AN INNOVATIONS APPROACH

TO

DISCRETE-TIME DETECTION THEORY

WITH

APPLICATION TO RADAR

BY

© Peter Aish Seymour Metford, B.Sc., M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

September 1984

AN INNOVATIONS APPROACH TO DISCRETE-TIME DETECTION THEORY WITH APPLICATION TO RADAR

DOCTOR OF PHILOSOPHY (1984)
(Electrical Engineering)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: AN INNOVATIONS APPROACH TO DISCRETE-TIME DETECTION THEORY WITH APPLICATION TO RADAR

AUTHOR: Peter Aish Seymour Metford, B.Sc. (University of Western Ontario), M.Sc. (McMaster University)

SUPERVISOR: Dr. S. Haykin

NUMBER OF PAGES: xiii, 148

ABSTRACT

A very rapidly convergent solution (in the form of a likelihood ratio test) for the problem of detecting a discrete-time stochastic process in additive white Gaussian noise is derived.

This likelihood ratio test is then applied to the problem of moving-target (aircraft) detection by airport surveillance radar systems. Using real radar data, the receiver operating characteristics are obtained for two different adaptive implementations of this likelihood ratio test, and also for the three versions of the Moving Target Detection algorithms presently in use in modern radar systems.

The better of the two adaptive implementations employs Kalman prediction error tapped delay-line filters and attains a minimum of 3 dB average performance improvement relative to the Moving Target Detection algorithms.

ACKNOWLEDGEMENTS

I want to thank my supervisor, Dr. Simon Haykin for his unfailing support and encouragement. I consider it a privilege to have worked with Simon.

A great deal of appreciation is due to Dr. Desmond Taylor for his contructive remarks and suggestions concerning the mathematical portion of this thesis.

I also wish to thank Dr. R. deBuda, the other member of my Ph.D. supervisory committee, for his helpful comments.

The cooperation given me by Mr. J. Taylor of Westinghouse and Mr. C. Muehe of Lincoln Laboratories in the release of the design details of the MTD algorithms is acknowledged.

The trials and tribulations of the TRACS technical support staff at CFB Trenton during the data recording are appreciated.

The assistance given by the CRL engineering and technical staff must be recognized.

Also are my thanks to my colleagues, from amongst whom we made a lot of friends, and from whom I learned a great deal.

Finally, I wish simply to thank my wife Aggie.

TO AGGIE

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AP	The Floating Point Systems AP-120B Array Processor
AR	Autoregressive process
ARMA	Autoregressive Moving Average process
ASR	Airport Surveillance Radar
AWGN	Additive White Gaussian Noise
CFAR	Constant False Alarm Rate
CPI	Coherent Processing Interval
CRL	Communications Research Laboratory
dB	Decibel
FFT	Fast Fourier Transform
GPIOP	The Floating Point Systems General Programmable Input-Output Processor
IBDA	Innovations Based Detection Algorithm
IR	Innovations Representation
LLR	Natural Logarithm of the Likelihood Ratio
LR	Likelihood Ratio
LSL	Least Squares Lattice
MMSE	Minimum Mean-Square Error
MTD	Moving Target Detector
MTI	Moving Target Indicator
pdf	Probability Density Function

P_D	Probability of Detection
P_{FA}	Probability of False Alarm
ROC	Receiver Operating Characteristic
SCNR	Signal Power to (Clutter plus Receiver Noise) Power Ratio
TDL	Tapped Delay Line Filter
WGN	White Gaussian Noise
ZVF	Zero Velocity Filter

LIST OF SYMBOLS

$\{x(n)\}$, $\{y(n)\}$, $\{z(n)\}$ - arbitrary stochastic processes

$\{s(n)\}$ - a signal process

$\{c(n)\}$ - a (clutter) stochastic process

$\{w(n)\}$ - a white Gaussian noise process

$\{\nu(n)\}$ - an innovations process

$\{\tilde{x}(n|H_i)\}$ - a prediction error process (assuming Hypothesis H_i)

$\{\rho(n)\}$ - a variance-normalized partial sum sequence of the innovations process

$\{\rho'(n)\}$ - a variance-normalized partial sum sequence of a stationary innovations sequence

$\hat{x}(n|H_i)$ - the optimal prediction (in a MMSE sense) of $x(n)$ assuming H_i

$\{\xi(n)\}$ - a conditionally-known process

$\{h(n)\}$ - the set of taps for a tapped delay-line filter

$\underline{R}_X$ - the autocorrelation matrix of the stochastic process $x(n)$

$\underline{R}_{XY}$ - the cross-correlation matrix of the stochastic processes $\{x(n)\}$ and $\{y(n)\}$

$\underline{C}$ - the covariance matrix of the scaled partial sums sequence of the WGN process $\{w(n)\}$

σ_w^2	-	the variance of the WGN process $\{w(n)\}$
$\sigma_s^2(n)$	-	the variance of the sum of the first n terms of the innovations process
$\sigma^2(n\|H_i)$	-	the prediction error power given Hypothesis H_i
$\sigma_b^2(n)$	-	variance of a backward prediction error process
$\underline{a}$	-	a vector of forward linear prediction coefficients
$\underline{a}'$	-	a time reversed vector of forward linear prediction coefficients
$\underline{b}$	-	a vector of backward linear prediction coefficients
$\underline{b}'$	-	a time reversed vector of backward linear prediction coefficients
M	-	the order of a linear prediction error filter
μ	-	the damping constant employed in the LSL algorithm
r,q	-	the variances of white noise processes employed in the Kalman prediction error filter
E[ ]	-	a probabilistic expectation
P,Q,R	-	probability measures
P[ ]	-	the probability of an event under the probability measure P

$P(\ )$	-	a probability distribution function
$p(\ )$	-	a probability density function
$\phi(x)$	-	the zero-mean, unit-variance Gaussian distribution function
Λ	-	a likelihood ratio
$\mathcal{F}$	-	a sigma-field of events
$\stackrel{\Delta}{=}$	-	"is defined as"

LIST OF FIGURES

Figure		Page
5.7	The ROC for the MTD-1 (weather clutter dominated radar data)	115
5.8	The ROC for the MTD-2 (weather clutter dominated radar data)	116
5.9	The ROC for the MTD-3 (weather clutter dominated radar data)	117
5.10	The ROC for the IBDA (LSL) (weather clutter dominated radar data)	118
5.11	The ROC for the IBDA (KALMAN) (weather clutter dominated radar data)	119
5.12	The ROC for the MTD-1 (ground clutter dominated radar data)	120
5.13	The ROC for the MTD-2 (ground clutter dominated radar data)	121
5.14	The ROC for the MTD-3 (ground clutter dominated radar data)	122
5.15	The ROC for the IBDA (LSL) (ground clutter dominated radar data)	123
5.16	The ROC for the IBDA (KALMAN) (ground clutter dominated radar data)	124
5.17	The performance comparisons relative to the MTD-1 (weather clutter dominated radar data)	125
5.18	The performance comparisons relative to the MTD-1 (ground clutter dominated radar data)	126

CHAPTER 1

INTRODUCTION

1.1 A Statement of the Research Problem

The purpose of this research is to solve the generic problem of the detection of a discrete-time, not necessarily Gaussian, stochastic process imbedded in additive white Gaussian noise (AWGN). This specific problem is chosen since it may be used as a basic building block in the construction of the detection problems inherent in many discrete-time communications systems, pulsed radar and sonar systems.

This detection problem is formulated in terms of statistical hypothesis testing where

Hypothesis H_1: received data = the stochastic signal process plus white Gaussian noise (1.1a)

Hypothesis H_o: received data = white Gaussian noise (WGN) (1.1b)

A solution to this problem must decide, on the basis of the received data, which of the two hypotheses is correct.

It is not sufficient merely to obtain a solution to this problem. The performance of the derived detection algorithm must be evaluated and compared to existing high performance detection algorithms. These performance comparisons are made by applying the detection algorithm to a specific detection problem for which existing detection algorithms are available.

The detection problem which was chosen for performance evaluations is the detection of aircraft by airport surveillance radar (ASR) systems. This problem was selected for several reasons. First: it utilizes the previously mentioned building block concept. Second: high performance detection algorithms (currently in use on some ASR systems) were made available for the performance comparisons. Third: the statistics of the received data are highly dynamic, which tests the robustness of the adaptive implementation of the derived detection algorithm. Fourth: the statistics of the received data process can not be assumed to be Gaussian. Fifth: real data is used for the performance analyses, resulting in strong and meaningful conclusions.

1.2 The Contributions of the Thesis

The thesis may be conceptually divided into three parts. In the first part, consisting of Chapter 2, the innovations representation (IR) is discussed so as to provide the reader with an introduction to this powerful tool.

In Chapter 3, a formal solution to the problem of detecting a stochastic signal process in AWGN is presented. This Innovations Based Detection Algorithm (IBDA) is then applied to a number of specific Gaussian detection problems as a demonstration of its usefulness in solving detection problems.

