds to the case of the target always being positioned at the center of the beam. Multipath effects over rough terrain are quite difficult to model and are ignored. The margin is referenced to a 1 m^2 target, and the margins for other target sizes can be found by shifting the base line up to the levels shown on the right hand scale of the plot. The approximate RCS ranges for birds and aircraft are indicated. The effects of applying *sensitivity time control* (STC) is shown as well. The radar allows the selection of two preprogrammed curves, and another curve is available for manual entry should the need arise. The specific curves shown are those in use at CFB Moose Jaw. It should be noted that the different polarizations operational at the two radar sites has some effect on bird detectability. Vaughn [1985] noted that some reflectivity measurements using horizontal polarization were as much as 4 dB stronger than those using vertical polarization.

The radar clearly has the capability to detect bird clutter, even with STC turned on, although small birds and bird flocks with radar cross-sections of less than 50 cm^2 will likely be missed. The "anti-bird" curve (labelled *STC 2 severe*) is clearly positioned to eliminate most but the strongest bird clutter. The dip around 24 nautical miles is notable: it arises from the site adaptation process which attempts to evenly distribute the false alarm rate over the entire surveillance area. The radar is primarily used for terminal control, and the Moose Jaw Military Terminal Control Area (MTCA) extends mostly over an area south of the base where several hills are located. The rather strong ground clutter returns originating from these hills are partially controlled with the extra attenuation.

Figure B.1 represents the best that can be expected from the radar. Unfortunately targets are rarely found exactly in the center of the beam, and the beam pattern factors F_t and F_r must be applied for a more realistic picture. Figure B.2 has

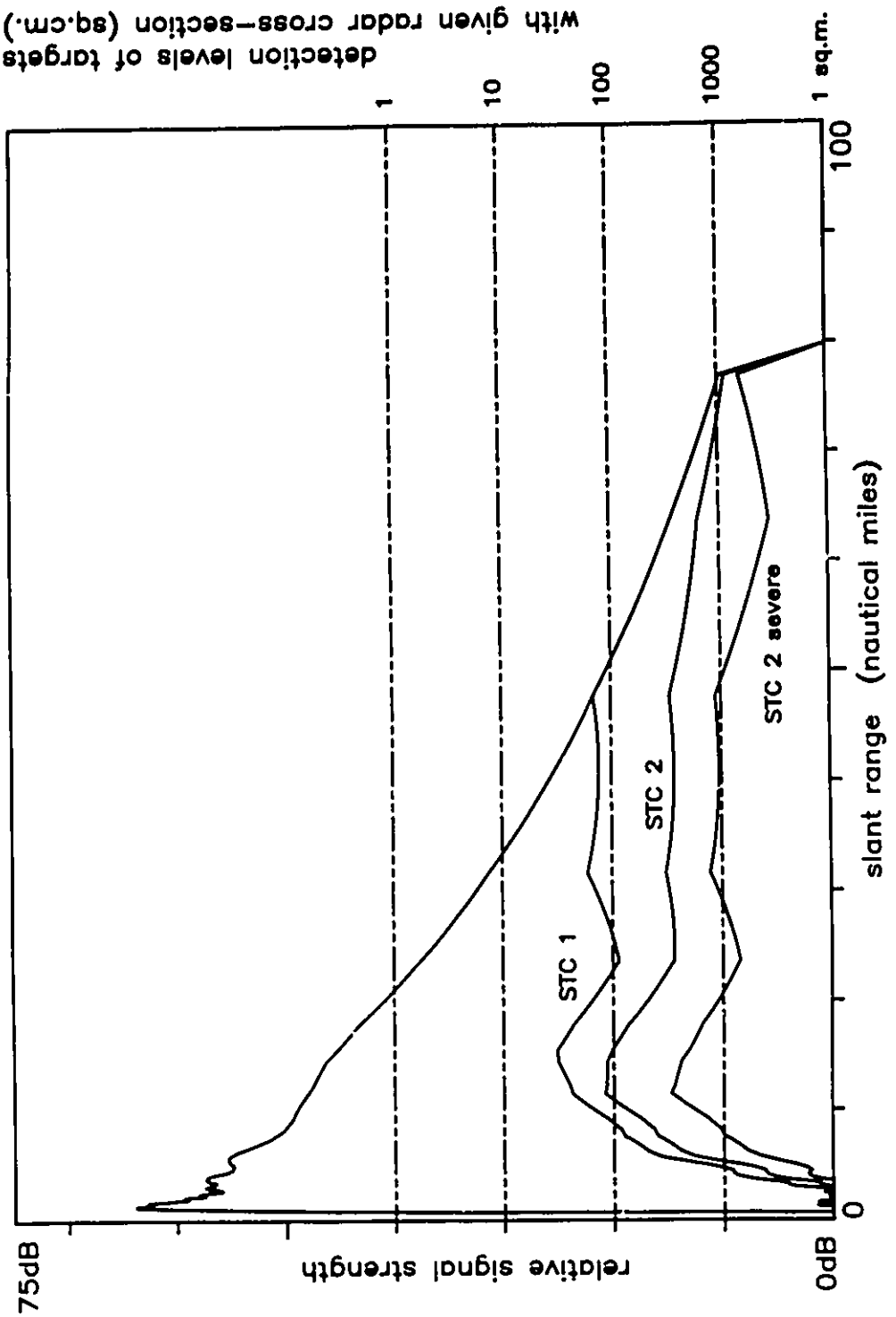


Figure B.2 TRACS–ASR detection performance at 6000 ft above sea level

The beam centre is assumed to be at a 2° elevation angle.

been computed using the modified csc^2 pattern, a 4/3 earth radius assumption, and an altitude of 6000 ft above sea level (asl). This level corresponds to approximately 4200 ft above ground (agl) at Moose Jaw, and represents a mean migration altitude for birds. The sharp signal drop at 79 nmi is due to the radar horizon, and the *cone of silence* is evident inside the 5 mile range of the radar. Note that at that altitude the minimum slant range is 4200 ft, or 0.7 nmi, and the plot can therefore not reach zero miles. The beam center ($F_t = 0$ dB) is assumed to be at an elevation angle of 2°, although the exact value is not known (it is adjustable). The bird detection capability is not materially reduced from its maximum (as in Figure B.1) for this case, although the minimum range is increased somewhat. Note that when STC is turned off very small targets with cross-sections of less than 1 m^2 become detectable inside the 20 mile range.

To obtain a better appreciation of the range/altitude performance of the radar, the contour plot in Figure B.3 shows the regions where small radar targets are most detectable. The lines represent the contours where $D_0 = D_r = 0$ dB with STC 1 selected. Lines of constant elevation based on the 4/3 earth radius assumption are superimposed. Due to the height of the radar above ground, the horizon is slightly below 0° elevation. It should be stressed here again, that this contour plot represents the ideal condition without multipath considerations. Favorite bird migration altitudes are between 1000 ft and 10,000 ft above ground, although higher and lower altitudes have been observed. Frequently, however, unfavorable winds at higher altitude will keep migration low and long-range detection will not be possible. Minimum detection altitudes also rise south of Moose Jaw due to the presence of hills with peak elevations of 2600 to 2900 ft above sea level. These hills raise the horizon to elevation angles of 0.2 to 0.3°, and some of the closer hills even raise the angles to as much as 0.7°. From

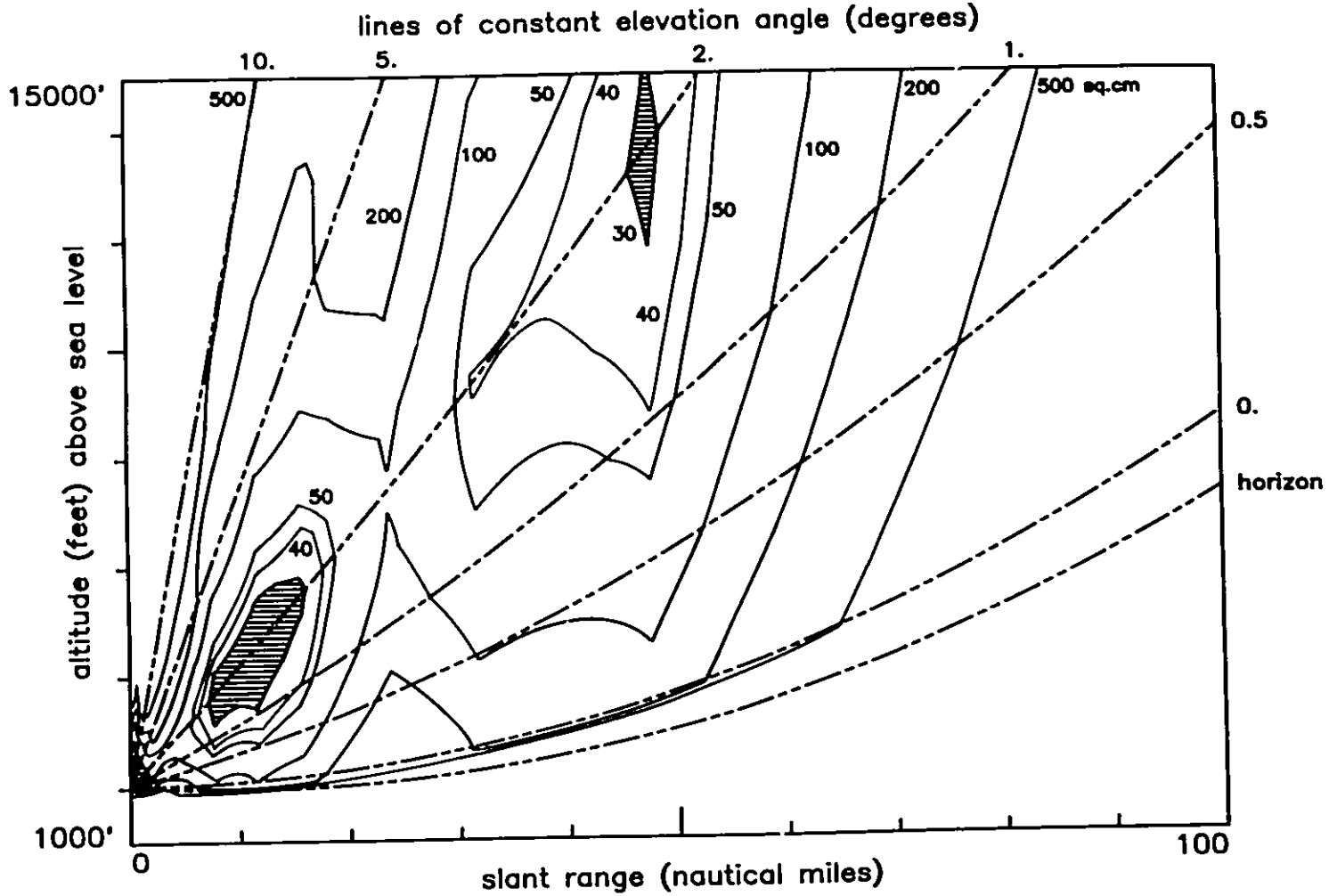


Figure B.3 TRACS–ASR minimum target detection performance using STC 1

The contours enclose regions where targets with indicated radar cross-section may be detected.

Figure B.3 it can be seen that this limits the visibility of a target at 6000 ft asl to about 35 to 40 miles.

It is also interesting to note the expected levels for rain clutter. Rain is fairly evenly distributed within the radar resolution cell and therefore no longer decreases in strength according to R^4. The volume of the cell increases with range proportional to R^2 as the beam widens in both azimuth and elevation. The resulting R^2 law means that the STC curve is not matched to it and, if used, will actually cause an increase in the observed clutter strength with range until it is no longer effective. This phenomenon is shown in Figure B.4. The maximum clutter strength is thus expected to be found at about 47 miles. The expected signal levels for several different precipitation rates is plotted in Figure B.5. The top curve represents a fairly heavy rain (16 mm/hr) and it is unlikely that much stronger clutter will commonly be observed unless STC is turned off.

B.2 Dynamic Range Limitations

The thermal noise floor places a limit on the small target detection performance. The 10-bit A/D word size places an upper limit on the signal strength which, when exceeded, will distort the signal and destroy the spectral information required for successful classification. The maximum dynamic range of a 10-bit A/D sampling linear bipolar data is 54 dB, assuming that the 0 dB level represents a signal voltage Δ equal to one quantization step. In the presence of a random signal the quantization noise has an average power given by $\Delta^2/6$, and is 7.8 dB down from this level. To maintain small signal fidelity, however, quantization noise must not dominate. In fact, for the 10 dB processing gain to be available the noise at the integrator must be completely decorrelated (white). Quantization noise itself only

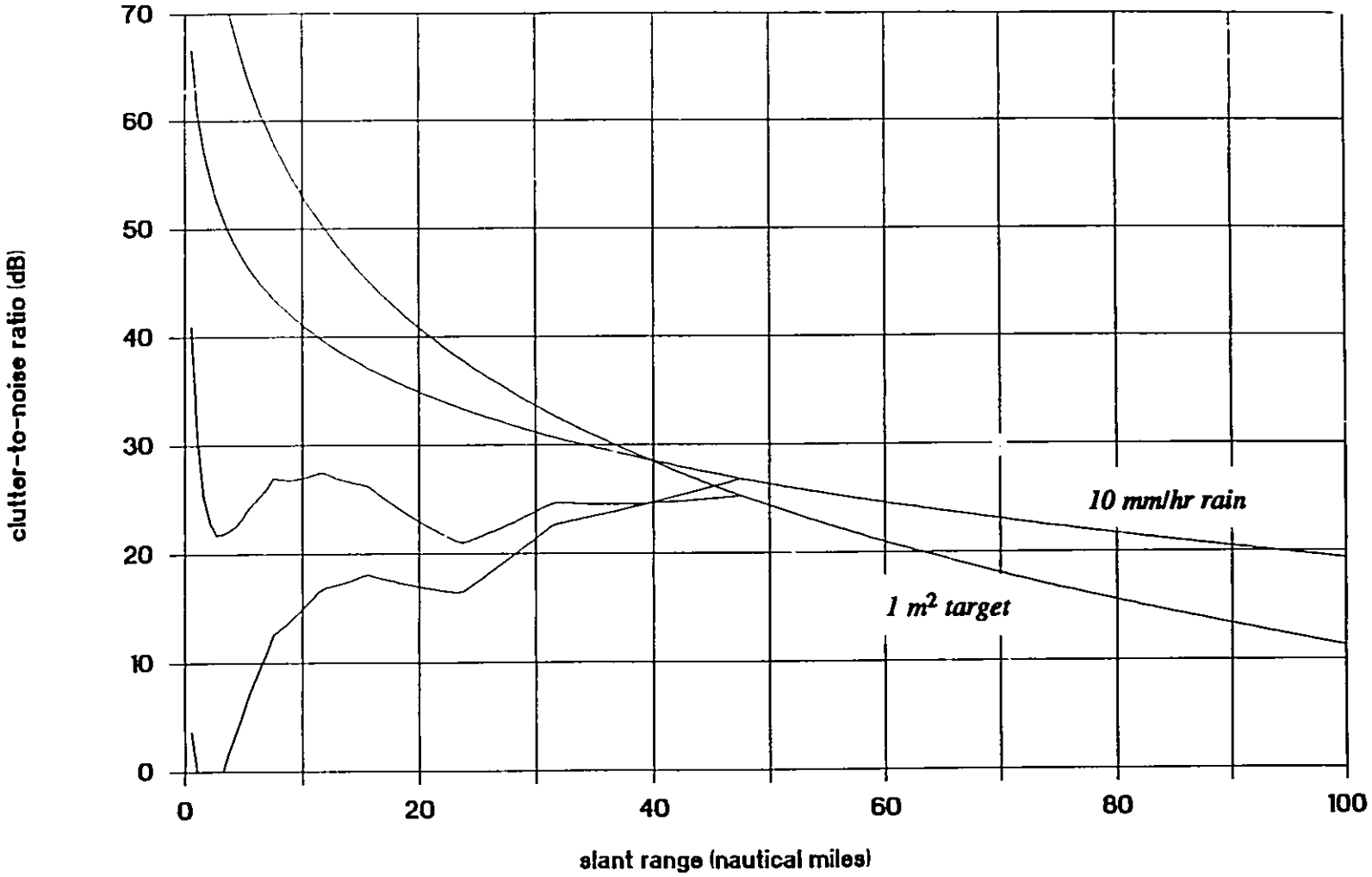


Figure B.4 TRACS–ASR 10 mm/hr rain clutter levels versus range

Expected clutter strengths compared to a 1 m^2 target. Lower curves are for STC 1 turned on.