In the third part, consisting of Chapter 4 and Chapter 5, the IBDA is applied to the non-Gaussian problem of target (aircraft) detection by ASR systems. Two fully adaptive implementations are discussed and realised via software. Performance comparisons are made with three versions of the Moving Target Detection (MTD) algorithms that are currently in use.

In Chapter 2, a tutorial outline of the IR and its application to estimation theory and to detection theory is presented. The section on estimation theory includes a discussion of the whitening approach to estimation theory first used by Bode and Shannon [5], and also an introduction to the Kalman approach to estimation theory pioneered by Kalman [35] and by Kalman and Bucy [36]. As pointed out by Kailath [27], this whitening concept is now recognized to be an application of the IR. The Kalman approach to estimation theory is included since, as demonstrated by Kailath [27], the IR is a powerful tool in expanding the domain of application of Kalman filtering.

The innovations approach to detection theory is outlined in section 2.3. This is an important discussion since the theoretical contributions of this thesis reside upon this use of the IR which was pioneered by Kailath [32, 33]. In this paper [32] Kailath greatly extended the domain of application of detection theory. Specifically the solution to the continuous-time analogue of the detection problem (eqn. 1.1) is derived and shown to have an estimator-correlator structure. The simplicity and breadth of application of this result is in striking contrast with the methods, such as the Karhunen-Loeve expansions, used by classical detection theory (Helstrom [22], Van Trees [64], Whalen [66]) to solve their detection problems posessing non-Gaussian statistics.

To place this thesis in perspective with Kailath's results, it should be noted that much of the mathematical formalism introduced by Kailath [32] is dependent upon the continuous-time properties of the detection problem which is examined in his paper [32]. Hence, his analysis is not directly transferable to the discrete-time detection problems with which this research is concerned. However, there is no doubt that this thesis' use of the

IR as a tool for solving detection problems is due to Kailath's pioneering work.

The basic mathematical results are presented in Chapter 3. Specifically, in section 3.1, the logarithm of the likelihood ratio (LLR) is derived for the detection problem:

$$\text{Hypothesis } H_1: x(n) = z(n)+w(n), \quad n=0,\ldots, N-1 \qquad (1.2a)$$

$$\text{Hypothesis } H_0: x(n) = w(n), \quad n=0,\ldots, N-1 \qquad (1.2b)$$

where $\{x(n)\}$ is the received data, $\{z(n)\}$ is a stochastic process (which need not be Gaussian), and $\{w(n)\}$ is the usual WGN process (Metford et al [44]).

The LLR is derived under the condition that both the stochastic signal process $\{z(n)\}$ and the WGN process $\{w(n)\}$ are strict-sense stationary. Discussions of the definition of strict-sense stationary may be found in Doob [9] and Papoulis [50]. From further analysis of the form of the LLR when applied to some specific detection problems (i.e. the detection of a deterministic signal in AWGN) it is obvious that this requirement of strict-sense stationarity is only a sufficient condition.

This LLR may be written in the, by now, classic estimator - correlator form as first discussed by Kailath [32]. Essentially the LLR is formed by calculating a best estimate of the stochastic signal process. Since this estimate is calculated by assuming that the signal process has been received, (i.e. Hypothesis H_1 is correct) it may be more properly referred to as a pseudo-estimate, a term introduced by Kailath [27]. This estimate is then correlated with the received data, in much the same manner as the known signal is correlated with the received data in the correlation receiver (Van Trees [65]).

Thus, any implementation of this LLR as applied to a given detection problem will have an estimator, whose structure will be dictated by some statistical model of the stochastic signal process. If the statistics of the signal process are not known or are dynamic, then the estimation necessarily must be adaptive, resulting in an adaptive detection algorithm.

The LLR for the detection problem (eqn. 1.2) is referred to in the thesis as the Innovations Based Detection Algorithm (IBDA). The word "algorithm" is employed since the solution to this detection problem may be utilized so as to solve a wide variety of related detection problems.

A detection problem of particular interest in this thesis is the detection of a stochastic signal process in additive, non-Gaussian coloured noise. This detection problem is formulated as:

Hypothesis H_2: $x(n)=s(n)+c(n)+w(n)$, $n=0,\ldots,N-1$ (1.3a)

Hypothesis H_1; $x(n)= \quad c(n)+w(n)$, $n=0,\ldots,N-1$ (1.3b)

where $\{x(n)\}$ is the received data, $\{s(n)\}$ is the stochastic signal process, $\{c(n)\}$ is some non-Gaussian coloured stochastic process and $\{w(n)\}$ is the usual WGN process.

This is the general hypothesis testing formulation of the radar target detection problem in the presence of the often non-Gaussian clutter from the surrounding radar environment and the corruptive effects of the receiver front-end, (Van Trees [65]). As proven in section 3.2, under certain very loose restrictions, the detection problem (eqn. 1.3) may be expanded through the introduction of a dummy hypothesis into the two coupled detection problems.

Hypothesis H_2: $x(n)=s(n)+c(n)+w(n)$, $n=0,\ldots,N-1$ (1.4a)

Hypothesis H_o: $x(n)= \quad w(n)$, $n=0,\ldots,N-1$ (1.4b)

and

Hypothesis H_1: $x(n)=c(n)+w(n)$, $n=0,\ldots,N-1$ (1.5a)

Hypothesis H_o: $x(n)= \quad w(n)$, $n=0,\ldots,N-1$ (1.5b)

Each of these two detection problems (eqn. 1.4) and (eqn. 1.5) is solved through application of the IBDA. The LLR for the original detection problem (eqn. 1.3) is the difference between the LLRs for the two coupled detection problems (eqn. 1.4) and (eqn. 1.5).

This solution may be compared to the formalism developed by Schwartz et al [12, 54, 55]. Schwartz examined the discrete-time detection problem where the conditional probability density function (pdf) of the data process given the signal process has a rather general exponential form. Hence, the detection problem considered by Schwartz can be described as the detection of a stochastic signal in additive noise where the noise statistics are from a family containing the Gaussian, gamma, beta, binomial and Poisson distributions [12].

The focus of this research of Schwartz et al is at least partly to ascertain the formal connection between the discrete-time detection problem (eqn. 1.2) and its continuous-time analogue as studied by Kailath [32]. As a result, the LLR as formulated by Schwartz is not geared for ease of computer implementation.

In section 3.3, the IBDA is applied to a number of specific Gaussian detection problems. These include the classic problems of a) the detection of a known signal in AWGN (Helstrom [22], Van Trees [64], Whalen [66]), and b) the detection of a Gaussian process in AWGN (Scholtz et al [53], Porat and Friedlander [51]).

The other two specific Gaussian detection problems which are solved through an application of the IBDA include c) the detection of a deterministic signal in coloured Gaussian noise (deLong and Hofstetter [8], Kay [38]) and d) the detection of a Gaussian signal process in coloured Gaussian noise (Therrien [63]). The significance of the detection problem (c) is that it is the basis of the detection model used in the design of the MTD algorithms (Taylor [62]) that set the performance standards against which the IBDA is tested.

For all of these specific Gaussian detection problems to which the IBDA is applied, the resulting LLRs have the same structure as the LLRs derived using other methods. This is of particular interest with reference to the MTD design model since it is demonstrated that the IBDA, when applied to the MTD detection model, generates the basic MTD algorithm.

Two different fully adaptive implementations of the IBDA (Metford and Haykin [45]) as applied to the radar target detection problem are discussed in Chapter 4. A brief description of ASR systems followed by a discussion of the usually non-Gaussian statistical properties of the radar data is contained in section 4.1. In section 4.2 the LLR for this detection problem is derived using the IBDA as generalized to handle complex baseband signals. This LLR is then reformulated into a recursive mode for efficient adaptive implementation.

The actual details of the adaptive implementation of the LLR test are discussed in section 4.3. Within this section is included the consideration of the structure of the prediction sections of the processes, and the calculations of the a priori prediction error powers and the adaptive threshold. Considerable experimentation was required in order to establish a stable and robust implementation. One noteworthy feature is the use of the median statistic (Ataman et al [3]) in the calculation of the adaptive threshold.

The two implementations differ only in the algorithm used in the prediction calculation. The first implementation uses the least squares adaptive prediction error lattice filter due to Morf (Friedlander [13], Schichor [58])). The other uses the Kalman prediction error tapped delay-line filter (Zhang and Haykin [69]).

The MTD algorithms (Karp and Anderson [37], Muehe [47], Muehe et al [48], Taylor [61], [62]) are discussed in section 4.4. The MTD algorithms are all designed for high performance in both weather and ground clutter dominated radar environments.