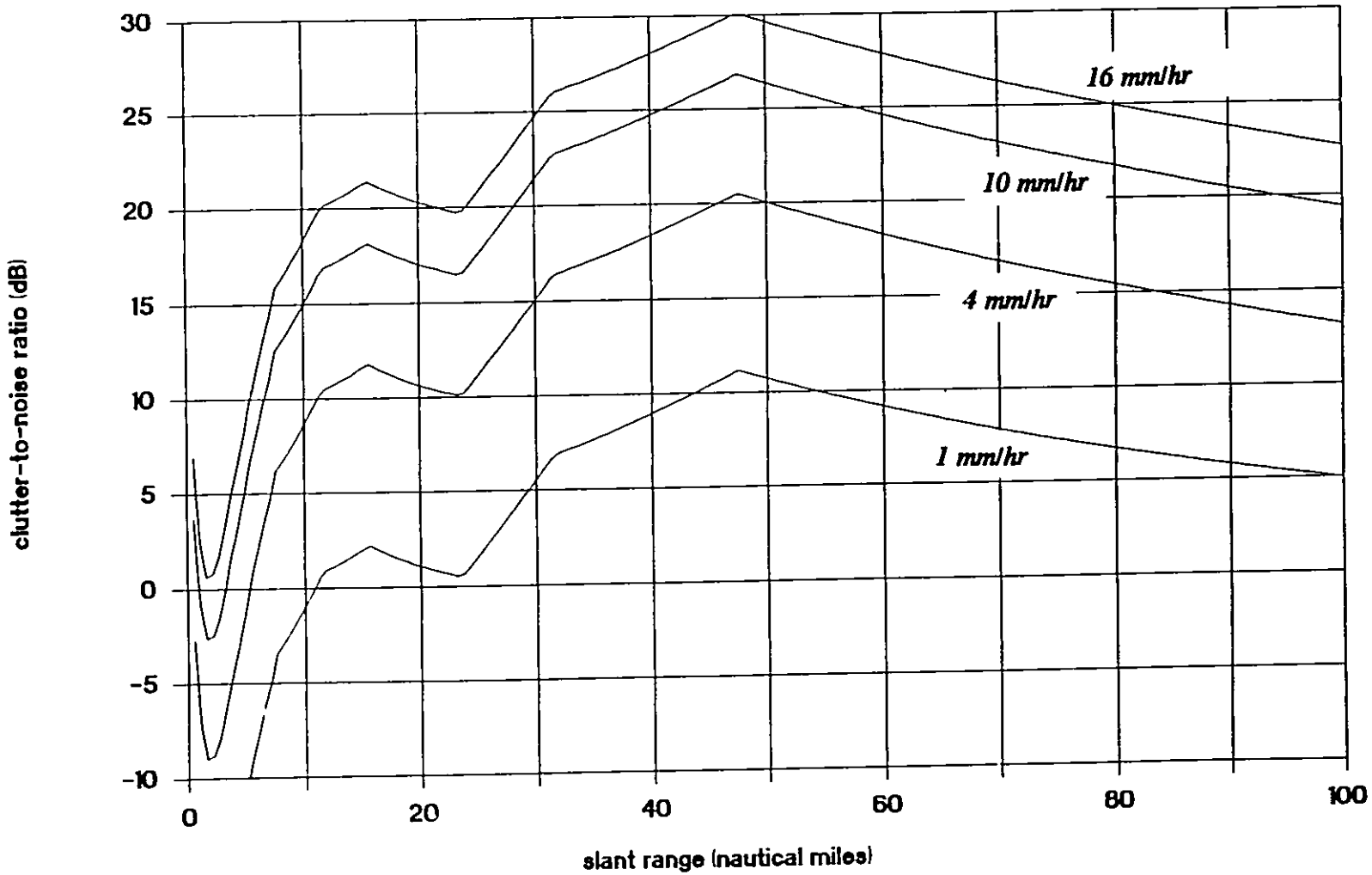


Figure B.5 TRACS–ASR comparative rain clutter levels versus range

Expected clutter strengths for 1, 4, 10, and 16 mm/hr rain fall with STC 1 turned on.

displays this characteristic when the signal to be digitized is random, and thus the thermal noise must be considerably stronger. The radar in Moose Jaw is adjusted such that the thermal to quantization noise ratio is 14.4 dB (12.1 dB in Trenton: the difference lies mainly in the noisier Q channel). Of course, this lowers the dynamic range to 46.3 dB (49.6 dB in Trenton) and overload becomes more likely. When STC 1 is selected, RCS values of 100 to 200 m^2 will saturate the A/D inside 50 miles.

To increase the SNR of bird clutter and to improve the detectability of small targets, STC must be turned off. Unfortunately, this also increases the danger of A/D overload significantly. In fact, as very small targets such as insects become visible birds could potentially exceed the available dynamic range. Table B.1 shows

RANGE	100 cm^2	500 cm^2	2800 cm^2	SNR: 46 dB
45 nmi	6 dB	13 dB	21 dB	93 m^2
20 nmi	21 dB	28 dB	35 dB	3.2 m^2
15 nmi	26 dB	33 dB	40 dB	1.0 m^2
12 nmi	30 dB	37 dB	44 dB	4100 cm^2
10 nmi	33 dB	40 dB	48 dB	1900 cm^2
8 nmi	37 dB	44 dB	52 dB	790 cm^2
7 nmi	39 dB	46 dB	54 dB	460 cm^2
6 nmi	42 dB	49 dB	57 dB	250 cm^2
5 nmi	45 dB	52 dB	60 dB	120 cm^2
4 nmi	49 dB	56 dB	64 dB	48 cm^2
3 nmi	54 dB	61 dB	69 dB	15 cm^2

Table B.1 Typical signal–to–noise ratios for bird clutter with STC off

The last column shows the maximum radar cross-section tolerable by the 46 dB dynamic range of the 10-bit A/D converter.

typical SNR values for three bird clutter sizes (most clutter is expected to range from 10 to 500 cm^2 [Vaughn, 1985; Rabinowitz *et al.*, 1985], although some studies [Barry *et al.*, 1973] have reported bird flock cross-sections of up to 2 m^2, with a mean of 2800 cm^2), as well as the maximum tolerable RCS for clutter situated in the center of the beam when STC is turned off.

Appendix C

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.*, 1973; Blokpoel, 1974], 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 knots, 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 smaller 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 targets travelling in the north-west direction (Figure C.1). 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 [Vaughn, 1985; Rabinowitz *et al.*, 1985].

Flight altitudes depend largely on the weather, particularly wind and cloud cover, but are expected to be mostly below 10,000 ft agl [Richardson, 1971, 1972; Blokpoel, 1974; Blokpoel and Burton, 1975; Blokpoel, 1976]. 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 (see Figure B.3). 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 travelling below 3000 ft agl and in large flocks (200 cm^2) at most times. On the other hand, the 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 travelling as high as 6000 to 8000 ft agl. In those cases, however, direction of travel tended to differ from those at

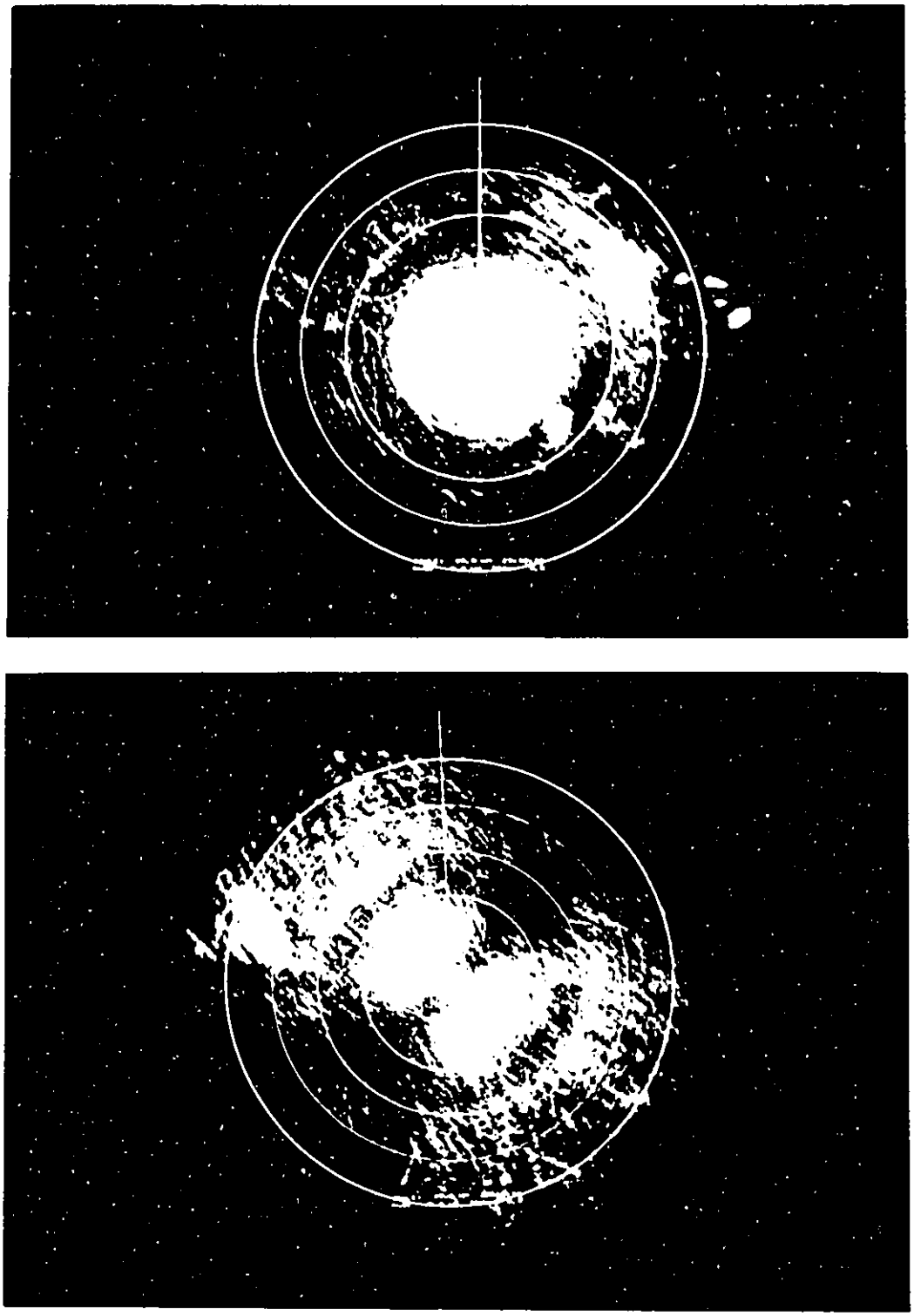


Figure C.1 PPI time exposure photographs of bird clutter

The top photograph was taken from unprocessed video, the bottom from MTI video. Time exposure was 5 minutes. Range rings are spaced at 10 nmi.

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 those conditions.

Analysis of data from the tapes also pointed towards 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. Signal-to-noise ratios ranged from virtually undiscernible to 15 dB when STC was turned on (curve 1). Clutter strengths of up to 35 dB where found when turned off. Doppler frequencies ranged from zero to up to 400 Hz, corresponding to an upper speed of travel of about 85 knots. Airspeeds are generally not this high [Blokpoel, 1974; Richardson, 1974, 1979], but groundspeeds of this magnitude are possible during periods of favorable winds [Barry *et al.*, 1973; Muehe *et al.*, 1974]. 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 Figure C.2 is 326°!). Perhaps none of these arguments constitute a "proof" of the observed clutter being caused by birds, but it surely is strong

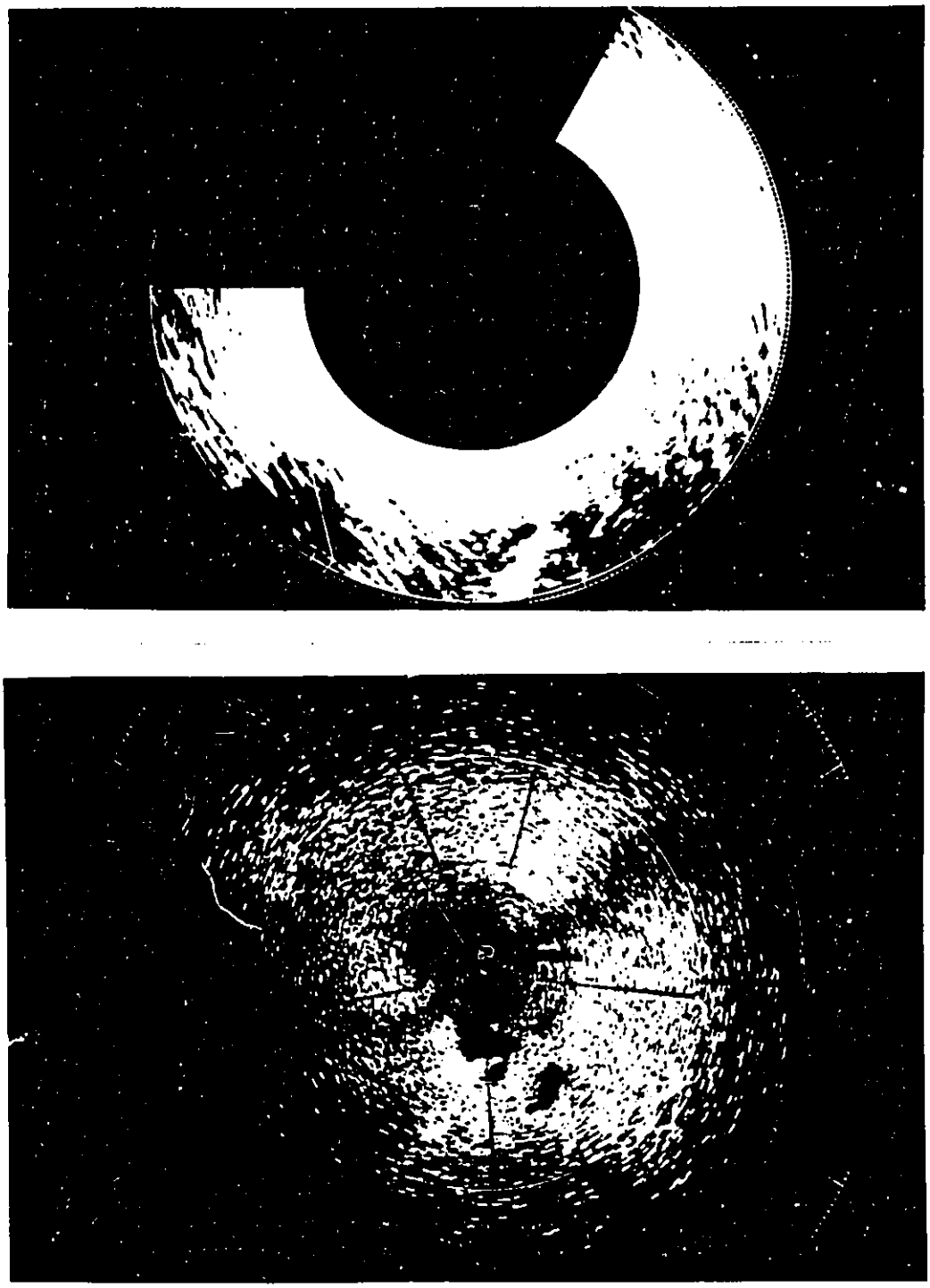


Figure C.2 Simulated PPI images of processed radar data including bird clutter

The top image shows 32 scans integrated non-coherently. The bottom is a single scan with ground and zero doppler clutter removed.

evidence. An alternate explanation for its source is certainly not immediately apparent.

Non-radar 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, 1973; Blokpoel and Gauthier, 1975], 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.

Appendix D

VALUE OF THE REFLECTION COEFFICIENTS FOR A SINUSOID IN NOISE

In this appendix an expression is derived for the value of the reflection coefficients

$$\hat{\rho}_m = \frac{-2 \sum_{n=m+1}^{N} f_n^{(m-1)} b_{n-1}^{*(m-1)}}{\sum_{n=m+1}^{N} [|f_n^{(m-1)}|^2 + |b_{n-1}^{(m-1)}|^2]} \tag{D.1}$$

as $N \longrightarrow \infty$, where N is the number of data samples used in the estimates. Mathematically, this in not an *expectation* of the value of $\hat{\rho}_m$, but rather the "true" value computed using the expectations or true values of the autocorrelation coefficients. It can be shown [Kay and Makhoul, 1983] that, as N becomes large, $\hat{\rho}_m$ does converge on the true value as

$$\rho_m = \frac{-2\, E[f_n^{(m-1)} b_{n-1}^{*(m-1)}]}{E[|f_n^{(m-1)}|^2 + |b_{n-1}^{(m-1)}|^2]} \tag{D.2}$$

Hence, with respect to this true value, $\hat{\rho}_m$ is consistent but not necessarily unbiased.

Let the time series x_n represent samples of a complex sinusoid in noise

$$x_n = A\, exp(jn\phi) + w_n \tag{D.3}$$

where ϕ is the frequency of the sinusoid normalized to the sampling frequency f_s. The

algebra of substituting this expression into (D.2) and the lattice equations (5.8), (5.9), (5.10) will quickly become cumbersome, hence the following analysis will proceed in two steps. First, ϕ is set to zero such that $x'_n = A + w'_n$, where w'_n is an independent measurement noise sequence distributed according to $\mathcal{N}(w{:}0,W)$. Second, the entire time series is heterodyned by the frequency ϕ. This step involves multiplying the time series x'_n by the complex sinusoid (with unity magnitude) $exp(jn\phi)$. The autocorrelation function $R'(m)$ of x'_n is defined as

$$R'(m) = E[x'_n x'^{*}_{n-m}] = A^2 + E[w'_n w'^{*}_{n-m}] \tag{D.4}$$

hence

$$R'(0) = A^2 + W \quad \text{and} \quad R'(m) = A^2, \; m > 1 \tag{D.5}$$

For simplicity of notation no distinction will be made between the estimated and theoretical values of the reflection coefficients. They are related to the autocorrelation values as [Box and Jenkins, 1976; Haykin and Kesler, 1979]

$$\rho_m = -\frac{\sum_{i=0}^{m-1} h_i^{(m-1)} R(m-i)}{\sum_{i=0}^{m-1} h_i^{*(m-1)} R(i)} \tag{D.6}$$

Using (D.5)

$$\rho'_m = -\frac{A^2 \sum_{i=0}^{m-1} h'^{(m-1)}_i}{W + A^2 \sum_{i=0}^{m-1} h'^{*(m-1)}_i} \tag{D.7}$$

where $h'^{(m)}_0 = 1$. Using Levinson's recursion

$$h_i^{(m)} = h_i^{(m-1)} + \rho_m h_{m-i}^{*(m-1)} \qquad i = 1, 2, \ldots, m \tag{D.8}$$

the summation terms may be written as

$$\sum_{i=0}^{m} h_i'^{(m)} = \sum_{i=0}^{m} h_i'^{(m-1)} + \rho_m' \sum_{i=0}^{m} h_{m-i}'^{*(m-1)} \tag{D.9}$$

The $h_i^{(m)}$ are the coefficients of the m^{th}-order prediction error filter (PEF). Since, for real valued time series, the PEF coefficients (which are equal to the negative of the autoregressive model parameters) are also real valued, and $h_m'^{(m-1)} = 0$, one can write

$$\sum_{i=0}^{m} h_i'^{(m)} = (1 + \rho_m') \sum_{i=0}^{m-1} h_i'^{(m-1)} = \prod_{i=1}^{m} (1 + \rho_i') \tag{D.10}$$

Resubstituting in (D.7)

$$\rho_m' = - \frac{n_0 \prod_{i=1}^{m-1} (1 + \rho_i')}{1 + n_0 \prod_{i=1}^{m-1} (1 + \rho_i')} \tag{D.11}$$

where n_0 is the signal-to-noise ratio A^2/W. Noting that (using D.11)

$$n_0 \prod_{i=1}^{m-2} (1 + \rho_i') = - \frac{\rho_{m-1}'}{1 + \rho_{m-1}'} \tag{D.12}$$

the following recursive relation is obtained

$$\rho'_m = \frac{\rho'_{m-1}}{1 - \rho'_{m-1}} \tag{D.13}$$

This recursion is initialized from (D.11)

$$\rho'_1 = -\frac{n_0}{1 + n_0} \tag{D.14}$$

whic : then leads to

$$\rho'_m = -\frac{n_0}{1 + m\ n_0} \tag{D.15}$$

The next step is to heterodyne the time series using $exp(jn\phi)$. When setting $x_n = x'_n exp(jn\phi)$, the autoregressive model, which is equivalent to (D.3) as $M \longrightarrow \infty$

$$x_n = \sum_{i=1}^{M} \gamma_i x_{n-i} + w_n \tag{D.16}$$

can be rewritten as

$$x'_n exp(jn\phi) = \sum_{i=1}^{M} \gamma_i x'_{n-i} exp(j(n-i)\phi) + w'_n exp(jn\phi) \tag{D.17}$$

with

$$\gamma_i = -h_i^{(M)} = \gamma'_i exp(ji\phi) = -h_i'^{(M)} exp(ji\phi) \tag{D.18}$$

The autocorrelation function becomes

$$R(m) = E[x_n x^*_{n-m}] = A^2 exp(jm\phi) + E[w_n w^*_{n-m}] \tag{D.19}$$

hence

$$R(0) = A^2 + W \qquad \text{and} \qquad R(m) = A^2 exp(jm\phi), \quad m \geq 1 \tag{D.20}$$

From (D.18) and (D.6)

$$\rho_m = -\frac{\sum_{i=0}^{m-1} h_i'^{(m-1)} exp(ji\phi) R(m-i)}{\sum_{i=0}^{m-1} h_i'^{*(m-1)} exp(-ji\phi) R(i)} \tag{D.21}$$

Using (D.20)

$$\rho_m = -\frac{A^2 \sum_{i=0}^{m-1} h_i'^{(m-1)}}{W + A^2 \sum_{i=0}^{m-1} h_i'^{*(m-1)}} exp(jm\phi) \tag{D.22}$$

Comparing (D.22) with (D.7) and (D.15) gives the final result as

$$\rho_m = \rho_m' exp(jm\phi) = -\frac{n_0}{1 + m\, n_0} exp(jm\phi) \tag{D.23}$$

Figure D.1 shows a set of MEM spectra based on this result. It is instructive to evaluate ρ_1 for the case of multiple complex sinusoids in noise, each having a different frequency. The resulting time series is defined as

$$x_n = \sum_{k=1}^{K} A_k\, exp(jn\phi_k) + w_n \tag{D.24}$$

and the autocorrelation values become

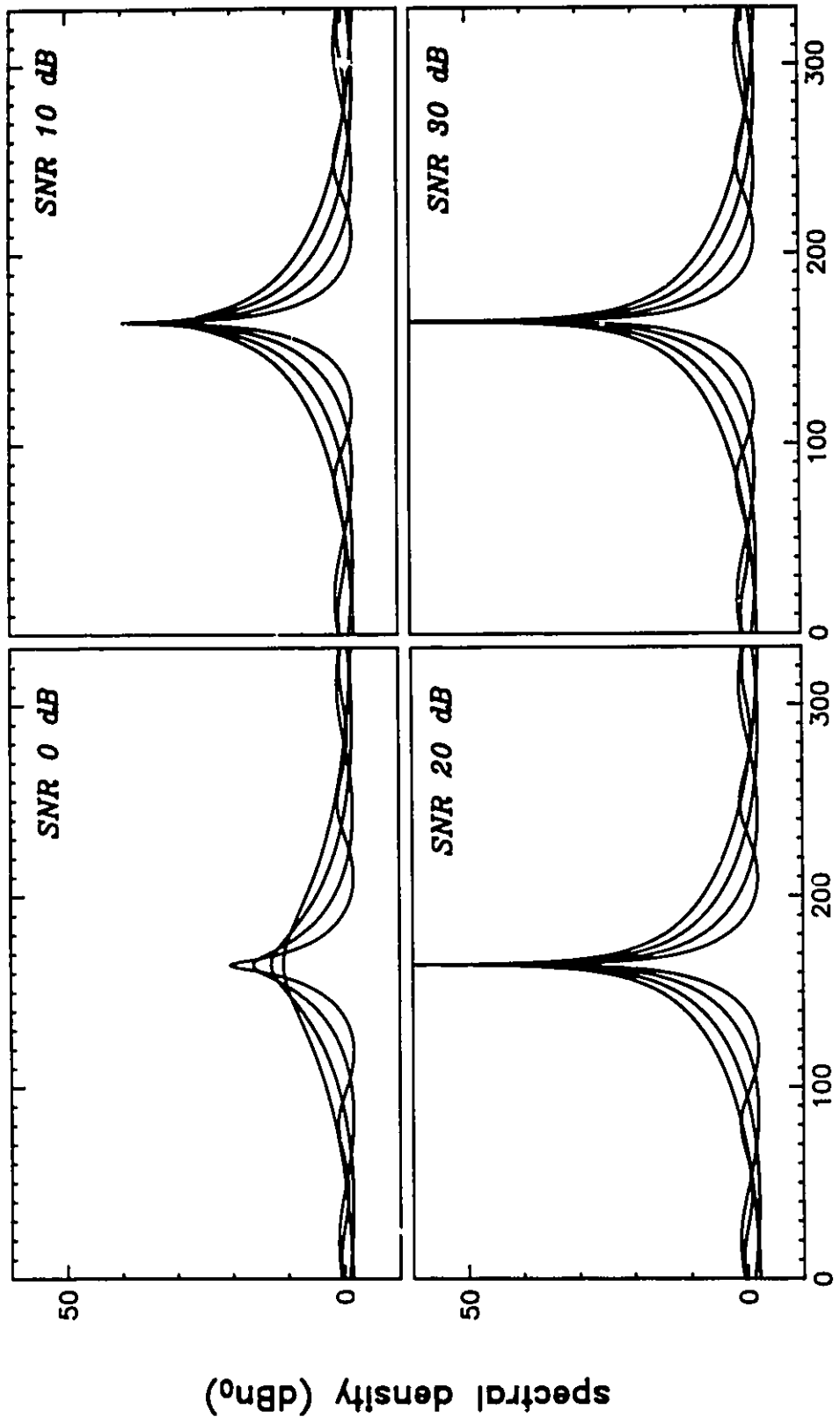


Figure D.1 Maximum entropy spectra based on theoretical values for the reflection coefficients 2^{nd}-, 3^{rd}-, 5^{th}-, and 9^{th}-order spectra are computed for the SNR values shown.

$$R(m) = E[x_n x^*_{n-m}] = \sum_{k=1}^{K} A_k^2 \, exp(jm\phi_k) + E[w_n w^*_{n-m}] , \qquad (D.25)$$

thus

$$R(0) = \sum_{k=1}^{K} A_k^2 + W \quad \text{and} \quad R(m) = \sum_{k=1}^{K} A_k^2 \, exp(jm\phi_k), \quad m > 1 \qquad (D.26)$$

The initial conditions for the lattice filter are $f_n^{(0)} = b_n^{(0)} = x_n$. From (D.2)

$$\rho_1 = \frac{-2\,E[x_n x^*_{n-1}]}{E[|x_n|^2 + |x_{n-1}|^2]} = -\frac{R(1)}{R(0)} \qquad (D.27)$$

Substituting (D.26)

$$\rho_1 = -\frac{\sum_{k=1}^{K} A_k^2 \, exp(j\phi_k)}{\sum_{k=1}^{K} A_k^2 + W} \qquad (D.28)$$

This can be written as

$$\rho_1 = \sum_{k=1}^{K} q_k \rho_{1k} \qquad (D.29)$$

where

$$\rho_{1k} = -\frac{n_{0k}}{1 + n_{0k}} exp(j\phi_k) \quad \text{and} \quad q_k = \frac{A_k^2 + W}{R(0)} \qquad (D.30)$$

A_k and ϕ_k represent the magnitude and frequency of the complex sinusoids present, and $n_{0k} = A_k^2/W$. This weighted average shows that ρ_1 is dominated by the strongest signal. Note that in the case of two sinusoids and $R(0) >> W$, ρ_1 lies on the line in the ρ-plane joining the two complex coefficients ρ_{11} and ρ_{12}. In the case of three or more sinusoids, ρ_1 lies within the region bounded by the lines joining the coefficients ρ_{1k}.

This implies that the magnitude of ρ_1 is always less than that for the strongest sinusoid in the group, that is $|\rho_1| \leq |\rho_{1_k}|$ for all k. Consequently, the SNR value computed using $R(0)$ and the known measurement noise W, then substituted in (D.23), is only an *upper bound* for the true magnitude of ρ_1, which obviously can be much less. Conversely, the signal-to-noise ratio arising from (D.23) given the estimate $\hat{\rho}_1$ per (D.1) is always less than, or equal to, the true SNR.

Appendix E

REFLECTION COEFFICIENT SENSITIVITY TO QUADRATURE ERRORS

Several errors can occur during the sampling of coherent quadrature data. Alignment errors in the in-phase and quadrature channels can result in unequal voltage gains, a phase difference other than the required 90°, and different noise levels. The first two factors can significantly alter the spectrum of the data and, consequently, the reflection coefficients themselves. In this appendix the effects of these errors on the spectrum are derived first. The sensitivity of the first reflection coefficient to changes in gain and phase errors is computed and some heuristic arguments are presented regarding the sensitivity of higher-order coefficients. Finally an algorithm to find the needed correction factors without calibration signals is described.

E.1 Effects of Gain and Phase Errors on the Spectrum

Given the signal phasor X with magnitude S and phase angle ϕ, the true in-phase and quadrature components are given by (see Figure E.1)

$$I = S\cos\phi, \qquad Q = S\sin\phi \tag{E.1}$$

Defining the voltage gain ratio between the channels as k_e, the gain error as e, and the Q channel phase error as δ, the apparent in-phase and quadrature components become

$$I' = k_e I = (1 + e) I \tag{E.2}$$

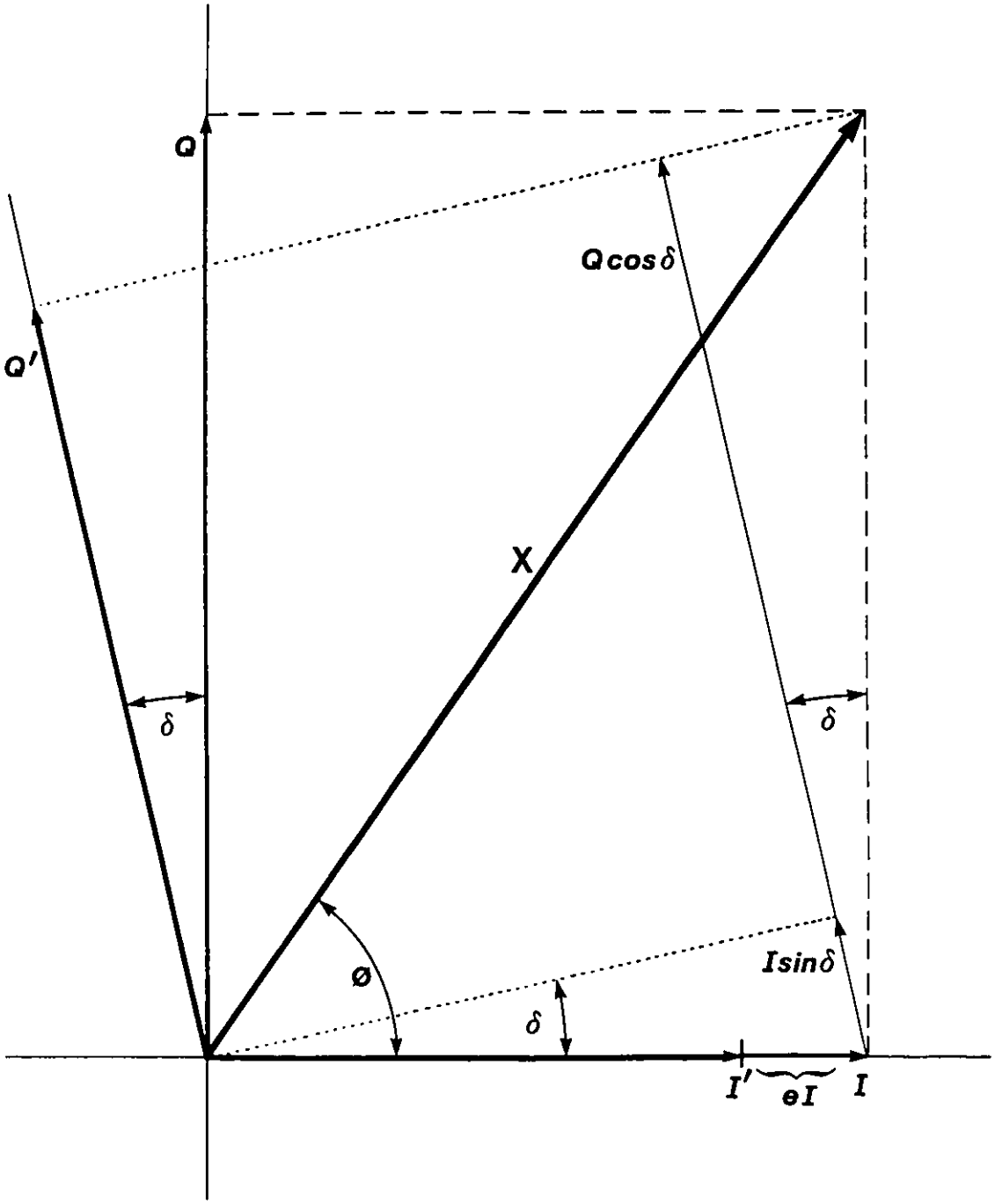


Figure E.1 Quadrature error geometry

The error free I and Q components are derived from vector x and phase angle ϕ. Quadrature phase error δ and gain error e lead to the measured values I' and Q'.