Systems performance evaluations and comparisons are made in Chapter 5. Real data was recorded from a modern coherent ASR installation and used for performance evaluations. All performance evaluations are made in terms of the probability of false alarm (P_{FA}) as a function of both the target signal power to the clutter plus receiver noise power ratio (SCNR) and the target doppler, for a fixed probability of detection (P_D). These receiver operating characteristics are obtained for all of the detection algorithms for both ground clutter and weather clutter dominated radar environments.

These results are summarized in Fig. 5.17 and Fig. 5.18 in the form of relative performance curves for $P_{FA} = 10^{-4}$ and $P_D = 0.8$. The results prove that when properly implemented, the IBDA can provide superior performance, albeit at the cost of increased computational complexity and memory requirements.

CHAPTER 2

THE INNOVATIONS REPRESENTATION

The Innovations Representation (IR) of a stochastic process $\{y(t)\}$ is a representation of that process as the output of a causal and causally invertible filter driven by a white noise process $\{\nu(t)\}$. The requirement that the filter be causally invertible ensures that the white noise process (the innovations process $\{\nu(t)\}$) is probabilistically equivalent to the original process $\{y(t)\}$. Thus, the innovations process can be used instead of the original process for statistical problems of detection and estimation. The use of the IR in this fashion greatly simplifies the formulation and the structure of detection and estimation algorithms, as the innovations process is uncorrelated and, for a large number of applications, Gaussian.

This probabilistic equivalence between the white process $\{\nu(t)\}$ and the original process $\{y(t)\}$ is the reason for calling the white process $\{\nu(t)\}$ the innovations process. Innovation denotes "newness" and this quality is represented by the whiteness of the process $\{\nu(t)\}$ where any redundant information in the form of correlation of the process $\{y(t)\}$ has been removed. Hence only new information is retained in the innovations process $\{\nu(t)\}$. The term "innovations process" has been attributed to Wiener and Masani (Kailath [27]).

Details of the IR are discussed in section 2.1. Application of the IR to estimation theory and to detection theory is made in section 2.2 and section 2.3 respectively.

The contents of this tutorial chapter lean heavily upon the stimulating series of papers on the IR and its applications by Kailath et al [1,14,15,24,25, 26,27,28,29,30,31,32,33,34]

2.1 Mathematics of the Innovations Process

Given a stochastic process $\{y(t),\ t\varepsilon T\}$, the innovations process $\{\nu(t),\ t\varepsilon T\}$ is defined (Kailath [25] as:

$$\nu(t) \quad = \quad y(t) - \hat{y}(t), \quad t\varepsilon T \tag{2.1}$$

The unique optimum prediction, in a minimum mean square error (MMSE) sense, is defined as the conditional expectation (Doob [9]):

$$\hat{y}(t) \quad \overset{\Delta}{=} \quad E[y(t)\,|\,y(t'),t' < t]\ , \quad t\varepsilon T \tag{2.2}$$

From the definition of conditional expectation (Doob [9]), it follows that the prediction error $(y(t)-\hat{y}(t))$ is orthogonal to (uncorrelated with) every finite variance function of the set of data $\{y(t'), t'<t\}$ (Frost and Kailath [14]).

If the process $\{y(t)\}$ is Gaussian, then the optimal MMSE prediction $\hat{y}(t)$ is a linear function of the set of data $\{y(t'), t'\leq t\}$ (Doob [9]) Hence, the innovations process is also Gaussian.

The stochastic process $\{y(t)\}$ is now defined in a form which is of interest to both detection and estimation theory where:

$$y(t) = z(t)+w(t), \quad t\varepsilon T \tag{2.3}$$

In general, $\{w(t), t\varepsilon T\}$ is a white Gaussian process, and the stochastic process $\{z(t), t\varepsilon T\}$ is statistically independent of the white Gaussian process $\{w(t)\}$ (Kailath [32]). If t is continuous, the resulting innovations process is Gaussian with the same variance as that of the white Gaussian process $\{w(t)\}$ (Kailath [32]). This important (and remarkable) property is fundamental to many of the applications of the IR to continuous-time detection and estimation problem. However, it should be noted that for discrete-time processes, the innovations process is only asymptotically Gaussian with a different variance than that of the original Gaussian process $\{w(t)\}$ (Metford et al [44]).

2.2 The Innovations Approach to Estimation Theory

A statement of a generic estimation problem is as follows. Consider the observation process $\{y(t)\}$ of a signal process $\{z(t)\}$ in additive white noise $\{w(t)\}$ such that:

$$y(t) = z(t)+w(t), \quad T_i \leq t \leq T_f \tag{2.4}$$

Usually it is assumed that the signal process $\{z(t)\}$ is statistically independent of the noise process $\{w(t)\}$ (Kailath [27]). It will also be assumed, for simplicity, that all processes are stationary. An essential requirement is that the covariance $R_y(\tau) = E[y(t)y(t+\tau)]$ is positive definite.

The quantity $x(t)$ which is to be estimated is assumed to be related in a deterministic and linear manner to the signal process $\{z(t)\}$. Hence, the linear estimation problem is the determination of the random variable $x(t)$ defined in terms of a linear operation on the known observations $\{y(t)\}$, where:

$$\hat{x}(t) \overset{\Delta}{=} \int_{T_i}^{T_f} d\tau h(t-\tau)y(\tau), \quad T_i \leq t \leq T_f \tag{2.5}$$

such that the mean square of the estimation error $(x(t)-\hat{x}(t))$ is a minimum.[1]

---

[1] Two other common criteria for defining optimality (Van Trees [64]) are the absolute value of the error, which results in the median estimate, and a uniform cost function, which results in the maximum a posteriori estimate. Sherman [55] shows that for the Gaussian problem, all three criteria result in identical estimates.

From section 2.1, this optimal MMSE estimate implies that

$$E[(x(t)-\hat{x}(t))y(t')] = 0, \quad T_i \leq t,\ t' \leq T_f \tag{2.6}$$

By substituting eqn. 2.5 into eqn. 2.6 it is easily shown that the optimal estimate is the solution of the integral equation:

$$R_{xy}(t) = \int_{T_i}^{T_f} d\tau h(t-\tau) R_y(\tau), \quad T_i \leq t \leq T_f \tag{2.7}$$

where the cross covariance $R_{xy}(t) = E[x(\tau)y(t-\tau)]$. This equation is readily solved for the particular estimation problem where $T_i = -\infty$ and $T_f = +\infty$.

A much more interesting problem arises when the final observation time T_f corresponds to the time at which the estimate is desired. Hence $T_f = t$ and eqn. 2.7 becomes:

$$R_{xy}(t) = \int_0^{\infty} d\tau h(\tau) R_y(t-\tau), \quad 0 \leq t \leq \infty \tag{2.8}$$

This equation is commonly referred to as the Wiener-Hopf equation (Van Trees [64]).

If we now set $x(t) = z(t+\alpha)$ where α is a constant, then for $\alpha < 0$, $\alpha = 0$, $\alpha > 0$ the resulting estimation problems are referred to as smoothing, filtering, and prediction, respectively.

Wiener [67] solved eqn. 2.8 for the filtering problems through his method of spectral factorisation under the condition that the power spectrum $S_y(f)$ is a rational function. Bode and Shannon [5] used whitening filters to obtain a more transparent solution than Wiener's. This method was later realized to be the continuous-time version of the original Wold-Kolmogorov estimation technique (Kailath [27]). These whitening filter approaches are now referred to as the innovations approach to estimation theory (Kailath [27]).

To illustrate the conceptual simplicity of the IR approach to estimation theory, assume that a realisable and reversible whitening linear filter $\{h_1(t)\}$ has been found. As discussed in section 2.1, the estimated quantity can be written in terms of the uncorrelated output $\{\nu(t)\}$ (the innovations process) of the whitening filter. Hence the optimal estimation filter $\{h_2(t)\}$ in terms of the innovations process $\{\nu(t)\}$, is the solution of the now trivial integral equation, where, writing $R_{x\nu}(t)=E[x(\tau)\nu(t-\tau)]$, then, from eqn. 2.8:

$$R_{x\nu}(t) = \int_0^\infty d\tau h_2(\tau)R_\nu(t-\tau) \tag{2.9}$$

$$= \int_0^\infty d\tau h_2(\tau)\delta(t-\tau)$$

$$= h_2(t)$$

Thus the impulse response of the estimation filter $h_2(t)$ (in terms of the innovation process) is written:

$$h_2(t) = R_{x\nu}(t)$$

But $R_{x\nu}(t) = \int_{-\infty}^{0} d\tau h_1(\tau) R_{xy}(t-\tau)$ (2.10)

where $\{h_1(\tau)\}$ is the impulse response of the whitening filter. Thus the entire optimum estimation filter is just a cascade of the whitening filter $\{h_1(t)\}$ and the filter $\{h_2(t)\}$ specified by eqn. 2.10.

As pointed out by Kalman [35], the above methods for specifying the Wiener filter are subject to a few limitations. These include the often difficult tasks of synthesizing the estimation filter's impulse response from the observed data and the various generalizations of the solutions of the Wiener problem.