$$Q' = S\sin(\phi-\delta) = S\sin\phi\cos\delta - S\cos\phi\sin\delta$$

$$Q' = Q\cos\delta - I\sin\delta \tag{E.3}$$

The required correction factors can immediately be seen to be

$$I = k_1 I'; \qquad k_1 = \frac{1}{k_e} \tag{E.4}$$

$$Q = k_2 (Q' + k_3 I); \quad k_2 = \frac{1}{\cos\delta}; \quad k_3 = \sin\delta \tag{E.5}$$

A technique for finding the values k_e and δ is given later in this appendix. To find the effect of these two errors on the spectrum, the apparent received signal V' must be broken down into its various frequency components. The true signal V is assumed to be a single complex sinusoid representing a Doppler target with amplitude A, frequency ω, and arbitrary phase ψ

$$V = A\cos(\omega t+\psi) + jA\sin(\omega t+\psi) \tag{E.6}$$

$$V' = k_e A\cos(\omega t+\psi) + jA\sin(\omega t+\psi-\delta) \tag{E.7}$$

In exponential notation

$$V' = \frac{A}{2}\left[k_e e^{j(\omega t+\psi)} + k_e e^{-j(\omega t+\psi)} + e^{j(\omega t+\psi-\delta)} - e^{-j(\omega t+\psi-\delta)}\right]$$

$$V' = \frac{A}{2}\left[(k_e+e^{-j\delta})e^{j(\omega t+\psi)} + (k_e-e^{j\delta})e^{-j(\omega t+\psi)}\right] \tag{E.8}$$

The relative power in the positive frequency component is found to be (left hand term of Equation E.8)

$$P'_p = \frac{A^2}{4}[1 + 2k_e cos\delta + k_e^2] \tag{E.9}$$

and in the negative frequency component (right hand term of Equation E.8)

$$P'_n = \frac{A^2}{4}[1 - 2k_e cos\delta + k_e^2] \tag{E.10}$$

From these equations it is clear that some of the power from the true target or positive frequency component is transferred into the negative frequency component ("mirror image" or false target). Using Equations (E.9) and (E.10), a false-target-level ratio can be defined [see also Sinsky and Wang, 1974]

$$FT(\text{dB}) = 10\ log \frac{P'_n}{P'_p} = 10\ log \left[\frac{1 + k_e^2 - 2k_e cos\delta}{1 + k_e^2 + 2k_e cos\delta} \right] \tag{E.11}$$

Figure E.2 shows false-target levels for some typical error ranges found in the TRACS-ASR system. Referring back to Equations (E.9) and (E.10), note that the total power does not remain constant ($P = A^2$):

$$P' = P'_p + P'_n = \frac{A^2}{2}[1 + k_e^2] \tag{E.12}$$

The noise power W' changes by a similar amount, leaving the total signal-to-noise ratio constant. The shift of power to the false target thus causes a net reduction in the true target SNR. This reduction can be expressed as a loss in SNR

$$LSN(\text{dB}) = 10\ log\ (P'_p/W') - 10\ log\ (P/W) \tag{E.13}$$

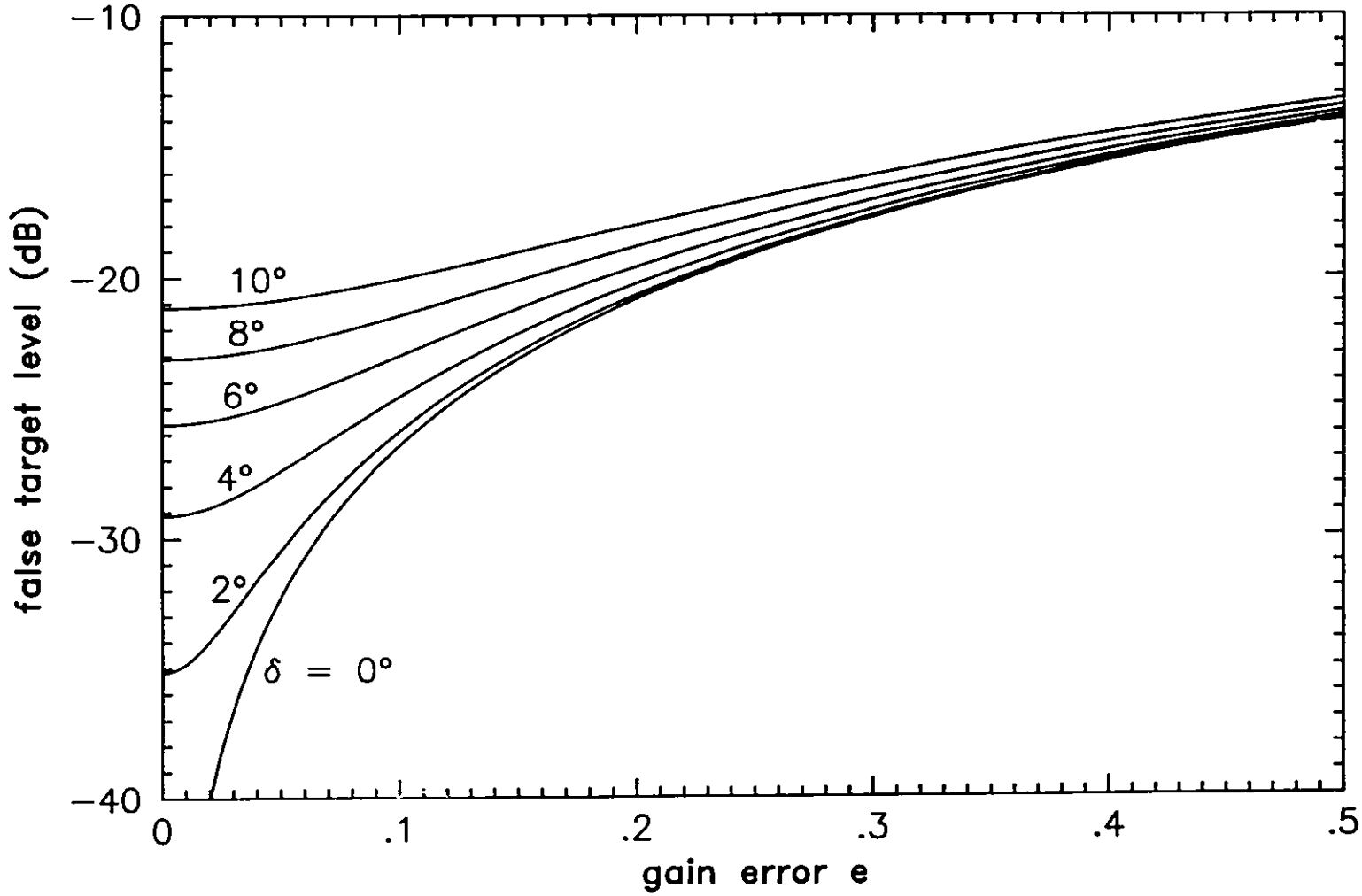


Figure E.2 False target levels resulting from quadrature errors

Note that the total noise power is given by $W' = \frac{W}{2}[1 + k_e^2]$, thus

$$LSN(\text{dB}) = 10\, log \left[\frac{1 + k_e^2 + 2k_e cos\delta}{2\,(1 + k_e^2)} \right] \tag{E.14}$$

This loss is plotted in Figure E.3 for several typical error values.

E.2 Reflection Coefficient Sensitivity

The first-order reflection coefficient computed using the Burg formula is given as

$$\rho_1 = \frac{-2 \sum_{n=2}^{N} x_n\, x_{n-1}^*}{\sum_{n=2}^{N} [\,|x_n|^2 + |x_{n-1}|^2]} = \frac{RN}{RD} \tag{E.15}$$

where x_n is the n^{th} complex data sample and N is the number of samples in the data set. Denoting the error in x_n' resulting from the gain and phase misalignments as Δx_n, the apparent reflection coefficient becomes

$$\rho_1' = \frac{-2 \sum_{n=2}^{N} [x_n + \Delta x_n]\ [x_{n-1}^* + \Delta x_{n-1}^*]}{\sum_{n=2}^{N} [\,|x_n + \Delta x_n|^2 + |x_{n-1} + \Delta x_{n-1}|^2]} = \frac{RN + \Delta RN}{RD + \Delta RD}$$

$$\rho_1' = \frac{RN - 2 \sum_{n=2}^{N} [\Delta x_n x_{n-1}^* + x_n \Delta x_{n-1}^* + \Delta x_n \Delta x_{n-1}^*]}{RD + \sum_{n=2}^{N} [\,|\Delta x_n|^2 + |\Delta x_{n-1}|^2 + 2\Re(x_n \Delta x_n^* + x_{n-1} \Delta x_{n-1}^*)]} \tag{E.16}$$

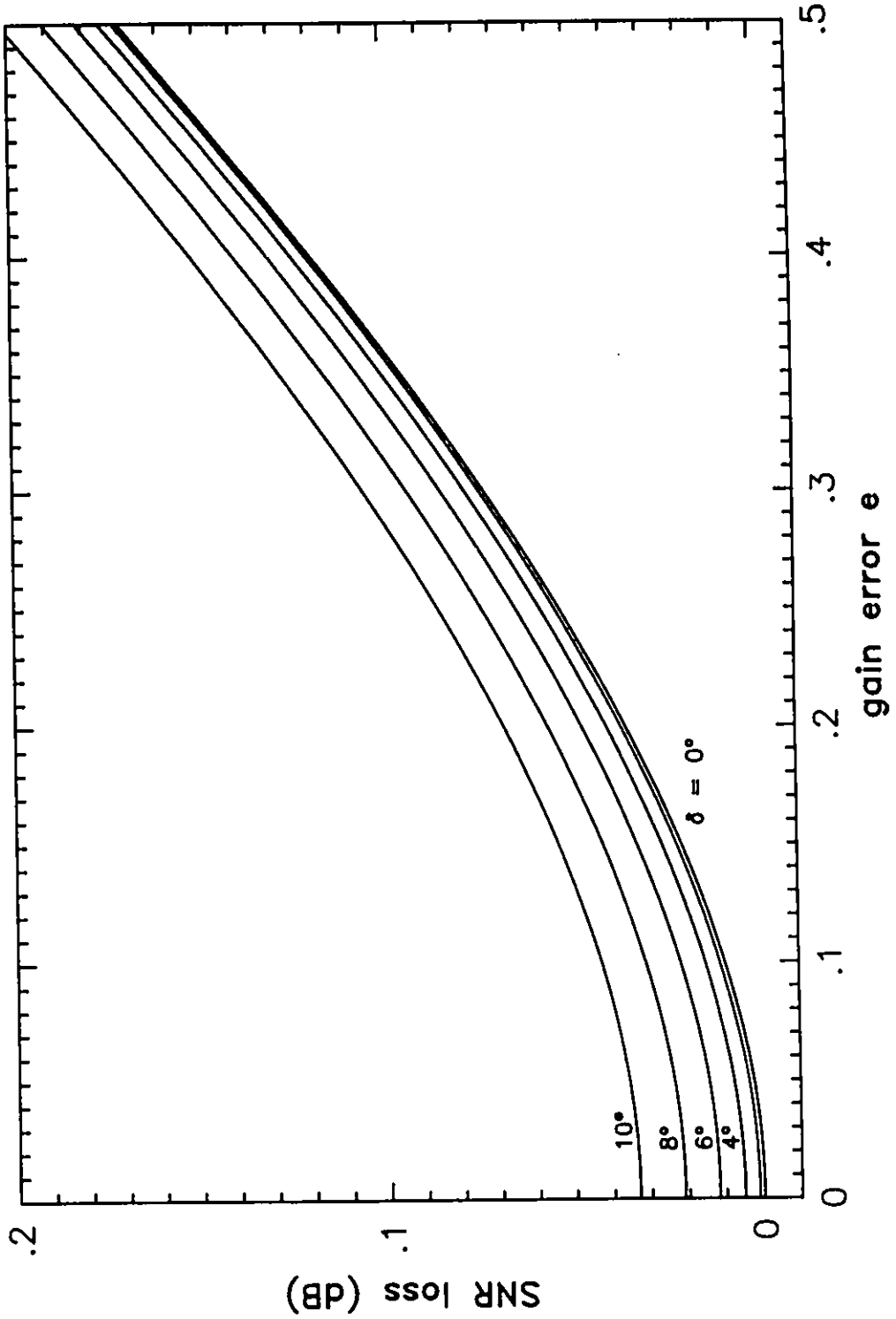


Figure E.3 Loss of true target signal–to–noise ratio as a function of quadrature errors

Expressing x_n and x_n' in terms of their I and Q components (E.2 and E.3)

$$x_n' = a_n' + jb_n' = x_n + \Delta x_n = a_n(1 + e) + j(b_n cos\delta - a_n sin\delta)$$
$$= [a_n + jb_n] + [a_n e + j(b_n(cos\delta - 1) - a_n \tag{E.17}$$

The error terms of Equation (E.16) can then be rewritten as

$$\Delta RN = -2 \sum_{n=2}^{N} [a_n a_{n-1}(k_e^2 - 1 + sin^2\delta) - b_n b_{n-1} sin^2\delta - (a_n b_{n-1} + a_{n-1} b_n) sin\delta\, cos\delta +$$
$$+ j(a_{n-1} b_n - a_n b_{n-1})(k_e cos\delta - 1)] \tag{E.18}$$

$$\Delta RD = \sum_{n=2}^{N} [(a_n^2 + a_{n-1}^2)(k_e^2 - 1) + (a_n^2 + a_{n-1}^2 - b_n^2 - b_{n-1}^2) sin^2\delta -$$
$$- (a_n b_n + a_{n-1} b_{n-1}) sin2\delta] \tag{E.19}$$

These terms are clearly data dependent and cannot be evaluated for any general case. Further insights may be gained, however, by assuming the signal to be a complex sinusoid in white noise

$$a_n = Acos(n\phi) + w_{In}; \qquad b_n = Asin(n\phi) + w_{Qn} \tag{E.20}$$

where $\phi = \omega\Delta t$, the phase difference between adjacent samples. The summation signs are then replaced with expectation operators

$$E[a_n] = \tfrac{1}{2}A^2 + W_I; \qquad E[b_n] = \tfrac{1}{2}A^2 + W_Q \tag{E.21}$$
$$E[a_n b_n] = E[\tfrac{1}{2}A^2 sin(2n\phi)] = 0 \tag{E.22}$$

$$E[a_n a_{n-1}] = \tfrac{1}{2}A^2 cos\phi \tag{E.23}$$

$$E[b_n b_{n-1}] = \tfrac{1}{2}A^2 cos\phi = E[a_n a_{n-1}] \tag{E.24}$$

$$E[a_n b_{n-1}] = -\tfrac{1}{2}A^2 sin\phi \tag{E.25}$$

$$E[a_{n-1} b_n] = \tfrac{1}{2}A^2 sin\phi = -E[a_n a_{n-1}] \tag{E.26}$$

Equations (E.18) and (E.19) thus become (recall $P = A^2$)

$$\Delta RN = -P\,[(k_e^2-1)cos\phi + j2(k_e cos\delta-1)sin\phi] \tag{E.27}$$

$$\Delta RD = W\,[(n_0+1-d_0)(k_e^2-1) - 2d_0 sin^2\delta] \tag{E.28}$$

where

$$n_0 = \frac{P}{W_Q + W_I} \quad \text{and} \quad d_0 = \frac{W_Q - W_I}{W_Q + W_I}$$

are defined as the signal-to-noise ratio and the channel differential noise factor respectively. Similarly the expected values for the true reflection coefficient can be evaluated

$$RN = -2\,E[(a_n a_{n-1}+b_n b_{n-1}) + j(b_n a_{n-1}+a_n b_{n-1})] = -2\,P\,[cos\phi + jsin\phi] \tag{E.29}$$

$$RD = E[a_n^2 + b_n^2 + a_{n-1}^2 + b_{n-1}^2] = 2\,[P + W] \tag{E.30}$$

hence

$$\rho_1' = -\frac{n_0\,[(k_e^2+1)cos\phi + j(2k_e cos\delta)sin\phi]}{(n_0+1)(k_e^2+1) + d_0(cos2\delta-k_e^2)} \tag{E.31}$$

Note that if the differential noise factor is zero, that is the noise in both channels is equal, then

$$\rho_1' = -\frac{n_0}{n_0+1}\left[\cos\phi + j\frac{2k_e}{k_e^2+1}\cos\delta\sin\phi\right] \tag{E.32}$$

which shows that only the imaginary part is affected by gain and phase errors; in fact since both $|2k_e/(k_e^2+1)|$ and $|\cos\delta|$ are always less than one, ρ_1' is driven away from the unit circle toward the real axis. Most of the information of interest is contained in the magnitude and phase which are simply, from Equation (E.31)

$$r' = |\rho_1'| = \frac{n_0\sqrt{[(k_e^2+1)\cos\phi]^2 + [(2k_e\cos\delta)\sin\phi]^2}}{(n_0+1)(k_e^2+1) + d_0(\cos 2\delta - k_e^2)} \tag{E.33}$$

$$\phi' = \arg(\rho_1') = \tan^{-1}\left[\frac{2k_e}{k_e^2+1}\cos\delta\tan\phi\right] \tag{E.34}$$