Kalman [35] sidestepped these problems by modelling the random signal process $\{y(t)\}$ as being generated by passing white noise through a (possibly time varying) lumped linear dynamic system. Complete knowledge of this system is assumed.

This approach emphasises the concept of states and state transitions. For example, in an aerospace application, the states of the system might be the position and velocity of a satellite. The state transitions can easily be realized through an application of kinematics. The observed data from which estimates of the satellite's position and velocity must be inferred from is the telemetry data.

The problem is now written (Kalman [35]):

$$\underline{x}(t+1) = \underline{\phi}(t+1;t)\underline{x}(t)+\underline{u}(t) \tag{2.11}$$

$$\underline{y}(t) = \underline{M}(t)\underline{x}(t) \tag{2.12}$$

where $\underline{u}(t)$ is a Gaussian random process of n-vectors with zero mean, $\underline{x}(t)$ is the state n-vector, $\underline{y}(t)$ is a state observation p-vector, $\underline{\phi}(t+1;t)$ is the state transition nxn matrix and $\underline{M}(t)$ is the measurement pxn matrix.

Given the observed values of $\underline{y}(t_o),\ldots,\underline{y}(t)$, the problem is to find an estimate $\hat{x}(t_1|t)$ of $\underline{x}(t_1)$ which is optimal in a MMSE sense.

Kalman [35] directly solved the prediction and filtering problems within this framework using the orthogonal properties of the innovations process. The dynamics of the state transitions (eqn. 2.11) result in a set of recursive relations for solving the estimation problem.

Kalman and Bucy [36] considered the continuous-time version of the original Kalman formulation. However, the innovations approach was not explicitly used. Instead, the orthogonal properties of the MMSE estimates were employed to formulate the appropriate Wiener-Hopf equation. The structure of the state equations were used to transform the solution of the Wiener-Hopf equation into the solution of a non-linear differential equation.

Kailath [27] refined the innovations approach and successfully applied it directly to the Kalman-Bucy problem to obtain not only the prediction and filtered estimates [27], but also the smoothed estimates [28]. The solutions are obtained in exactly the same manner as used for the discrete-time problem.

The innovations approach was used by Kailath [1,14, 15,24,25,26,27,28,29,30,31,32,33,34] to examine many generalisations of the estimation problem (eqn. 2.4), including coloured background noise and non-linear estimates. In all cases, the resulting equations give a great deal of physical significance and understanding to the pertinent problem.

2.3 The Innovations Approach to Detection Theory

Statistical decision theory is approached from the viewpoint of statistical hypothesis testing (Helstrom [22], Van Trees [64], Whalen [66]). Specifically, a choice must be made between two hypotheses through the use of some decision-making criterion. As a concrete example, the problem of the detection of a signal process in additive noise would be formulated as:

$$\text{Hypothesis } H_1: \quad x(t) = z(t)+w(t), \quad 0 \le t \le T \qquad (2.13a)$$

$$\text{Hypothesis } H_o: \quad x(t) = \quad w(t), \quad 0 \le t \le T \qquad (2.13b)$$

where $\{x(t)\}$ is the received data process, $\{z(t)\}$ is the signal process, and $\{w(t)\}$ is the additive noise.

This decision-making problem is solved using Bayes rule (Papoulis [50]). Under Hypothesis H_1, Bayes rule takes the form:

$$P[H_1|\{x(t)\}]P[\{x(t)\}] = P[H_1]P[\{x(t)\}|H_1] \qquad (2.14)$$

where $P[H_1]$ is the a priori probability of Hypothesis H_1, $P[H_1|\{x(t)\}]$ is the a posteriori probability of Hypothesis H_1 given that the data $\{x(t)\}$ has been received, and $P[\{x(t)\}|H_1]$ is the likelihood of receiving the data $\{x(t)\}$ given that Hypothesis H_1 is true.

Dividing eqn. 2.14 by a similar expression for Hypothesis H_o yields an equation of fundamental importance to detection theory, namely:

$$\frac{P[H_1|\{x(t)\}]}{P[H_o|\{x(t)\}]} = \frac{P[H_1]}{P[H_o]} \cdot \frac{P[\{x(t)\}|H_1]}{P[\{x(t)\}|H_o]} \qquad (2.15)$$

The second term on the RHS of eqn. 2.15 is called the likelihood ratio and will be denoted:

$$\Lambda_{H_1,H_o} \stackrel{\Delta}{=} \frac{P[\{x(t)\}|H_1]}{P[\{x(t)\}|H_o]} \qquad (2.16)$$

From Doob [9] it is obvious that the likelihood ratio may also be written in terms of conditional n-fold joint probability distributions where, defining the n-vector $\underline{x} = (x(t_1), \ldots, x(t_n))$, and given an arbitrary n-vector $\underline{\alpha} = (\alpha_1, \ldots \alpha_n)$, then:

$$\Lambda_{H_1,H_o} = \frac{P(\underline{x} \leq \underline{\alpha} | H_1)}{P(\underline{x} \leq \underline{\alpha} | H_o)} \tag{2.17}$$

Also, if the conditional n-fold joint probability density functions are well behaved (Feller [11]), then:

$$\Lambda_{H_1,H_o} = \frac{p(\underline{x} | H_1)}{p(\underline{x} | H_o)} \tag{2.18}$$

In theory, at least, the decision as to which of the two hypotheses is true is made on the basis of whether or not the ratio of the a posteriori probabilities is greater or less than some preset threshold. The threshold is set according to the desired decision-making criterion such as Bayes, Neyman-Pearson or Minimax (Van Trees [64]). However, the calculation of the ratio of the a posteriori probabilities requires knowledge of the a priori probabilities which may or may not be available. Interesting discussions of the problem of the a priori probabilities may be found in Lehmann [42] and Woodward [68].

Irrespective of the a priori probabilities, the major computational burden for decision-making resides in the calculation of the LR. As a result, detection problems are formulated in terms of a LR test (Van Trees [64]). In addition, because of the preponderance of Gaussian detection problems, usually the natural logarithm of the likelihood ratio (LLR) is implemented.

A classic problem is the detection of a known signal $\{s(t)\}$ in AWGN. The additive noise is stated as being an uncorrelated Gaussian process since this is a good model of the corruptive effects of the thermal noise generated at the front end of many receiver structures (Whalen [66]). For this detection problem the LLR is essentially a correlation of the known signal $\{s(t)\}$ with the received data process $\{x(t)\}$, as shown by:

$$\ln\Lambda_{H_1,H_o} = \frac{2}{N_o}\int_o^T dt\, s(t)x(t) - \frac{1}{N_o}\int_o^T dt\,(s(t))^2 \qquad (2.19)$$

where $N_o/2$ is the value of the two-sided power spectral density (Haykin [20]) of the WGN $\{w(t)\}$. This LLR is usually termed a matched filter or a correlation receiver (Van Trees [64]).

This detection problem has frequently been generalised to the detection of a known signal in coloured Gaussian noise (Whalen [66]) and to the detection of a Gaussian process in Gaussian noise (Helstrom [22]).

The innovations approach to detection theory

(Kailath [32], [33]) led to a significant breakthrough in problems involving the detection of non-Gaussian processes in AWGN.

To apply the innovations method to the detection problem (eqn. 2.13) it is assumed that the signal process $\{z(t)\}$ is a (not necessarily Gaussian) finite energy stochastic process (Kailath [24]). It is also assumed that the AWGN process $\{w(t)\}$ is statistically independent of the signal process $\{z(t)\}$. If the assumption is made that the received process $\{x(t) = z(t)+w(t)\}$ (i.e. that Hypothesis H_1 is true), then the optimal MMSE estimate

$$\hat{z}(t) = E[z(t)|x(t'),t'<t] \tag{2.20}$$

is formed. The resulting innovations process $\{\nu(t)\}$ where

$$\nu(t) = x(t)-\hat{z}(t) \tag{2.21}$$

is Gaussian with the same variance as the noise process $\{w(t)\}$, if t is continuous (Kailath [37]). As a result, the above detection problem (eqn. 2.13) may be replaced by the probabilistically identical detection problem:

$$\text{Hypothesis } H_1: \quad x(t) = z(t)+w(t)$$
$$= \hat{z}(t)+\nu(t) \tag{2.22a}$$

$$\text{Hypothesis } H_o: \quad x(t) = w(t)$$
$$= \nu(t) \tag{2.22b}$$

This detection problem (eqn. 2.22) is the detection of a conditionally-known process $\{\hat{z}(t)\}$ in AWGN. The resulting LLR (Kailath [32]):

$$\ln\Lambda_{H_1,H_o} = \int dt\hat{z}(t)x(t) - \frac{1}{2}\int dt(\hat{z}(t))^2 \qquad (2.23)$$

bears a startling resemblance to the LLR (eqn. 2.19) for the detection of a known signal in AWGN. The structure of this LLR has the form of an estimator-correlator. It should be noted that the first integral on the RHS of eqn. 2.23 is an Ito integral (Kailath [32]) which obeys different rules than those for ordinary integrals.