Note also that the phase angle of ρ_1' is independent of d_0 and n_0. Since the term in the brackets is always less than or equal to $\tan\phi$, with equality occurring only when no errors are present, any error will drive the angle away from $\pm\pi/2$ and towards 0 (0 or $\pm\pi$ if the four quadrant arctangent is used). This is illustrated in Figure E.4. Figure E.5 shows the sensitivity of the magnitude expressed in terms of the apparent SNR. From Equation (E.32) the magnitude r of the error free ρ_1' is seen to be $n_0/(n_0+1)$, hence this apparent SNR is defined as

$$u_0 = \frac{r'}{1 - r'} \tag{E.35}$$

The sensitivity of r to gain and phase errors can be found by differentiating Equation (E.33). To simplify this task define

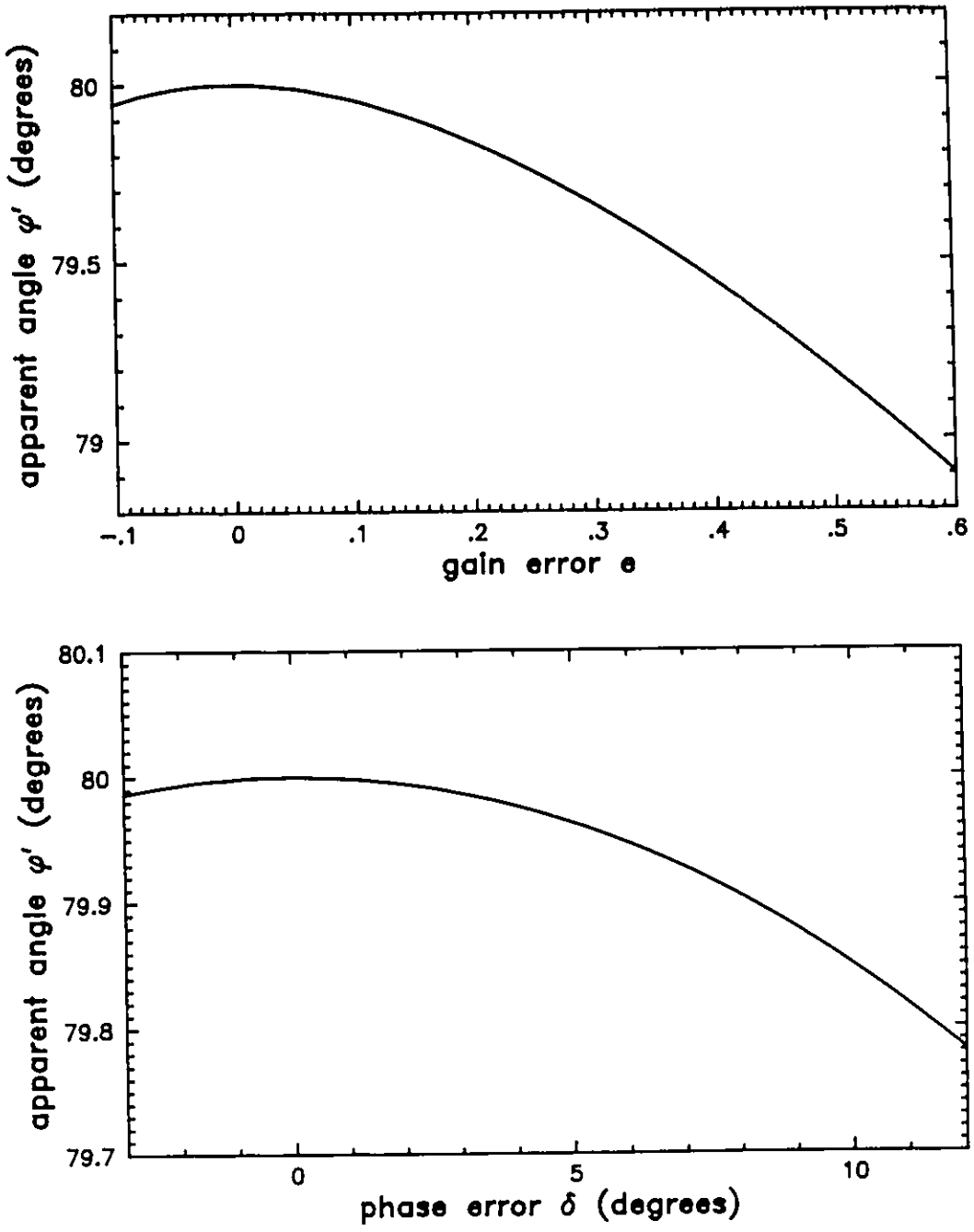


Figure E.4 Effects of gain and phase errors on the apparent phase of ρ_1

Plot of the deviation of the apparent phase angle ϕ' from its true value ($\phi = 80°$) as a function of gain error (top: $\delta = 0°$) and phase error (bottom: $e = 0$).

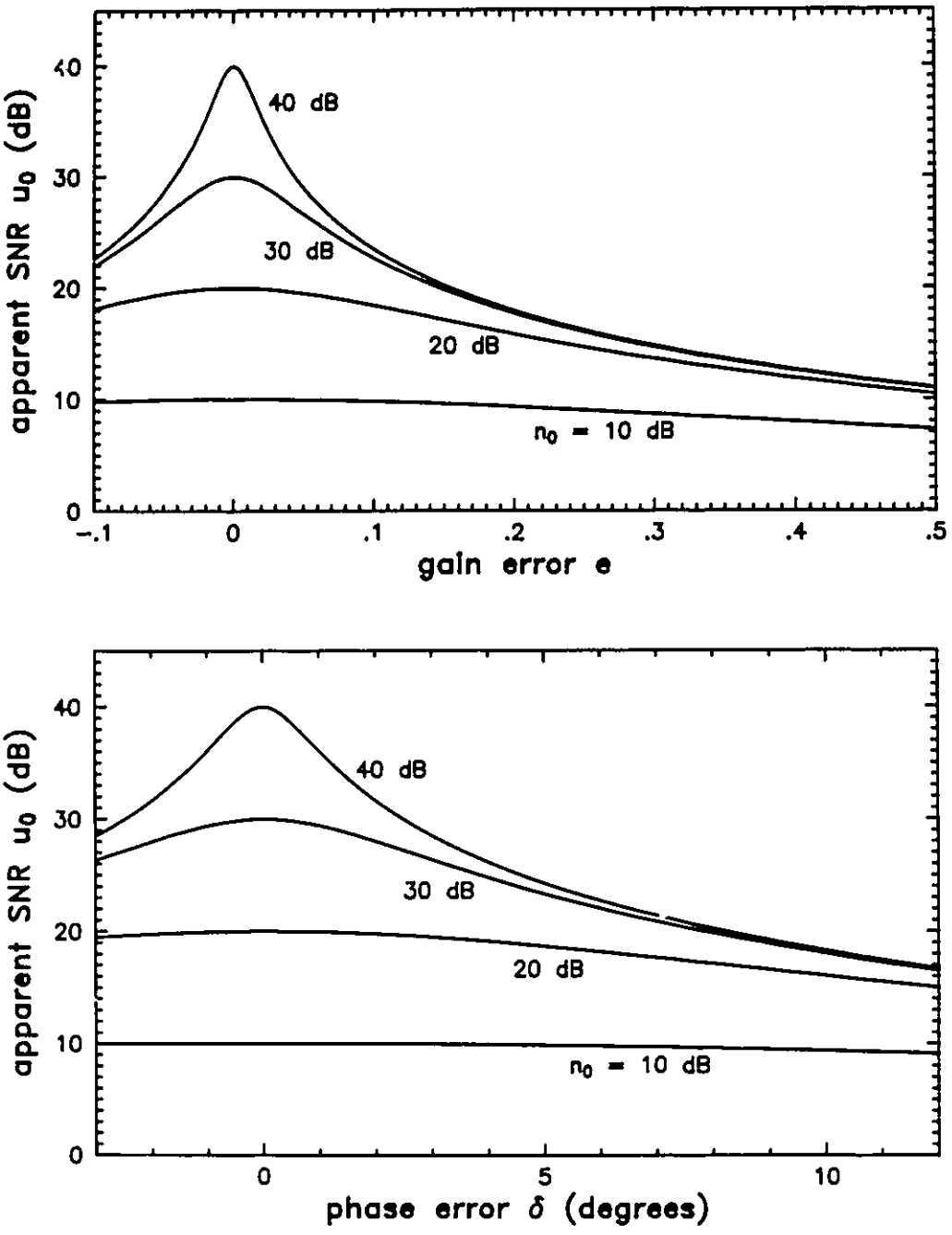


Figure E.5 Effects of gain and phase errors on the magnitude of ρ_1

Plot of the deviation of the apparent SNR u_0 from its true value n_0 as a function of gain error (top: $\delta = 0°$) and phase error (bottom: $e = 0$).

$$U = \frac{n_0}{(n_0+1)(k_e^2+1) + d_0(cos2\delta - k_e^2)} \tag{E.36}$$

$$V = \sqrt{[(k_e^2+1)cos\phi]^2 + [(2k_e cos\delta)sin\phi]^2} \tag{E.37}$$

thus

$$r' = UV; \quad \partial r' = U\partial V + V\partial U \tag{E.38}$$

and

$$\frac{\partial r'}{\partial k_e} = 2r'k_e \left[\frac{(k_e^2+1)cos^2\phi + (2cos^2\delta)sin^2\phi}{V^2} - \frac{U(n_0+1-d_0)}{n_0} \right] \tag{E.39}$$

$$\frac{\partial r'}{\partial \delta} = 2r'sin2\delta \left[\frac{k_e^2 sin^2\phi}{V^2} - \frac{Ud_0}{n_0} \right] \tag{E.40}$$

The sensitivity to variations in the phase error is seen to be very low around values of $\delta = 0$ regardless of any gain errors or differential noise values. The requirements for low sensitivity with respect to the gain error are somewhat tighter as is evident from Figure E.5. At $k_e = 1$ and $\delta = 0$

$$U = \frac{n_0}{2(n_0 + 1)} \quad \text{and} \quad V = 2$$

hence $r' = r$ and

$$\frac{\partial r'}{\partial k_e} = r \frac{d_0}{n_0 + 1} \tag{E.41}$$

In terms of the apparent SNR u_0 (Equation E.35) $\partial r' = (1-r')^2 \, \partial u_0$ and

$$\partial u_0 = n_0 d_0 \partial k_e \quad \text{or} \quad \partial SNR(\text{dB}) = 4.34 d_0 \partial k_e \tag{E.42}$$

Clearly a low sensitivity requires that the differential noise factor be zero. It is interesting to note that if d_0 is not zero, the magnitude of ρ_1' is not at its maximum when the gain and phase errors are zero (see Figure E.6). A higher apparent SNR will occur when the errors suppress the noisier channel to achieve a reduction in noise power W' which is greater than the increase in the false-target level. This phenomenon has an implication when trying to determine the correction factors as described in the next section.

The effects of these errors on the higher-order coefficients can be analyzed in a similar manner, but the computations become extremely long and tedious. The following heuristic arguments may provide some insights. It can be shown that the expected value of the reflection coefficient ρ_m for the single-complex-sinusoid-in-noise case is given by

$$\rho_m = -\frac{n_0}{mn_0+1} exp(jm\phi) \tag{E.43}$$

which provides an upper magnitude limit of $1/m$ for large signal-to-noise ratios. It is clear, however, that for m complex sinusoids ρ_m must be located much closer to the unit circle in order for all m spectral peaks to be resolved [Stehwien, 1983]. Since the phase errors can be modeled as the addition of a second sinusoid, ρ_2 will move close to the unit circle, especially for large SNRs. It is more difficult to assess the change in the phase angle of ρ_2, but it needs to move considerably in order to correctly position the second spectral peak. Experimental observations, albeit for a pulse modulated sinusoid, suggest that change magnitudes greater than $\pi/2$ are possible for typical error

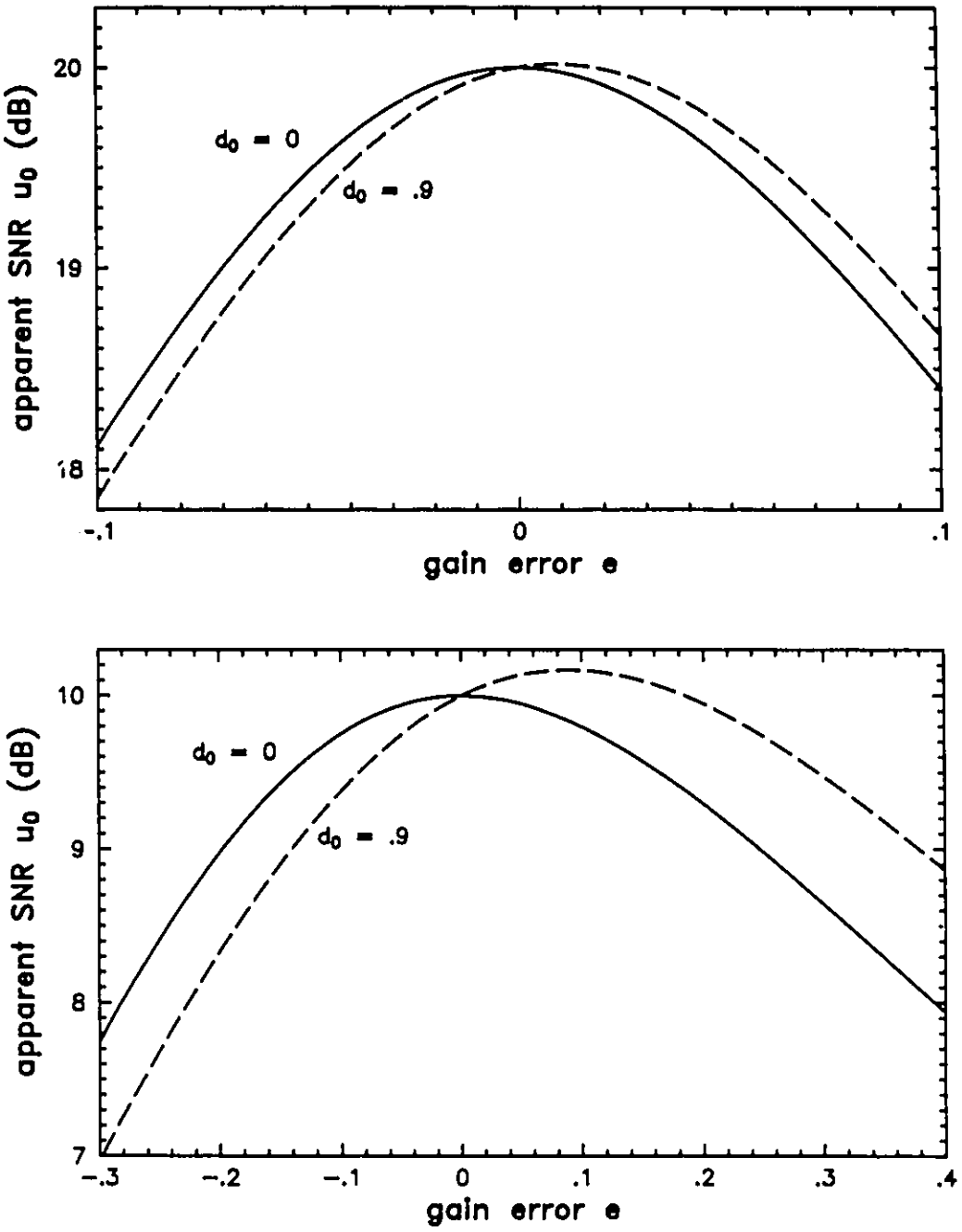


Figure E.6 Effects of unequal noise powers in the *I* and *Q* channels

The apparent SNR u_0 can exceed the true SNR n_0 at nonzero gain errors. This effect is most significant at low SNR values (bottom plot: n_0 = 10 dB, top plot: n_0 = 20 dB).

values. Because the true location of even higher-order coefficients is more and more obscured, as little error as possible must remain before attempting to characterize and classify the data. Any residual error will, however, place an upper limit on the order for which useful information remains accessible. Unless this limit can be found, very high-order coefficients must be used with caution.

E.3 Determining the Gain and Phase Error

The plots in Figures E.4 and E.5 imply that if the signal is known a priori to be a single complex sinusoid in noise, then a method for finding the correction factors k_1, k_2, and k_3 is to simply look for a maximum in r' or ϕ' relative to the unknown errors e and δ. A Doppler-shifted test signal would serve this purpose nicely – it allows direct and precise phase and gain measurements – but may not always be available. Actual targets, however, will be modulated not only with the azimuth beam pattern, but also with any internal target motion and scintillation. In practice they can still be used since such modulation will not by itself contribute to apparent false targets. Care must be taken, however, to ensure that the target Doppler will be completely within one side of the spectrum, that is no d.c. value or aliasing should be present. A frequency such that $\phi = \pm\pi/2$ is ideal. From Figure E.5 it is also clear that as high an SNR as possible should be used to take advantage of the very sharp peak in u_0. The presence of differential noise requires an adjustment of k_e after finding the maximum magnitude (ϕ' is independent of d_0: see Equation E.34). From Equation (E.39) the maximum of r' for $\delta = 0$ is found at

$$k_e^2 = \frac{n_0 + 1 + d_0(1 + \cot^2\phi)}{n_0 + 1 - d_0(1 + \cot^2\phi)} \tag{E.44}$$

At a phase angle equal to $\pi/2$ this reduces to

$$k_e^2 = \frac{n_0 + 1 + d_0}{n_0 + 1 - d_0} \qquad \text{(E.45)}$$

For large n_0 the peak in r occurs, as expected from Figure E.5, very close to $k_e = 1$ or $e = 0$. Hence steps 2 to 5 below are not required if $n_0 >> d_0(1+cot^2\phi)$, such that $k_m \approx 1$ within the required precision. With this knowledge the procedure for finding the correction factors can now be summarized:

1. Find values for k_e' and δ which maximize r or u_0 using Equations (E.4), (E.5), and (E.15). Conventional gradient or search techniques may be used. Minimizing $|\pm\pi/2 - \phi'|$ is not as effective since the peak is quite flat (see Figure E.4).
2. Initialize k_m to 1.
3. Set $k_e = k_m k_e'$ and compute $d_0 = (W_Q - W_I)/(W_Q + W_I)$ where $W_I = W_I'/k_e^2$. W_Q and W_I' are the measured relative noise powers.
4. Apply the updated correction factors to the data (Equation E.5) and recompute ρ_1, $cot\phi = \Re(\rho_1)/\Im(\rho_1)$, and $n_0 = |\rho_1|/(1-|\rho_1|)$.
5. Update k_m using Equation (E.44) and go back to step 3 until convergence is achieved.

Figure E.7 shows sample spectra before and after applying the correction factors obtained with this procedure on aircraft targets extracted from the recorded TRACS–ASR data. The factors varied somewhat from target to target and tape to tape,

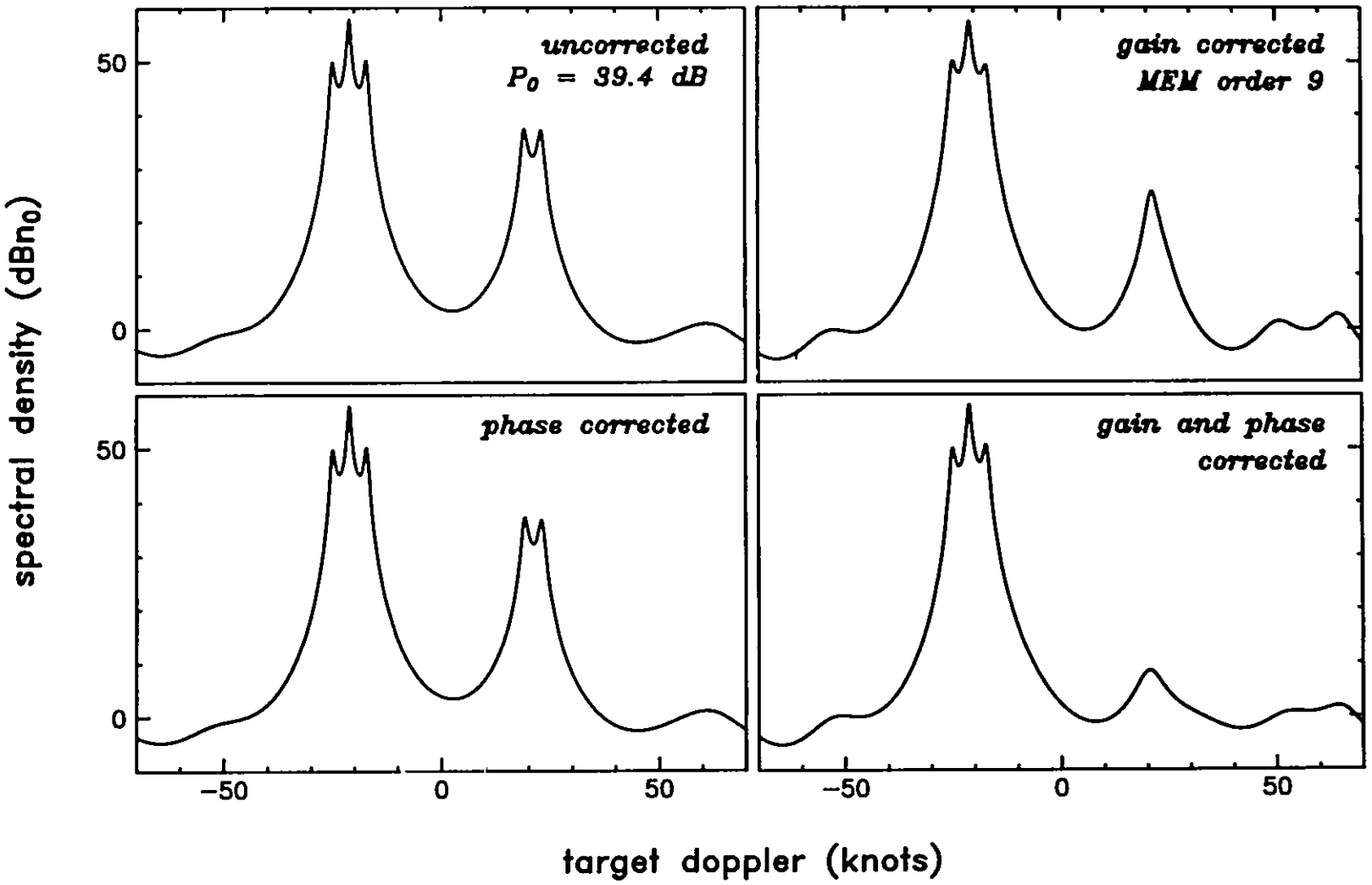


Figure E.7a Sample spectrum of an aircraft target before and after quadrature error correction.

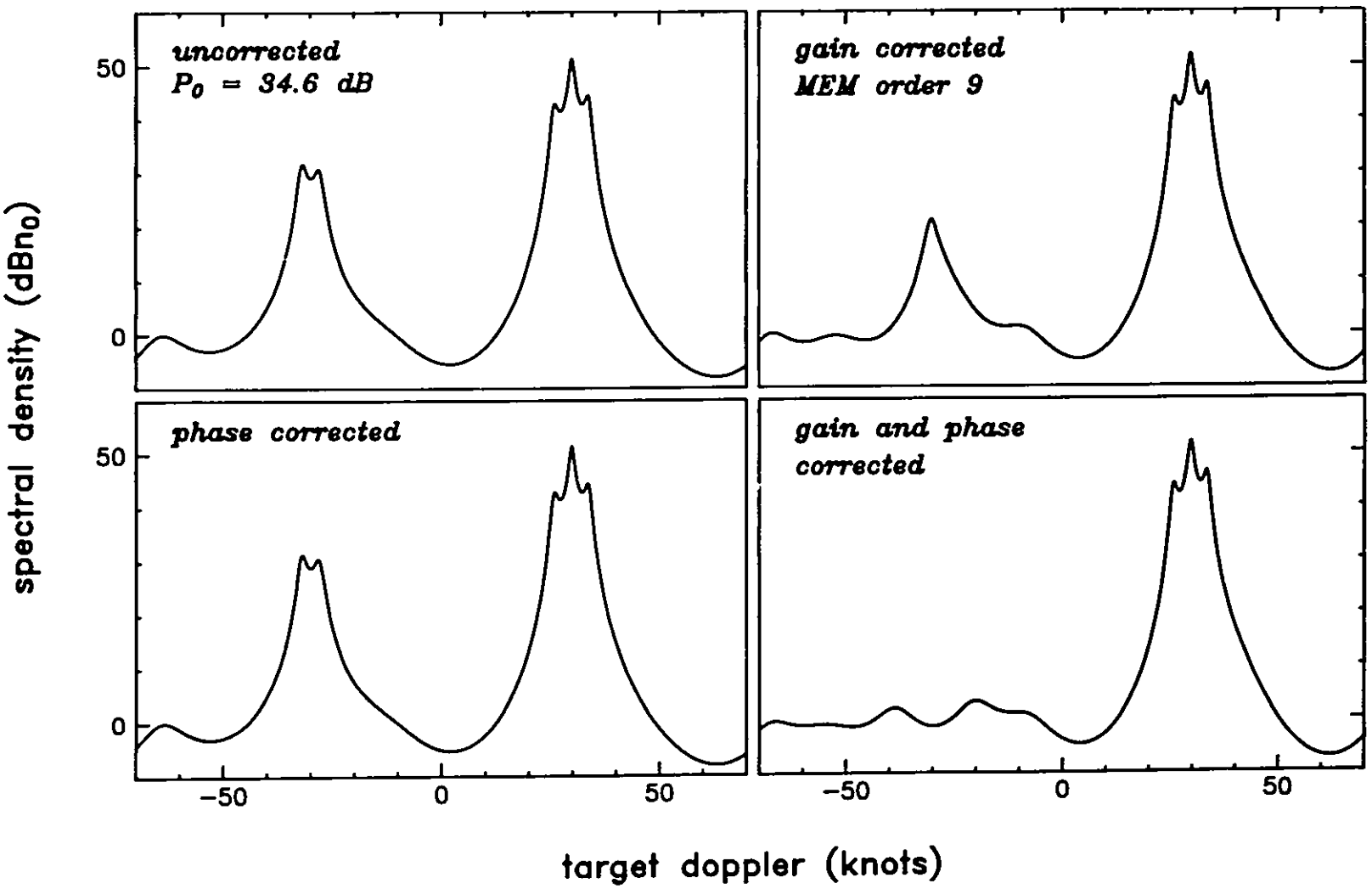


Figure E.7b Sample spectrum of an aircraft target before and after quadrature error correction.

but they were well within the regions of low sensitivity. All classification results presented in this thesis were based on the application of these factors to the data.

REFERENCES AND BIBLIOGRAPHY

Agnel, J.F. and E.N.S.T. "Modélisation autorégressive et classification de fouillis de sols. Application à la reconnaissance.", Proceedings of the International Conference on Radar, pp. 109–114, Versailles, France, May 1984

Agnel, J.F. "Elimination de Fouillis en Radar et Détection Séquentielle des Cibles: une approche par la reconnaissance des formes et le filtrage en treillis", These de Docteur Ingenieur, Ecole Nationale Superieure des Telecommunications, Paris, France, March 1985

Agrawala, A.K. (Ed.) Machine Recognition of Patterns, IEEE Press, New York, 1977

Anderson, T.W. An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, New York, 1984

Andrews, D.F. "A Note on the Selection of Data Transformations", Biometrika, vol. 58, no. 2, pp. 249–254, 1971

Andrews, D.F. "Plots of High-Dimensional Data", Biometrics, vol. 28, pp. 125–136, March 1972

Armstrong, G.M. and Donaldson, R.J. "Plan Shear Indicator for Real-Time Doppler Radar Identification of Hazardous Storm Winds", Journal of Applied Meteorology, vol. 8, pp. 376–383, June 1969

Atlas, D. "Radar Studies of Meteorological 'Angel' Echoes", Journal of Atmospheric and Terrestrial Physics, vol. 15, pp. 262–287, 1959

Atlas, D. "Possible Key to the Dilemma of Meteorological 'Angel' Echoes", Journal of Meteorology, vol. 17, no. 2, pp. 95–103, April 1960

Atlas, D., Browning, K.A., Donaldson, R.J., and Sweeney, H.J. "Automatic Digital Radar Reflectivity Analysis of a Tornadic Storm", Journal of Applied Meteorology, vol. 2, pp. 574–593, 1963

Atlas, D. and Wexler, R. "Wind Measurement by Coherent Radar with a Dual Beam Pattern", Journal of Applied Meteorology, vol. 4, pp. 598–606, 1965

Ballantyne, E.H. "The Addition of Height to a Weather Surveillance Radar PPI", Technical Report MWT–9, Stormy Weather Group, McGill University, Montreal, Quebec, December 1974

Barnett, V. (Ed.) Interpreting Multivariate Data, John Wiley & Sons, Chichester, 1981

Barry, J.R., Carter, B.K., Erdahl, R.J., Harris, R.L., Miller, J.T., Smith, G.D., and Barnes, R.M. "Angel Clutter and the ASR Air Traffic Control Radar", Final Report under Federal Aviation Administration contract DOT–FA72WA–2705, Applied Physics Laboratory, John Hopkins University, Silver Spring, Maryland, February 1973

Barton, D.K. (Ed.) Radars Volume 2: The Radar Equation, Artech House, Dedham, Massachusetts, 1974

Barton, D.K. (Ed.) Radars Volume 5: Radar Clutter, Artech House, Dedham, Massachusetts, 1975

Barton, D.K. (Ed.) Radars Volume 6: Frequency Agility and Diversity, Artech House, Dedham, Massachusetts, 1977

Barton, D.K. (Ed.) Radars Volume 7: CW and Doppler Radar, Artech House, Dedham, Massachusetts, 1978

Barton, D.K. "Land Clutter Models for Radar Designs and Analysis", Proceedings of the IEEE, vol. 73, no. 2, pp. 198–204, February 1985

Bellon, A. "Use of Digital Radar Data for a Short Term Precipitation Forecast", Scientific Report MW–80, Stormy Weather Group, McGill University, Montreal, Quebec, November 1973

Bendat, J.S. and Piersol, A.G. Engineering Applications of Correlation and Spectral Analysis, John Wiley and Sons, New York, 1980

Blackmer, R.H., Duda, R.O., and Reboh, R. "Application of Pattern Recognition Techniques to Digitized Weather Radar Data", Final Report Covering the Period 31 March 1972 to 30 April 1973 under NOAA National Weather Service Contract NWS 1–36072, Stanford Research Institute, Menlo Park, California, May 1973

Blackwell, F., Houghton, E.W., and Wilmot, T.A. "Analysis and Classification of Bird Flight and Echo Data Obtained by Radar", presented at the Bird Strike Committee Europe Conference, Frankfurt, June 1974

Blake, L.V. Radar Range–Performance Analysis, Lexington Books, Lexington, Massachusetts, 1980

Blokpoel, H. "Bird Migration Forecasts for Military Operations", Occasional Paper No. 16, Canadian Wildlife Service, Ottawa, 1973

Blokpoel, H. "Migration of Lesser Snow and Blue Geese in spring across southern Manitoba, Part 1: Distribution, chronology, directions, numbers, heights and speeds", Report Series Number 28, Canadian Wildlife Service, Ottawa, 1974

Blokpoel, H. and Gauthier, M.C. "Migration of Lesser Snow and Blue Geese in spring across southern Manitoba, Part 2: Influence of the weather and prediction of major flights", Report Series No. 32, Canadian Wildlife Service, Ottawa, 1975

Blokpoel, H., Heyland, J.D., Burton, J., and Samson, N. "Observations of the Fall Migration of Greater Snow Geese across Southern Quebec", Canadian Field–Naturalist, vol. 89, no. 3, pp. 268–277, 1975

Blokpoel, H. and Burton, J. "Weather and Height of Nocturnal Migration in Eastcentral Alberta: A Radar Study", Bird–Banding, vol. 46, no. 4, pp. 311–328, Autumn 1975

Blokpoel, H. "Radar Studies of Airborne Bird Movements", presented at the Workshop on Remote Sensing of Wildlife, Quebec City, November 1975

Blokpoel, H. Bird Hazards to Aircraft, Clarke, Irwin & Company Limited, Ottawa, 1976

Blokpoel, H. and Richardson, W.J. "Weather and Spring Migration of Snow Geese across Southern Manitoba", Oikos, vol. 30, no. 2, pp. 350–363, 1978

Blokpoel, H. and Gauthier, M.C. "Weather and the Migration of Canada Geese across Southeastern Ontario in Spring 1975", Canadian Field–Naturalist, vol. 94, no. 3, pp. 293–299, 1980

Box, G.E.P. and Jenkins, G.M. Time Series Analysis: Forecasting and Control, Holden-Day, Oakland, California, 1976

Brookner, E. (Ed.) Radar Technology, Artech House, Dedham, Massachusetts, 1977

Brookner, E. (Ed.) Aspects of Modern Radar, Artech House, Boston, Massachusetts, 1988

Burg, J.P. "Maximum Entropy Spectral Analysis", Proceedings of the 37th Meeting of the Society of Exploration Geophysicists, Oklahoma City, Oklahoma, October 1967

Burg, J.P. "A New Analysis Technique for Time Series Data", presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, Netherlands, 1968

Burg, J.P., Luenberger, D.G., and Wenger, D.L. "Estimation of Structured Covariance Matrices", Proceedings of the IEEE, vol. 70, no. 9, pp. 963–974, September 1982