The importance of this estimator-correlator structure is that it gives a physical feeling for these non-Gaussian detection problems. This leads to the suggestion that even if the needed optimal MMSE prediction $\{z(t)\}$ is not available, or known, a good realisable approximation might well suffice.

CHAPTER 3

THE INNOVATIONS BASED DETECTION ALGORITHM

A major contribution of this thesis is the derivation and implementation of the LLR for the detection of a discrete-time stochastic process imbedded in AWGN. This detection problem is formulated as an hypothesis testing problem.

Specifically, given a received real data process $\{x(n),\ n=0,\ldots,\ N-1\}$, then the discrete-time detection problem:

$$\text{Hypothesis } H_1:\ x(n) = z(n)+w(n),\quad n=0,\ldots,\ N-1 \qquad (3.1a)$$

$$\text{Hypothesis } H_0:\ x(n) = \qquad w(n),\quad n=0,\ldots,\ N-1 \qquad (3.1b)$$

is defined, where $\{z(n),\ n=0,\ldots,N-1\}$ is assumed to be statistically independent of the AWGN process $\{w(n),\ n=0,\ldots,N-1\}$. In addition, both of these two stochastic processes are assumed to be strict-sense stationary (Doob [9]).

In section 3.1, a transformation of the discrete-time innovations process is defined and shown, under appropriate conditions, to satisfy a generalised central limit theorem. This transformed process is introduced into the discrete-time detection problem (eqn. 3.1) by means of a dummy hypothesis. The LLR is then formulated assuming that this transformed innovations process is Gaussian (Metford et al [44]). This LLR is referred to within the remainder of the thesis as the Innovations Based Detection Algarithm (IBDA).

In section 3.2 the IBDA is applied to the problem of detecting a stochastic signal in non-Gaussian coloured noise. The results of section 3.1 and section 3.2 are then applied, in section 3.3, to some specific detection problems. These include: a) the detection of a deterministic signal in AWGN, b) the detection of a coloured Gaussian process in AWGN, c) the detection of a deterministic signal in coloured Gaussian noise, and d) the detection of a coloured Gaussian process in coloured Gaussian noise. These specific applications are presented so as to demonstrate the usefulness of the IBDA as a tool for solving many detection problems.

A summary of this chapter is presented in section 3.4.

3.1 The Derivation of the Innovations Based Detection Algorithm

3.1.1 The Reformulation of the Discrete-Time Detection Problem

Defining the sigma-field of events $\mathcal{F}_n \overset{\Delta}{=} \sigma\{x(k), k \leq n\}$, the innovations process will be denoted henceforth as $\{\tilde{x}(n|H_1),\ n=0,\ldots,N-1\}$ where:

$$\tilde{x}(n|H_1) = x(n)-E[x(n)|\mathcal{F}_{n-1},H_1]$$

$$= (z(n)-\hat{x}(n|H_1))+w(n) \qquad (3.2)$$

where the prediction $\hat{x}(n|H_1) = E[x(n)|\mathcal{F}_{n-1},H_1]$ is optimal in a MMSE sense, given the past data $\{x(k),\ k \le n\}$ and assuming that Hypothesis H_1 is correct.

Defining the process $\{y(n),\ n=0,\ldots,N-1\}$ by

$$y(n) = \frac{1}{A(n)} \sum_{k=0}^{n} x(k), \quad n=0,\ldots,N-1 \qquad (3.3)$$

where $\{A(n),\ n=0,\ldots,N-1\}$ is a set of arbitrary finite real constants, then the above detection problem (eqn. 3.1) may be transformed into the equivalent detection problem:

$$\text{Hypothesis } H_1':\ y(n)=\frac{1}{A(n)} \sum_{k=0}^{n} (z(k)+w(k)),\quad n=0,\ldots,N-1 \qquad (3.4a)$$

$$\text{Hypothesis } H_o':\ y(n)=\frac{1}{A(n)} \sum_{k=0}^{n} w(n),\quad n=0,\ldots N-1 \qquad (3.4b)$$

Consider now the process $\{\rho(n),\ n=0,\ldots,N-1\}$, where

$$\rho(n) = y(n)-\frac{1}{\sigma_s(n)} \sum_{k=0}^{n} \hat{x}(k|H_1),\quad n=0,\ldots,N-1 \qquad (3.5)$$

and $\sigma_s^2(n)$ is the variance of the sum of the first n terms of the innovations process $\{\tilde{x}(k|H_1),\ k=0,\ldots,N-1\}$ given that Hypothesis H_1 is correct. If in eqn. 3.3 the set

$\{A(n)\}$ is defined such that:

$$A(n) = \sigma_s(n), \quad n=0,\ldots,N-1$$

then

$$\rho(n) = \frac{1}{\sigma_s(n)} \sum_{k=0}^{n} x(k) - \frac{1}{\sigma_s(n)} \sum_{k=0}^{n} \hat{x}(k|H_1)$$

$$= \frac{1}{\sigma_s(n)} \sum_{k=0}^{n} x(k) - \hat{x}(k|H_1)$$

$$= \frac{1}{\sigma_s(n)} \sum_{k=0}^{n} \tilde{x}(k|H_1) \tag{3.6}$$

where $\sigma_s^2(n) = E[(\sum_{k=0}^{n} \tilde{x}\ (k|H_1))^2|H_1]$ (3.7)

$$= \sum_{k=0}^{n} E[(z(k)-\hat{x}(k|H_1)^2]+n\sigma_w{}^2$$

and $\sigma_w{}^2 \overset{\Delta}{=} E[(w(k))^2]$, i.e. $\sigma_w{}^2$ is the variance of the AWGN process. Hence, the process $\{\rho(n),\ n=0,\ldots,N-1\}$ is composed of variance-normalised partial sums of the discrete-time innovations process $\{\tilde{x}(k|H_1),\ n=0,\ldots,N-1\}$.

The detection problem of eqn. 3.4 may now be written as:

Hypothesis H_1': $y(n)=(\frac{1}{\sigma_s(n)} \sum_{k=0}^{n} \hat{x}(k|H_1))+\rho(n)$, $n=0,\ldots,N-1$ (3.8a)

Hypothesis H_o': $y(n)=(\frac{1}{\sigma_s(n)} \sum_{k=0}^{n} w(n))$, $n=0,\ldots N-1$ (3.8b)

Denoting by $\overline{H}_o$ the dummy hypothesis $y(n) = \rho(n)$, $n=0, \ldots, N-1$, and then using the chain rule of LR's (Halmos [18]), the LR for the original detection problem (eqn. 3.1) may be written as:

$$\Lambda_{H_1,H_o} = \Lambda_{H_1',H_o'} = \Lambda_{H_1',\overline{H}_o} \Lambda_{\overline{H}_o,H_o'} \qquad (3.9)$$

where $\Lambda_{H_1',\overline{H}_o}$ is the LR for the detection problem:

$$\text{Hypothesis } H_1': y(n)=\left(\frac{1}{\sigma_s(n)} \sum_{k=0}^{n} \hat{x}(k|H_1)\right)+\rho(n), \quad n=0,\ldots,N-1 \qquad (3.10a)$$

$$\text{Hypothesis } \overline{H}_o : y(n)=\rho(n), \qquad n=0,\ldots,N-1 \qquad (3.10b)$$

and $\Lambda_{\overline{H}_o,H_o'}$ is the LR for the detection problem:

$$\text{Hypothesis } \overline{H}_o: y(n)=\rho(n), \qquad n=0,\ldots,N-1 \qquad (3.11a)$$

$$\text{Hypothesis } H_o':y(n)=\left(\frac{1}{\sigma_s(n)} \sum_{k=0}^{n} w(k)\right), \qquad n=0,\ldots,N-1 \qquad (3.11b)$$

There are two reasons for this reformulation of the original detection problem (eqn. 3.1). First, under Hypothesis H_1' (eqn. 3.10a), $y(n)$ is now expressed as the sum of a known function of the (known) past data and the process $\{\rho(n)\}$. Second, under Hypothesis $\overline{H}_o$ (eqn. 3.10b), the process $y(n)$ is expressed in terms of the process $\{\rho(n)\}$ composed of variance-normalised partial sums of the innovations process $\{\tilde{x}(n|H_1)\}$. Since the innovations process $\{\tilde{x}(n|H_1)\}$ is by definition uncorrelated and already posesses a Gaussian component (eqn. 3.2), it might be expected that the process $\{\rho(n)\}$ satisfies some form of a central limit theorem.

3.1.2 A Proof of the Asymptotic Normality of the Variance-Normalised Partial Sums of the Discrete-Time Innovations Process

In this section a proof is presented of the asymptotic normality of the process $\{\rho(n)\}$ composed of variance-normalised partial sums of the discrete-time innovations process $\{\nu(n)\}$. In addition, it is proven that the rate of convergence is $1/\sqrt{n}$ for large n.