Burington, R.S. and May, D.C. Handbook of Probability and Statistics with Tables, McGraw-Hill Book Company, New York, 1970

Chan, H.C. and Haykin, S. "An Adaptive MTI Processor Using FIR and IIR Filters", CRL Internal Report Series No. CRL–38, Communications Research Laboratory, McMaster University, Hamilton, Ontario, May 1976

Chan, H.C. and Haykin, S. "Applications of the Maximum Entropy Method in Radar Signal Processing", CRL Internal Report Series No. CRL–62, Communications Research Laboratory, McMaster University, Hamilton, Ontario, March 1979

Chen, C.H. "On Information and Distance Measures, Error Bounds, and Feature Selection", Information Sciences, vol. 10, pp. 159–173, 1976

Chen, C.H. "Spectral Resolution of Fougere's Maximum Entropy Spectral Analysis", Proceedings of the IEEE, vol. 69, no. 7, pp. 839–841, July 1981

Chen, C.H. Nonlinear Maximum Entropy Spectral Analysis Methods for Signal Recognition, Research Studies Press, Chichester, England, 1982

Childers, D.G. (Ed.) Modern Spectrum Analysis, IEEE Press, New York, 1978

Chow, C.K. "On Optimum Recognition Error and Reject Tradeoff", IEEE Transactions on Information Theory, vol. IT–16, no. 1, pp. 41–46, January 1970

Cook, C.E. and Bernfeld, M. Radar Signals, Academic Press, New York, 1967

Cox, D.R. and Small, N.J.H. "Testing Multivariate Normality", Biometrika, vol. 65, no. 2, pp. 263–272, 1978

Currie, B.W., Carter, C.R., and Haykin, S.S. "Signal Processing Applied to the Separation of Aircraft from Birds and Weather Using Air Traffic Control Radar", CRL Internal Report Series No. CRL–50, Communications Research Laboratory, McMaster University, Hamilton, Ontario, November 1977

Currie, B.W. and Haykin, S. "Radar Clutter Classifier Prototype for Use with the TRACS Airport Surveillance Radar, Part I", CRL Internal Report Series No. CRL–94, Communications Research Laboratory, McMaster University, Hamilton, Ontario, December 1981

Dennis, A.S., Smith, P.L., Boyd, E.I., and Musil, D.J. "Radar Observations of Hailstorms in Western Nebraska", Final report under National Science Foundation Grant GA–1518, Institute of Atmospheric Sciences, South Dakota School of Mines and Technology, Rapid City, South Dakota, March 1971

Dickinson, B.W. "Reflection Coefficient Estimation Using Cholesky Decomposition", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP–27, no. 2, pp. 146–149, April 1979

Dill, L.M. and Major, P.F. "Spatial Structure of Bird Flocks", Field Note No. 76, Associate Committee on Bird Hazards to Aircraft, National Research Council, Ottawa, June 1977

Donaldson, R.J. "Vortex Signature Recognition by a Doppler Radar", Journal of Applied Meteorology, vol. 9, pp. 661–670, August 1970

Draper, N.R. and C[illegible]R. "On Distributions and Their Transformation to Normality", J[illegible] of the Royal Statistical Society, vol. B–31, no. 3, pp. 472–476, [illegible]

Duda, R.O. and Hart, [illegible] Pattern Classification and Scene Analysis, John Wiley & Sons, New York, 1973

Duda, R.O. and Blackmer, R.H. "Application of Pattern Recognition Techniques to Digitized Weather Radar Data", Final Report Covering the Period 25 May 1971 to 31 March 1972 under NOAA National Weather Service Contract NWS 1–36072, Stanford Research Institute, Menlo Park, California, May 1973

Dunsmuir, W. "Estimation for Stationary Time Series When Data Are Irregularly Spaced or Missing" in Applied Time Series Analysis II, Findley, D.F. (Ed.), Academic Press, New York, 1981

Durrani, T.S. and Arslanian, A.S. "Signal Detection Performance of Lattice Processors", Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, pp. 703–706, Boston, Massachusetts, April 1983

Eastwood, E. Radar Ornithology, Methuen & Co Ltd, London, 1967

Eaves, J.L. and Reedy, E.K. (Ed.) Principles of Modern Radar, Van Nostrand Reinhold Company, New York, 1987

Fehlauer, J. and Eisenstein, B.A. "Feature Extraction by Identification of a Parameterized System Model", IEEE Transactions on Automatic Control, vol. AC–26, no. 2, pp. 577–580, April 1981

Feynman, J. and Fougere, P.F. "Eighty--Eight Year Periodicity in Solar-Terrestrial Phenomena Confirmed", Journal of Geophysical Research, vol. 89, no. A5, pp. 3023–3027, May 1984

Findley, D.F. (Ed.) Applied Time Series Analysis II, Academic Press, New York, 1981

Flock, W.L. and Green, J.L. "The Detection and Identification of Birds in Flight, Using Coherent and Noncoherent Radars", Proceedings of the IEEE, vol. 62, no. 6, pp. 745–753, June 1974

Foley, D.H. "Considerations of Sample and Feature Size", IEEE Transactions on Information Theory, vol. IT–18, no. 5, pp. 618–626, September 1972

Fougere, P.F., Zawalick, E.J., and Radoski, H.R. "Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis", Physics of the Earth and Planetary Interiors, vol. 12, pp. 201–207, 1976

Fougere, P.F. "A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis", Journal of Geophysical Research, vol. 82, no. 7, pp. 1051–1054, March 1977

Fougere, P.F. "A Solution to the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectrum Analysis of Complex Signals", Proceedings of the RADC Spectrum Estimation Workshop, pp. 77–84, Rome, New York, May 1978

Fougere, P.F. "Sunspots: Power Spectra and a Forecast", Solar–Terrestrial Predictions Proceedings, vol. 3, pp. A1–A9, December 1978

Fougere, P.F. "Spontaneous Line Splitting in Multichannel Maximum Entropy Power Spectra", First ASSP Workshop on Spectral Estimation, vol. 1, pp. 1.6.1–1.6.3, Hamilton, Ontario, August 1981

Fougere, P.F. "Spontaneous Line Splitting in Burg Maximum Entropy Power Spectra: A Review", Proceedings of the IEEE 2nd International Symposium on Computer Aided Seismic Analysis and Discrimination, pp. 76–77, August 1981

Fougere, P.F. "Estimation of the Power Spectrum of a Power-Law Process: FFT versus Maximum Entropy Technique", Proceedings of the IEEE 3rd International Symposium on Computer Aided Seismic Analysis and Discrimination, pp. 60–63, 1983

Fougere, P.F. "A Review of the Problem of Spontaneous Line Splitting in Maximum Entropy Power Spectral Analysis ", in Maximum–Entropy and Bayesian Methods in Inverse Problems, Smith, C.R. and Grandy, W.T. Jr. (Ed.), pp. 303–315, Reidel Publishing Company, 1985

Fougere, P.F. "On the Accuracy of Spectrum Analysis of Red Noise Processes Using Maximum Entropy and Periodogram Methods: Simulation Studies and Application to Geophysical Data", Journal of Geophysical Research, vol. 90, no. A5, pp. 4355–4366, May 1985

Fougere, P.F. "Maximum-Entropy Spectrum From a Non-Extendible Autocorrelation Function", Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1377–1379, Tokyo, Japan, April 1986

Fu, K.S. "Recent Developments in Pattern Recognition", IEEE Transactions on Computers, vol. C–29, pp. 845–854, 1980

Fukunaga, K. Introduction to Statistical Pattern Recognition, Academic Press, New York, 1972

Fukunaga, K. and Kessell, D.L. "Application of Optimum Error-Reject Functions", IEEE Transactions on Information Theory, vol. IT–18, no. 5, pp. 814–817, September 1972

Fukunaga, K. and Flick, T.E. "Classification Error for a Very Large Number of Classes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI–6, no. 6, pp. 779–788, November 1984

Fukunaga, K. and Flick, T.E. "A Test of the Gaussian–Ness of a Data Set Using Clustering", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI–8, no. 2, pp. 240–247, March 1986

Geotis, S.G. "Some Radar Measurements of Hailstorms", Journal of Applied Meteorology, vol. 2, pp. 270–280, April 1963

Gibson, J.D. "On Reflection Coefficients and the Cholesky Decomposition", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-25, no. 1, pp. 93–96, January 1977

Giri, N.C. Multivariate Statistical Inference, Academic Press, New York, 1977

Glick, N. "Additive Estimators for Probabilities of Correct Classification", Proceedings of the IEEE Conference on Pattern Recognition and Image Processing, pp. 304–317, Troy, New York, June 1977

Glick, N. "Additive Estimators for Probabilities of Correct Classification", Pattern Recognition, vol. 10, pp. 211–222, 1978

Glover, K.M. and Bishop, A.W. "Wind Measurement by Dual Beam Radar", Proceedings: 13th Radar Meteorology Conference, pp. 456–463, Montreal, August 1968

Glover, K.M. "A Precision Digital Video Integrator", Proceedings: 15th Radar Meteorology Conference, pp. 193–197, Champaign-Urbana, October 1972

Gnanadesikan, R. Methods for Statistical Data Analysis of Multivariate Observations, John Wiley & Sons, New York, 1977

Gunn, W.W.H. and Solman, V.E.F. "A Bird Warning System for Aircraft in Flight", in The Problems of Birds as Pests, Murton, R.K. and Wright, E.N. (Ed.), pp. 87–96, Academic Press, London, 1968

Hardy, K.R. "CPS-9-Radar Investigation of Clear-Air Convection", Proceedings: 13th Radar Meteorology Conference, pp. 236–241, Montreal, August 1968

Hardy, K.R. and Katz, I. "Probing the Clear Atmosphere with High Power, High Resolution Radars", Proceedings of the IEEE, vol. 57, no. 4, pp. 468–480, April 1969

Harrold, T.W. and Browning, K.A. "Identification of Preferred Areas of Shower Development by Means of High Power Radar", Quarterly Journal of the Royal Meteorological Society, vol. 97, pp. 330–339, 1971

Hawkes, C.D. and Haykin, S.S. "Modeling of Clutter for Coherent Pulsed Radar", IEEE Transactions on Information Theory, vol. IT–21, no. 6, pp. 703–706, November 1975

Hay, D.R. and Reid, W.M. "Radar Angels in the Lower Troposphere", Canadian Journal of Physics, vol. 40, pp. 128–138, 1962

Haykin, S.S. and Carter, C.R. "Radar Separation of Aircraft from Birds and Weather", CRL Internal Report Series No. CRL–28, Communications Research Laboratory, McMaster University, Hamilton, Ontario, April 1975

Haykin, S. and Kesler, S.B. "The Complex Form of the Maximum Entropy Method for Spectral Estimation", Proceedings of the IEEE, vol. 64, no. 5, pp. 822–823, May 1976

Haykin, S. "Spectral Estimation of Radar Clutter Using The Maximum Entropy Method", in Pattern Recognition and Signal Processing, Chen, C.H. (Ed.), pp. 323–366, Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1978

Haykin, S., Kesler, S.B., and Currie, B.W. "An Experimental Classification of Radar Clutter", Proceedings of the IEEE, vol. 67, no. 2, pp. 332–333, February 1979

Haykin, S. and Kesler, S. "Prediction-Error Filtering and Maximum-Entropy Spectral Estimation", in Nonlinear Methods of Spectral Analysis, Haykin, S. (Ed.), chapter 2, Springer-Verlag, Berlin, 1979

Haykin, S., Currie, B.W., and Kesler, S.B. "Maximum-Entropy Spectral Analysis of Radar Clutter", Proceedings of the IEEE, vol. 70, no. 9, pp. 953–962, September 1982

Haykin, S., Litva, J., Greenlay, T., and Stehwien, W. "Angle of Arrival Estimation Using Modified Forward-Backward Linear Prediction", Proceedings of the High Resolution Spatial Processing in Underwater Acoustics Workshop, Mississippi, October 1983

Haykin, S. "Radar Signal Processing", IEEE ASSP Magazine, vol. 2, no. 2, pp. 2–18, April 1985

Haykin, S. and Zhang, Q.T. "Model-dependent Signal Detection", Proceedings of the Third ASSP Workshop on Spectrum Estimation and Modeling, pp. 106–109, Boston, Massachusetts, November 1986

Hendry, A. and McCormick, G.C. "Polarization Properties of Precipitation Scattering", Bulletin of Radio And Electrical Engineering Division, National Research Council, vol. 21, no. 3, pp. 9–20, 1971

Hicks, J.J. and Angell, J.K. "Radar Observations of Breaking Gravitational Waves in the Visually Clear Atmosphere", Journal of Applied Meteorology, vol. 7, pp. 114–121, February 1968

Ho, Y.C. and Agrawala, A.K. "On Pattern Classification Algorithms – Introduction and Survey", Proceedings of the IEEE, vol. 56, no. 12, pp. 2101–2114, December 1968

Houghton, E.W., Blackwell, F., Ogilvie, M., and Wilmot, T.A. "A Radar Study of Wild Duck", presented at the 10th Bird Strike Committee Europe Conference, Stockholm, 1975

Howard, J.E., Birgenheier, R.A., and Young, G.O. "Applications of Learning Techniques to AMTI Radar Problems", Report No. AFAL–TR–72–240 Air Force Avionics Laboratory Contract No. F33615–71–C–1667, Project No. 4163, work performed by Hughes Aircraft Company, Culver City, California, June 1972

Hughes, G.F. "On the Mean Accuracy of Statistical Pattern Recognizers", IEEE Transactions on Information Theory, vol. IT–14, no. 1, pp. 55–63, January 1968

Hunt, F.R. "Automatic Radar Equipment to Determine Bird Strike Probability Part I. Night-time Passerine Migration", Field Note No. 69, Associate Committee on Bird Hazards to Aircraft, National Research Council, Ottawa, April 1975

Hunt, F.R. "The Probability of Bird-Aircraft Collisions based on Radar Data", Field Note No. 74, Associate Committee on Bird Hazards to Aircraft, National Research Council, Ottawa, November 1976

Hunt, F.R. "Automatic Radar Equipment to Determine Bird Strike Probability Part II. Migrating Water-Fowl Flocks", Field Note No. 75, Associate Committee on Bird Hazards to Aircraft, National Research Council, Ottawa, March 1977

Jain, A.K. and Waller, W.G. "On the Optimal Number of Features in the Classification of Multivariate Gaussian Data", Pattern Recognition, vol. 10, pp. 365–374, 1978

Jain, A.K. "Advances in Statistical Pattern Recognition", in Pattern Recognition Theory and Applications, Devijver, P.A. and Kittler, J. (Ed.), pp. 1–19, Springer-Verlag, Berlin, 1987

Jaynes, E.T. "On the Rationale of Maximum-Entropy Methods", Proceedings of the IEEE, vol. 70, no. 9, pp. 939–953, September 1982

Jenkins, G.M. and Watts, D.G. Spectral Analysis and its applications, Holden-Day, San Francisco, California, 1968

Jernigan, M.E. "Pattern Recognition", Course Notes, University of Waterloo, Waterloo, Ontario, September 1984

Jones, R.H. "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations", Technometrics, vol. 22, pp. 389–395, 1980

Jones, R.H. "Fitting a Continuous Time Autoregression to Discrete Data" in Applied Time Series Analysis II, Findley, D.F. (Ed.), Academic Press, New York, 1981

Kalayeh, H.M. and Landgrebe, D.A. "Predicting the Required Number of Training Samples", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI–5, no. 6, pp. 664–667, November 1983

Kalbfleisch, J.G. Probability and Statistical Inference, Volume 1: Probability, Springer-Verlag, New York, 1985

Kalmus, H.P. "Doppler Wave Recognition with High Clutter Rejection", IEEE Transactions on Aerospace and Electronic Systems (supplement), vol. AES-3, no. 6, pp. 334–339, November 1967

Kanal, L. "Patterns in Pattern Recognition: 1968 – 1974", IEEE Transactions on Information Theory, vol. IT–20, no. 6, pp. 697–722, November 1974

Kashyap, R.L. "Optimal Feature Selection and Decision Rules in Classification Problems with Time Series", IEEE Transactions on Information Theory, vol. IT–24, no. 3, pp. 281–288, May 1978

Kay, S.M. and Marple, S.L. "Spectrum Analysis - A Modern Perspective", Proceedings of the IEEE, vol. 69, no. 11, pp. 1380–1419, November 1981

Kay, S.M. and Makhoul, J. "On the Statistics of the Estimated Reflection Coefficients of an Autoregressive Process", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-31, no. 6, pp. 1447–1455, December 1983

Kesler, S.B. "Nonlinear Spectral Analysis of Radar Clutter", CRL Internal Report Series No. CRL–51, Communications Research Laboratory, McMaster University, Hamilton, Ontario, January 1978