Theorem 1

Denoting the zero-mean, unit-variance Gaussian probability distribution:

$$\phi(x) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{x} du \exp(-\tfrac{1}{2} u^2)$$

then $\lim_{n\to\infty} P[\rho(n) \le x] = \phi(x)$

Proof of Theorem 1

Although the random variables represented by the discrete-time innovations process $\{\tilde{x}(n|H_1), n=0,\ldots\}$ are uncorrelated, they are not identically distributed, nor are they statistically independent. This dependence conflicts with one of the necessary conditions for the classical central limit theorem (Loeve [43]). Since this dependency is weak, this proof relies upon a generalisation of the classical central limit theorem to the case of weakly dependent random processes.

Consider the n+1-fold joint probability distribution $P[\tilde{x}(o|H_1)\le a_o,\ldots,\tilde{x}(n|H_1)\le a_n]$. We now introduce a related probability distribution:

$$Q[\tilde{x}(o|H_1)\le a_o,\ldots\tilde{x}(n|H_1)\le a_n] \triangleq \prod_{k=o}^{n} P[\tilde{x}(k|H_1)\le a_k]$$

From Abbot and Blum [2], if for every $\varepsilon>0$, a positive integer n_o exists depending only on ε, and, if for every choice of non-negative integers $(i_o,\ldots,i_n)$ with $n_o\le i_o \ldots < i_n$, an n+1-fold joint probability distribution R exists which may depend on ε, n_o and n, then if

(1) $$\left|P[a_o<\tilde{x}(i_o|H_1)\le b_o,\ldots,a_n<\tilde{x}(i_n|H_1)\le b_n]-\prod_{k=o}^{n}P[a_k<\tilde{x}(i_k|H_1)\le b_k]\right|$$

$$<\varepsilon R[a_o\le x(o)\le b_o,\ldots,a_n\le x(n)\le b_n]$$

and if

(2) $\lim\limits_{n\to\infty} Q[\rho(n)\le x] = \phi(x)$

then $\lim\limits_{n\to\infty} P[\rho(n)\le x] = \phi(x)$

Condition (1) is satisfied if the random variables $\{\tilde{x}(n|H_1)\}$ are asymptotically statistically independent. Although this condition may not rigorously hold in all cases, in many physical systems we are justified in making this assumption. Condition (2) is easily satisfied since under Q, the random variables $\{\tilde{x}(n|H_1)\}$ are statistically independent. Using the classical central limit theorem (Loeve [43]), since from Eqn. 3.7 $\lim_{\to\infty} \sigma_s^2(n) \to \infty$, Then $\lim_{n\to\infty} Q[\rho(n) \leq x] = \phi(x)$. QED.

We now consider the rate of convergence of the probability distribution of $\rho(n)$ to the normal distribution $\phi(x)$.

Since the input data process $\{x(n)\}$ is assumed to be stationary then the innovations process $\{\tilde{x}(n|H_1)\}$ is asymptotically stationary. Choose some positive integer m to be sufficiently large, such that the process $\{\tilde{x}(n|H_1),\ n>m\}$ is stationary. Then consider the process $\{\rho'(n),\ n>m\}$, where:

$$\rho'(n) = \frac{1}{\sigma_s'(n-m)} \sum_{k=m}^{n} \tilde{x}(k|H_1)$$

and $\sigma_s'^2(n-m) = (n-m)\cdot(\sigma_e^2+\sigma_w^2)$, where $\sigma_e^2 = E[(z(k)-\hat{x}(k|H_1))^2]$, $k>m$

Lemma 1

$$\sup_x \; P[\rho'(n) \le x] - \phi(x) \le \frac{C_4}{\sqrt{n-m}}$$

where C_4 is a real and finite constant.

Proof of Lemma 1

From Statulevicius [60], if for any real and finite constants C_1, C_2, C_3, C_4 where

1) $\left| \sum_{k=m}^{n} \tilde{x}(k|H_1) \right| \le C_1$

2) $\sigma_s'^2(n-m) \ge \underline{C_2}(n-m)^{\lambda}$ for some $\lambda \ge \frac{1}{2}$
$n-m$

3) $\lim_{n\to\infty} G_3 \sigma_s'^2(n-m) \to \infty$

4) $\{\tilde{x}(n|H_1), n \ge m\}$ is asymptotically independent

then $\sup_x \left| P[\rho'(n) \le x] - \phi(x) \right| \le \dfrac{C_4}{\sqrt{\sigma_s'(n-m)}}$

Although condition (1) is in conflict with our assumption that the corruptive noise process $\{w(n)\}$ is Gaussian, it is obviously satisfied in an a posteriori sense in any physical system of interest. Conditions (2) and (3) are trivially upheld (Eqn. 3.7), while, condition (4) has already been assumed (see proof of Theorem 1). QED

Lemma 2

For every $\varepsilon>0$, $\lim_{n\to\infty} P[|\rho(n)-\rho'(n)| > \varepsilon] = 0$

Proof of Lemma 2

Since (Papoulis [50])

$$P[|\rho(n)-\rho'(n) > \varepsilon] \le \frac{E[|\rho(n)-\rho'(n)|^2]}{\varepsilon^2}$$

then

$$E[|\rho(n)-\rho'(n)|^2]=E[|\frac{1}{\sigma_s(n)}\sum_{k=0}^{n}\tilde{x}(k|H_1)-\frac{1}{\sigma_s'(n-m)}\sum_{k=m}^{n}\tilde{x}(k|H_1)|^2]$$

$$= E[|\frac{1}{\sigma_s(n)}\sum_{k=0}^{m-1}\tilde{x}(k|H_1)+\frac{\sigma_s'(n-m)-\sigma_s(n)}{\sigma_s(n)\ \sigma_s'(n-m)}\sum_{k=m}^{n}\tilde{x}(k|H_1)|^2]$$

$$= \frac{\sigma_s^2(m-1)}{\sigma_s^2(n)} + \frac{(\sigma_s'(n-m)-\sigma_s(n))^2}{\sigma_s^2(n)}$$

But $\lim_{n\to\infty}\frac{\sigma_s^2(m-1)}{\sigma_s^2(n)} = 0$ and $\lim_{n\to\infty}(\sigma_s^2(n-m)-\sigma_s^2(n)) = 0$

Hence $\lim_{n\to\infty} E[|\rho(n)-\rho'(n)|^2] = 0$, QED

Theorem 2

$$\lim_{n\to\infty}\ \sup_x |P[\rho(n)\le x]-\phi(x)| \le \frac{C}{\sqrt{n}}$$

Proof of Theorem 2

From Lemma 1, the probability distribution function of $\rho'(n)$ converges at a rate $1/\sqrt{n}$ to the Gaussian distribution. From Lemma 2, $\rho(n)$ converges in probability to $\rho'(n)$. Therefore, the probability distribution function of $\rho(n)$ for large n converges at a rate of $1/\sqrt{n}$ to the Gaussian distribution function. QED.

In summary it should be pointed out that the results of this section apply rigorously only to the probability distribution function of the stochastic process $\{\rho(n)\}$. However, following Feller [11], under the very mild constraint that the characteristic function of $\rho(n)$ be integrable, then it can be shown that these results apply also to the probability density function of $\rho(n)$. This condition is satisfied in almost all physical systems of interest.

Irrespective of the above comment though, since $\rho(n)$ is expected to be Gaussian for all practical purposes for a small number of samples, the detection problem (Eqn. 3.10) may be considered to be the detection of a conditionally-known signal in Gaussian noise, while the detection problem (Eqn. 3.11) may be considered to be the detection of one of two possible Gaussian processes.