Kesler, S.B. and Haykin, S. "The Maximum Entropy Method Applied to the Spectral Analysis for Radar Clutter", IEEE Transactions on Information Theory, vol. IT-24, no. 2, pp. 269–272, March 1978

Klass, P.J. "FAA to Evaluate New Air Route Radar", Aviation Week and Space Technology, vol. 99, no. 16, pp. 51–53, 15 October 1973 and:
Richardson, W.J. "Letters to the Editor: Radar 'Angels'", vol. 99, no. 22, p. 76, 26 November 1973

Kleiner, B., Martin R.D., and Thomson D.J. "Robust Estimation of Power Spectra", The Journal of the Royal Statistical Society Series B, vol. 41, no. 3, pp. 313–351, 1979

Kres, H. Statistical Tables for Multivariate Analysis, Springer-Verlag, New York, 1983

Krishnaiah, P.R. (Ed.) Multivariate Analysis III, Academic Press, New York, 1973

Krishnaiah, P.R. (Ed.) Multivariate Analysis IV, North-Holland Publishing Company, Amsterdam, 1977

Krishnaiah, P.R. and Kanal, L.N. (Ed's.) Handbook of Statistics, Volume 2: Classification, Pattern Recognition, and Reduction of Dimensionality, North-Holland Publishing Company, Amsterdam, 1982

Lachenbruch, P.A. and Mickey, R.M. "Estimation of Error Rates in Discriminant Analysis", Technometrics, vol. 10, pp. 1–11, 1968

Lhermitte, R.M. "Note on the Observation of Small-Scale Atmospheric Turbulence by Doppler Radar Techniques", Radio Science, vol. 4, no. 12, pp. 1241–1246, December 1969

Lissack, T. and Fu, K.S. "Error Estimation in Pattern Recognition via L^{α}-Distance Between Posterior Density Functions", IEEE Transactions on Information Theory, vol. IT–22, no. 1, pp. 34–45, January 1976

Long, M.W. Radar Reflectivity of Land and Sea, Lexington Books, Lexington, Massachusetts, 1975

MacGregor, J.F. "Optimal Choice of the Sampling Interval for Discrete Process Control", Technometrics, vol. 18, no. 2, pp. 151–160, May 1976, and Erratum, vol 19, no. 2, p. 224, May 1977

Macikunas, A. and Haykin, S. "Polarization as a Radar Discriminant" in Selected Topics in Signal Processing, Haykin, S. (Ed.), chapter 11 (pp 251–286), Prentice Hall, Englewood Cliffs, New Jersey, 1989

Markel, J.D. and Gray, A.H.,Jr. Linear Prediction of Speech, Springer-Verlag, Berlin, 1976

Marshall, J.S. and Hitchfeld, W. "Interpretation of the Fluctuating Echo from Randomly Distributed Scatterers. Part I", Canadian Journal of Physics, vol. 31, pp. 962–1009, 1953

Marshall, J.S. and Ballantyne, E.H. "Weather Surveillance Radar", Scientific Report MW–81, Stormy Weather Group, McGill University, Montreal, Quebec, November 1973

Martin, R.D. and Thomson, D.J. "Robust-Resistant Spectrum Estimation", Proceedings of the IEEE, vol. 70, no. 9, pp. 1097–1115, September 1982

Masry, E., Klamer, D., and Mirabile, C. "Spectral Estimation of Continuous-Time Processes: Performance Comparison Between Periodic and Poisson Sampling Schemes", IEEE Transaction on Automatic Control, vol. AC-23, no. 4, pp. 679–685, August 1978

Metford, P.A.S. "An Innovations Approach to Discrete-Time Detection Theory with Application to Radar", Ph.D. Thesis, McMaster University, Hamilton,Ontario, September 1984

Metford, P.A.S. and Haykin, S. "Experimental Analysis of an Innovations-based Detection Algorithm for Surveillance Radar", IEE Proceedings, vol. 132, Pt. F, no. 1, pp. 18–26, February 1985

Meyer, D.P. and Mayer, H.A. Radar Target Detection, Academic Press, New York, 1973

Miller, L.J. and Strauch, R.G. "A Dual Doppler Radar Method for the Determination of Wind Velocities within Precipitation Weather Systems", Remote Sensing of the Environment, vol. 3, pp. 219–235, 1974

Moorcroft, D.R. "Maximum-Entropy Spectral Analysis of Radio-Auroral Signals", Radio Science, vol. 13, no. 4, pp. 649–660, July-August 1978

Muehe, C.E., Drury, W.H., Hofstetter, E.M., Labitt, M., McCorison, P.B., and Sferrino, V.J. "New Techniques Applied to Air-Traffic Control Radars", Proceedings of the IEEE, vol. 62, no. 6, pp. 716–723, June 1974

Murton, R.K. and Wright, E.N. (Ed.) The problems of Birds as Pests, Academic Press, London, 1968

Nagy, G. "Candide's Practical Principles of Experimental Pattern Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI–5, no. 2, pp. 199–200, March 1983

Nathanson, F.E. and Reilly, J.P. "Radar Precipitation Echoes: Experiments on Temporal, Spatial, and Frequency Correlation", IEEE Transactions on Aerospace and Electronic Systems, vol. AES–4, no. 4, pp. 505–514, July 1968

Nathanson, F.E. Radar Design Principles, McGraw-Hill Book Company, New York, 1969

Ogura, H. and Yoshida, Y. "Spectral Analysis and Subtraction of Noise in Radar Signals", IEEE Transactions on Aerospace and Electronic Systems, vol. AES-17, no. 1, pp. 62–71, January 1981

Papoulis, A. Probability, Random Variables, and Stochastic Processes, McGraw-Hill Book Company, New York, 1955

Patrick, E.A. Fundamentals of Pattern Recognition, Prentice Hall, Inc., Englewood Cliffs, N.J., 1972

Plank, V.G. "A Meteorological Study of Radar Angels", Geophysical Research Papers No. 52, Geophysics Research Directorate, Air Force Cambridge Research Center, Bedford, Massachusetts, July 1956

Pollen, G.E. "Topics in Radar Clutter Modeling", Course Lecture Notes, Technology Service Corporation, Santa Monica, California, March 1973

Rabinowitz, S.J., Gager, C.H., Brookner, E., Muehe, C.E., and Johnson, C.M. "Applications of Digital Technology to Radar", Proceedings of the IEEE, vol. 73, no. 2, pp. 325–339, February 1985

Radoski, H.R., Fougere, P.F., and Zawalick, E.J. "A High Resolution Power Spectral Estimate: The Maximum Entropy Method", Air Force Surveys in Geophysics, no. 282, Air Force Cambridge Research Laboratories, Bedford, Massachusetts, February 1974

Radoski, H.R., Fougere, P.F., and Zawalick, E.J. "A Comparison of Power Spectral Estimates and Applications of the Maximum Entropy Method", Journal of Geophysical Research, vol. 80, no. 4, pp. 619–625, February 1975

Radoski, H.R., Zawalick, E.J., and Fougere, P.F. "The Superiority of Maximum Entropy Power Spectrum Techniques Applied to Geomagnetic Micropulsations", Physics of the Earth and Planetary Interiors, vol. 12, pp. 208–216, 1976

Richardson, W.J. and Haight, M.E. "Migration Departures from Starling Roosts", Canadian Journal of Zoology, vol. 48, no. 1, pp. 31–39, 1970

Richardson, W.J. "Spring Migration and Weather in Eastern Canada: A Radar Study", American Birds, vol. 25, no. 4, pp. 684–690, August 1971

Richardson, W.J. "Autumn Migration and Weather in Eastern Canada: A Radar Study", American Birds, vol. 26, no. 1, pp. 10–17, February 1972

Richardson, W.J. "Spring Migration over Puerto Rico and the Western Atlantic, A Radar Study", Ibis, vol. 116, pp. 172–193, 1974

Richardson, W.J. "Multivariate Approaches to Forecasting Day-to-day Variations in the Amount of Bird Migration", Proceedings Conference on the Biological Aspects of the Bird/Aircraft Collision Problem, pp. 309–329, Clemson, South Carolina, February 1974

Richardson, W.J. "Timing and Amount of Bird Migration in Relation to Weather: A Review", Oikos, vol. 30, no. 2, pp. 224–272, 1978

Richardson, W.J. "Southeastward Shorebird Migration over Nova Scotia and New Brunswick in autumn: A Radar Study", Canadian Journal of Zoology, vol. 57, pp. 107–124, 1979

Rihaczek, A.W. "The Maximum Entropy of Radar Resolution", IEEE Transactions on Aerospace and Electronic Systems, vol. AES-17, no. 1, p. 144, January 1981, and:
Jackson, P.L. "Radar Superresolution? Perhaps!" and Rihaczek, A.W. "Comments on 'Radar Superresolution? Perhaps!'", vol. AES-17, no. 5, pp. 734–735, September 1981

Riley, J.R. "Radar Cross Section of Insects", Proceedings of the IEEE, vol. 73, no. 2, pp. 228–232, February 1985

Rogers, R.R. "The Effect of Variable Target Reflectivity on Weather Radar Measurements", Quarterly Journal of the Royal Meteorological Society, vol. 97, pp. 154–167, 1971

Schaefer, G.W. "Bird Recognition by Radar: A Study in Quantitative Radar Ornithology", in The Problems of Birds as Pests, Murton, R.K. and Wright, E.N. (Ed.), pp. 53–86, Academic Press, London, 1968

Schaffner, M. "A Digital Sweep Integrator for Weather Radar", Proceedings: 12th Conference on Radar Meteorology, pp. 28–33, Norman, Oklahoma, October 1966

Schleher, D.C. (Ed.) MTI Radar, Artech House, Dedham, Massachusetts, 1978

Shore, J.E. "Minimum Cross-Entropy Spectral Analysis", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-29, no. 2, pp. 230–237, April 1981

Sinsky, A.I. and Wang, C.P. "Error Analysis of a Quadrature Coherent Detector Processor", IEEE Transactions on Aerospace and Electronic Systems, vol. AES–10, no. 6, pp. 880–883, November 1974

Skolnik, M.I. (Ed.) Radar Handbook, McGraw-Hill Book Company, New York, 1970

Skolnik, M.I. Introduction to Radar Systems, McGraw-Hill Book Company, New York, 1980

Skolnik, M.I. "Fifty Years of Radar", Proceedings of the IEEE, vol. 73, no. 2, pp. 182–197, February 1985

Smith, P.L., Hardy, K.R., and Glover, K.M. "Applications of Radar to Meteorological Operations and Research", Proceedings of the IEEE, vol. 62, no. 6, pp. 724–745, June 1974

Solman, V.E.F. (Ed.) "Studies of Bird Hazards to Aircraft", Report Series Number 14, Canadian Wildlife Service, Ottawa, 1971

Stables, E.R. and New, N.D. "Birds and Aircraft: The Problems", in The Problems of Birds as Pests, Murton, R.K. and Wright, E.N. (Ed.), pp. 3–16, Academic Press, London, 1968

Stearns, S.D. "On Selecting Features for Pattern Classifiers", Proceedings of the Third International Joint Conference on Pattern Recognition, pp. 71–75, Coronado, California, November 1976

Stehwien, W. "Parametric Spectral Analysis of Radar Clutter", M.Eng. Thesis, McMaster University, Hamilton, Ontario, September 1983

Stehwien, W. and Haykin, S. "Parametric Analysis of Radar Clutter", Proceedings of the International Electrical, Electronics Conference, vol. 1, pp. 136–139, Toronto, Ontario, September 1983

Stehwien, W. and Haykin, S. "Classification of Radar Clutter Based on Spectral Parameters", Proceedings of the International Conference on Radar, pp. 126–131, Versailles, France, May 1984

Stehwien, W. "A Clutter Classifier - Preliminary Results", CRL Internal Report Series No. CRL–135, Communications Research Laboratory, McMaster University, Hamilton, Ontario, December 1984

Stehwien, W. and Haykin, S. "Experimental Radar Clutter Data Evaluation for Clutter Classification Research", CRL Internal Report Series No. CRL–152, Communications Research Laboratory, McMaster University, Hamilton, Ontario, January 1986

Stehwien, W. and Haykin, S. "Statistical Classification of Radar CLutter", Proceedings of the 1986 IEEE National Radar Conference, pp. 101–106, Los Angeles, California, March 1986

Stehwien, W. and Haykin, S. "A Statistical Radar Clutter Classifier", Proceedings of the 1989 IEEE National Radar Conference, pp. 164–169, Dallas, Texas, March 1989

Taylor, J.W. and Brunins, G. "Design of a New Airport Surveillance Radar (ASR-9)", Proceedings of the IEEE, vol. 73, no. 2, pp. 284–289, February 1985

Thomson, D.J. "Spectrum Estimation Techniques for Characterization and Development of WT4 Waveguide - I", The Bell System Technical Journal, vol. 56, no. 9, pp. 1769–1815, November 1977

Thomson, D.J. "Spectrum Estimation Techniques for Characterization and Development of WT4 Waveguide - II", The Bell System Technical Journal, vol. 56, no. 10, pp. 1983–2005, December 1977

Thomson, D.J. "Spectrum Estimation and Harmonic Analysis", Proceedings of the IEEE, vol. 70, no. 9, pp. 1055–1096, September 1982

Tjostheim, D. "Recognition of Waveforms Using Autoregressive Feature Extraction", IEEE Transactions on Computers, vol. C–26, pp. 268–270, March 1977

Toussaint, G.T. "An Upper Bound on the Probability of Misclassification in Terms of the Affinity", Proceedings of the IEEE, vol. 65, no. 2, pp. 275–276, March 1977

Toussaint, G.T. "Probability of Error, Expected Divergence, and the Affinity of Several Distributions", Proceedings of the IEEE Conference on Pattern Recognition and Image Processing, pp. 299–303, Troy, New York, June 1977

Tufts, D.W. and Kumaresan, R. "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood", Proceedings of the IEEE, vol. 70, no. 9, pp. 975–989, September 1982

Ulrych, T.J. and Bishop T.N. "Maximum Entropy Spectral Analysis and Autoregressive Decomposition", Reviews of Geophysics and Space Physics, vol. 13, no. 1, pp. 183–200, February 1975

van Campenhout, J.M. "On the Peaking of the Hughes Mean Recognition Accuracy: The Resolution of an Apparent Paradox", IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-8, no. 5, pp. 390–395, May 1978

van den Bos, A. "Alternative Interpretation of Maximum Entropy Spectral Analysis", IEEE Transactions on Information Theory, vol. IT-17, no. 4, pp. 493–494, July 1971

van Trees, H.L. Detection, Estimation, and Modulation Theory, Part III: Radar–Sonar Signal Processing and Gaussian Signals in Noise, John Wiley & Sons, New York, 1971

Vaughn, C.R. "Birds and Insects as Radar Target: A Review", Proceedings of the IEEE, vol. 73, no. 2, pp. 205–227, February 1985

Watanabe, H., Kirimoto, T., and Kondo, M. "Trade-off Studies on Adaptive Filtering Algorithms Used for Adaptive Moving Target Indicators", Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, Tokyo, Japan, April 1986

Weber, P. and Haykin, S. "Detailed Survey of Pulse Doppler Radar Techniques", CRL Internal Report Series No. CRL–109, Communications Research Laboratory, McMaster University, Hamilton, Ontario, March 1983

Wilk, M.B. and Gnanadesikan, R. "Probability Plotting Methods for the Analysis of Data", Biometrika, vol. 55, no. 1, pp. 1–17, 1968

Wilson, J., Carbone, R., Baynton, H., and Serafin, R. "Operational Application of Meteorological Doppler Radar", Bulletin of the American Meteorological Society, vol. 61, no. 10, pp. 1154–1168, October 1980

Wozencraft, J.M. and Jacobs, I.M. Principles of Communication Engineering, John Wiley & Sons, New York, 1965

Young, I.T. "Further Consideration of Sample and Feature Size", IEEE Transactions on Information Theory, vol. IT–24, no. 6, pp. 773–775, November 1978

Young, T.Y, and Calvert, T.W. Classification, Estimation and Pattern Recognition, American Elsevier Publishing Company, Inc., New York, 1974

Young, T.Y. and Fu, K.S. (Ed's.) Handbook of Pattern Recognition and Image Processing, Academic Press, Orlando, Florida, 1986

Zawadzki, I.I. and Rogers, R.R. "Two-Dimensional Spectra of Precipitation Patterns by Coherent Optical Analysis", Technical Report MWT–5, Stormy Weather Group, McGill University, Montreal, Quebec, September 1969

Zawadzki, I.I. "Statistical Properties of Precipitation Patterns", Journal of Applied Meteorology, vol. 12, pp. 459–472, April 1973

Zrnic, D.S. and Doviak, R.J. "Velocity Spectra of Vortices Scanned with a Pulse-Doppler Radar", Journal of Applied Meteorology, vol. 14, pp. 1531–1539, December 1975

"United to Alter Radar to Spot Turbulence", Aviation Week and Space Technology, vol. 118, no. 16, p. 196, 18 April 1983