3.1.3 The Calculation of the Likelihood Ratio

The probability density function (pdf) for the random vector $\underline{x}$ consisting of the n random variables $(x(o),\ldots, x(N-1))$ may be written as:

$$p(\underline{x}) = \frac{1}{(2\pi)^{N/2}|\underline{R}_x(N-1)|^{1/2}} \exp[-\frac{1}{2}(x-\underline{\xi})R_x^{-1}(N-1)(\underline{x}-\underline{\xi})^T] \qquad (3.12)$$

where $\xi(i)$ is the mean of the i-th random variable $x(i)$ and $\underline{R}_x(N-1)$ is the NxN correlation matrix where its i,j-th element:

$$R_x(N-1;i,j) = E[x(i)x(j)], \quad i,j=0,\ldots,N-1$$

Similarly, the conditional pdf of $x(n)$ given $(x(o),\ldots, x(n-1))$ is written:

$$p(x(n)|x(n-1),\ldots,x(1)) = \frac{1}{\sqrt{2}}\frac{|\underline{R}_x(n-1)|^{1/2}}{|\underline{R}_x(n)|^{1/2}} \exp[-\frac{1}{2}(\underline{x}-\underline{\xi})(\underline{R}_x^{-1}(n)$$

$$-\underline{R}_x^{-1}(n-1))(\underline{x}-\underline{\xi})^T] \qquad (3.13)$$

where it is to be understood that the matrix $\underline{R}_x^{-1}(n-1)$ has been augmented by an extra column and an extra row of zero to make it dimensionally compatible with the matrix $\underline{R}_x^{-1}(n)$

Consider the process $\{y(k) = \xi(k) + \rho(k),\ k=0,\ldots, N-1\}$ where $\{\rho(k),\ k=0,\ldots,N-1\}$ are correlated Gaussian random variables with correlation matrix $\underline{R}_\rho(N-1)$, and $\{\xi(k),\ k=0,\ldots,N-1\}$ is a conditionally-known process, that is, $\xi(k) = f_k(y(o),\ldots,y(k-1))$ where f_k is some function of $(k-1)$ variables. Through the repeated application of Bayes rule the following identity is generated:

$$p(\underline{y})=p(y(N)\,|\,y(N-1),\ldots,y(0)),\ldots,p(y(1)\,|\,y(0))p(y(0)) \qquad (3.14)$$

which, through the use of the explicit form of the Gaussian conditional pdf (Eqn. 3.13), is rewritten as:

$$p(\underline{y})= \frac{1}{\sqrt{2\pi}}\,\frac{|\underline{R}_\rho(N-2)|^{1/2}}{|\underline{R}_\rho(N-1)|^{1/2}}\exp[-\frac{1}{2}(\underline{y}-\underline{\xi})(\underline{R}_\rho^{-1}(N-1)-\underline{R}_\rho^{-1}(N-2))(\underline{y}-\underline{\xi})^T]\ x$$

$$x \ldots \frac{1}{\sqrt{2\pi}}\,\frac{|\underline{R}_\rho(1)|^{1/2}}{|\underline{R}_\rho(2)|^{1/2}}\exp[-\frac{1}{2}(\underline{y}-\underline{\xi})(\underline{R}_\rho^{-1}(2)-\underline{R}_\rho^{-1}(1))(\underline{y}-\underline{\xi})^T]\ x$$

$$\frac{1}{\sqrt{2\pi}}\,\frac{1}{|\underline{R}_\rho(1)|^{1/2}}\exp[-\frac{1}{2}(\underline{y}-\underline{\xi})\underline{R}_\rho^{-1}(1)(\underline{y}-\underline{\xi})^T]$$

$$= \frac{1}{(2\pi)^{N/2}|\underline{R}_\rho(N-1)|^{1/2}}\exp[-\frac{1}{2}(\underline{y}-\underline{\xi})\underline{R}_\rho^{-1}(N-1)(\underline{y}-\underline{\xi})^T] \qquad (3.15)$$

Now consider the detection problem (Eqn. 3.10) and define:

$$\xi(n) = \frac{1}{\sigma_s(n)} \sum_{k=0}^{n} \hat{x}(k|H_1), \quad n=0,\ldots,N-1 \qquad (3.16)$$

Assuming that the random variables $\{\rho(n),\ n=0,\ldots,N-1\}$ are Gaussian, then

$$\Lambda_{H_1',\overline{H}_o} = \frac{p(\underline{y}|H_1')}{p(\underline{y}|\overline{H}_o)}$$

$$= \frac{\dfrac{1}{(2\pi)^{N/2}|\underline{R}_\rho(N-1)|^{1/2}} \exp[-\frac{1}{2}(\underline{y}-\underline{\xi})\underline{R}_\rho^{-1}(N-1)(\underline{y}-\underline{\xi})^T]}{\dfrac{1}{(2\pi)^{N/2}|\underline{R}_\rho(N-1)|^{1/2}} \exp[-\frac{1}{2}\underline{y}\ \underline{R}_\rho^{-1}(N-1)\underline{y}^T]}$$

$$= \exp[-\frac{1}{2}(\underline{y}-\underline{\xi})\underline{R}_\rho^{-1}(N-1)(\underline{y}-\underline{\xi})^T + \frac{1}{2}\ \underline{y}\ \underline{R}_\rho^{-1}(N-1)\underline{y}^T] \qquad (3.17)$$

where $R_\rho(N-1;k,k') = E[\rho(k)\rho(k')]$

$$= \begin{cases} \dfrac{\sigma_s^2(k)}{\sigma_s(k)\sigma_s(k')}, & k < k' \\ 1, & k = k' \\ \dfrac{\sigma_s^2(k')}{\sigma_s(k)\sigma_s(k')}, & k > k' \end{cases}$$

Now, denoting the correlation matrix of the Gaussian process $\{\frac{1}{\sigma_s(n)} \sum_{\ell=0}^{n} w(\ell),\ n=0,\ldots,N-1\}$ as $\underline{C}(N-1)$ where the k, k'-th element

$$C(N-1;k,k') = E[(\frac{1}{\sigma_s(k)} \sum_{\ell=0}^{k} w(\ell))(\frac{1}{\sigma_s(k')} \sum_{\ell'=0}^{k'} w(\ell'))]$$

$$\frac{\min(k+1,\ k'+1)\sigma_w^2}{\sigma_s(k)\ \sigma_s(k')}\ ,\quad k,k'=0,\ldots,N-1$$

and again assuming that the random variables $\{\rho(n),\ n=0,\ldots,N-1\}$ are Gaussian, the LR for the second detection problem (Eqn. 3.11) is easily written as:

$$\Lambda_{\overline{H}_o,H_o'} = \frac{p(\underline{y}|\overline{H}_o)}{p(\underline{y}|H_o'')}$$

$$= \frac{\dfrac{1}{(2\pi)^{N/2}|\underline{R}_p(n-1)|^{1/2}} \exp[-\frac{1}{2}\ \underline{y}\ \underline{R}_p^{-1}(N-1)\underline{y}^T]}{\dfrac{1}{(2\pi)^{N/2}|\underline{C}(N-1)|^{1/2}} \exp[-\frac{1}{2}\ \underline{y}\ \underline{C}^{-1}(N-1)\underline{y}^T]}$$

$$= \frac{|\underline{C}(N-1)|^{1/2}}{|\underline{R}_p(N-1)|^{1/2}} \exp[\frac{1}{2}\ \underline{y}\ \underline{C}^{-1}(N-1)\underline{y}^T - \frac{1}{2}\ \underline{y}\ \underline{R}_p^{-1}(N-1)\underline{y}^T] \qquad (3.18)$$

Thus, substituting Eqn. 3.17 and Eqn. 3.18 into Eqn. 3.9, Eqn. 3.9 is rewritten as:

$$\Lambda_{H_1,H_o} = \Lambda_{H_1',\bar{H}_o}\Lambda_{\bar{H}_o,H_o'}$$

$$= \frac{|\underline{C}(N-1)|^{1/2}}{|\underline{R}_\rho(N-1)|^{1/2}} \exp[-\frac{1}{2}(\underline{y}-\underline{\xi})\underline{R}_\rho^{-1}(N-1)(\underline{y}-\underline{\xi})^T + \frac{1}{2}\,\underline{y}\,\underline{C}^{-1}(N-1)\underline{y}^T] \tag{3.19}$$

However, it is possible to show (Appendix 1) that:

$$\underline{y}\,\underline{C}^{-1}(N-1)\underline{y}^T = \frac{1}{\sigma_w^2}\sum_{k=0}^{N-1} x^2(k), \tag{3.20}$$

with

$$|\underline{C}(N-1)| = \prod_{k=0}^{N-1} \frac{\sigma_w^2}{\sigma_s^2(k)} \tag{3.21}$$

and that

$$(\underline{y}-\underline{\xi})\underline{R}^{-1}(N-1)(\underline{y}-\underline{\xi})^T = \sum_{k=0}^{N-1} \frac{\tilde{x}^2(k|H_1)}{\sigma^2(k|H_1)} \tag{3.22}$$

with

$$|\underline{R}_\rho(N-1)| = \prod_{k=0}^{N-1} \frac{\sigma^2(k|H_1)}{\sigma_s^2(k)} \tag{3.23}$$

where the expectation of the k-th prediction error power (assuming that Hypothesis H_1 is correct) $\sigma^2(k|H_1)$ is introduced such that:

$$\sigma^2(k|H_1) \triangleq \sigma_s^2(k)-\sigma_s^2(k-1) = E[(x(k)-\hat{x}(k|H_1))^2|H_1] \tag{3.24}$$

Thus, upon substituting these relations (Eqn. 3.20, 3.21, 3.22, 3.23) into Eqn. 3.19, the LR for the original detection problem (Eqn. 3.1) is written as:

$$\Lambda_{H_1,H_o} = \left(\prod_{k=0}^{N-1} \frac{\sigma_w^2}{\sigma^2(k|H_1)}\right)^{1/2} \exp\left[\frac{1}{2}\sum_{k=0}^{N-1}{}'\left(\frac{x^2(k)}{\sigma_w^2} - \frac{\tilde{x}^2(k|H_1)}{\sigma^2(k|H_1)}\right)\right] \tag{3.25}$$

Hence, taking the logarithm of both sides, the LLR:

$$\ln \Lambda_{H_1,H_o} = \frac{1}{2}\sum_{k=0}^{N-1}\left[\ln\left(\frac{\sigma_w^2}{\sigma^2(k|H_1)}\right)+ \frac{x^2(k)}{\sigma_w^2} - \frac{\tilde{x}^2(k|H_1)}{\sigma^2(k|H_1)}\right] \tag{3.26}$$

This LLR may be rewritten in terms of an estimator-correlator structure.

The LLR of Eqn. 3.26 will be referred to as the Innovations Based Detection-Algorithm (IBDA). The needed parameters for the calculation of the LLR include the set of the expectations of the prediction error powers

$\{\sigma^2(k|H_1), k=0,\ldots,N-1\}$, the variance of the AWGN σ_w^2 and the prediction error powers. So as to emphasize the fact that the set of expectations $\{\sigma^2(k|H_1), k=1,\ldots,N\}$ are used as normalising constants in the calculation of the LLR, the set $\{\sigma^2(k|H_1), k=0,\ldots,N-1\}$ will be referred to as the a priori prediction error powers.

3.2 The Application of the Innovations Based Detection Agorithm to the Detection of a Stochastic Signal Process in Additive Non-Gaussian Noise

As discussed in Chapter 1, this detection problem is formulated as:

$$\text{Hypothesis } H_2\text{: } x(n) = s(n)+c(n)+w(n), \quad n=0,\ldots,N-1 \tag{3.27a}$$

$$\text{Hypothesis } H_1\text{: } x(n) = c(n)+w(n), \quad n=0,\ldots,N-1 \tag{3.27b}$$

where $\{x(n)\}$ is the received data, $\{s(n)\}$ is the signal process, $\{c(n)\}$ is some other stochastic process, and $\{w(n)\}$ is the usual WGN process. The inclusion of the WGN component is not particularly restrictive since in many detection problems, the corruptive effects of the sensor may be adequately modelled as additive WGN (Van Trees [65]).

In order for the application of the IBDA to this detection problem (Eqn. 3.27) to be valid, it will be assumed that all of the stochastic processes are stationary. Later, through an examination of the form of the LLRs resulting from the application of the IBDA to some specific detection problems, it will become apparent that this is a sufficient but not a necessary condition.

An important feature of this detection problem is that the signal process $\{s(n)\}$ and the noise process $\{c(n)\}$ may not be statistically independent. This situation can arise in both sonar and radar detection problems where the target response $\{s(n)\}$ and the clutter process $\{c(n)\}$ which results from reflections from the surrounding environment are both generated by the same transmitted pulse of energy. However, the assumption will be made that the AWGN process $\{w(n)\}$ is statistically independent of both the signal process $\{s(n)\}$ and the noise process $\{c(n)\}$.

Since the received data $\{x(n)\}$ must be a finite energy process in any physical system of interest, then it is obvious that the background noise process $\{c(n)\}$ and the process $\{s(n) + c(n)\}$ must also be finite energy. Since these processes are finite energy, then, despite the above mentioned statistical dependency, from Kadota and Shepp [23], the chain rule of LR's may be applied to the above detection problem (Eqn. 3.27) through the use of a dummy hypothesis:

Hypothesis H_o': $x(n) = w(n)$, $n=0,\ldots,N-1$

where $\{w(n)\}$ is the WGN process. Hence the two coupled detection problems are generated where:

$$\text{Hypothesis } H_2\text{: } x(n) = s(n)+c(n)+w(n), \quad n=0,\ldots,N-1 \tag{3.28a}$$

$$\text{Hypothesis } H_o\text{: } x(n) = w(n), \quad n=0,\ldots,N-1 \tag{3.28b}$$

and

$$\text{Hypothesis } H_1\text{: } x(n) = c(n)+w(n), \quad n=0,\ldots,N-1 \tag{3.29a}$$

$$\text{Hypothesis } H_o\text{: } x(n) = w(n), \quad n=0,\ldots,N-1 \tag{3.29b}$$

The LR for the original detection problem (Eqn. 3.27) is written as:

$$\Lambda_{H_2,H_1} = \frac{\Lambda_{H_2,H_o}}{\Lambda_{H_1,H_o}} \tag{3.30}$$

where Λ_{H_2,H_o} and Λ_{H_1,H_o} are the LRs for the detection problem (Eqn. 3.28) and (Eqn. 3.29) respectively. But the two detection problems (Eqn. 3.28) and (Eqn. 3.29) may each be solved through application of the IBDA. Hence from (Eqn. 3.25)

$$\Lambda_{H_2,H_1} = \frac{\left(\prod_{k=0}^{N-1} \frac{\sigma_w^2}{\sigma^2(k|H_2)}\right)^{1/2} \exp \frac{1}{2}\left[\sum_{k=0}^{N-1} \frac{x^2(k)}{\sigma_w^2} - \frac{\tilde{x}^2(k|H_2)}{\sigma^2(k|H_2)}\right]}{\left(\prod_{k=0}^{N-1} \frac{\sigma_w^2}{\sigma^2(k|H_1)}\right)^{1/2} \exp \frac{1}{2}\left[\sum_{k=0}^{N-1} \frac{x^2(k)}{\sigma_w^2} - \frac{\tilde{x}^2(k|H_1)}{\sigma^2(k|H_1)}\right]} \tag{3.31}$$

and

$$2\ \ln\Lambda_{H_2,H_1} = \sum_{k=0}^{N-1} [\ln \frac{\sigma^2(k|H_1)}{\sigma^2(k|H_2)} + \frac{\tilde{x}^2(k|H_1)}{\sigma^2(k|H_1)} - \frac{\tilde{x}^2(k|H_2)}{\sigma^2(k|H_2)}] \quad (3.32)$$

where $\{\tilde{x}(k|H_i),\ k=0,\ldots,N-1$ and $\{\sigma^2(k|H_i),\ k=0,\ldots,N-1$ are the prediction errors and the a priori prediction error powers, respectively, assuming Hypothesis H_i, $i=1,2$.

It is apparent that if, say Hypothesis H_1 is correct then the prediction $\tilde{x}(k|H_1)$ will be optimal in a MMSE sense while the prediction $\tilde{x}(k|H_2)$ will not be optimal. Hence the normalised prediction error $\tilde{x}^2(k|H_1)/\sigma^2(k|H_1)$ will be smaller than the normalised prediction error $\tilde{x}^2(k|H_2)/\sigma^2(k|H_2)$. This situation will be reversed if Hypothesis H_2 is correct. This difference in the relative sizes of the normalised prediction error powers is the basic mechanism by which the LLR operates.

3.3 Some Specific Applications of the Innovations Based Detection Algorithm

In this section, several different specific detection problems are solved through application of the IBDA. Comparison is made with the solutions to these problems as found in the literature. In section 3.3.1 and section 3.3.2 the IBDA is applied to the detection of a deterministic signal in AWGN and the detection of a coloured

Gaussian process in AWGN, respectively. Solutions are obtained through application of (Eqn. 3.26). In section 3.3.3 and section 3.3.4 the IBDA is applied to the detection of a deterministic signal in additive coloured Gaussian noise and the detection of a coloured Gaussian process in coloured Gaussian noise, respectively. Solutions are obtained through application of (Eqn. 3.32).

3.3.1 The Detection of a Deterministic Signal in Additive White Gaussian Noise

Consider the detection problem:

$$\text{Hypothesis } H_1 : x(n) = s(n)+w(n), \quad n=0,\ldots,N-1$$

$$\text{Hypothesis } H_o : x(n) = w(n), \quad n=0,\ldots,N-1$$

where $s(n)$ is a deterministic signal process and $w(n)$ is a WGN process of variance σ_w^2.

Since the signal process $\{s(n)\}$ is deterministic whereas the WGN process $\{w(n)\}$ is completely unpredictable, then the optimum prediction $\hat{x}(k|H_1) = s(k)$, $k=0,\ldots,N-1$. Hence, the a priori prediction error power (Eqn. 3.24):

$$\begin{aligned}\sigma^2(k|H_1) &= E[(x(k)-\hat{x}(k|H_1))^2|H_1] \quad, \\ &= E[(s(k)+w(k)-s(k))^2] \quad, \\ &= \sigma_w^2\end{aligned}$$

Thus, from (Eqn. 3.26):

$$\ln\Lambda_{H_1,H_o} = \frac{1}{2}\sum_{k=0}^{N-1} \ln\left(\frac{\sigma_w^{2}}{\sigma^2(k|H_1)}\right) + \frac{\tilde{x}^2(k)}{\sigma_w^{2}} - \frac{\tilde{x}^2(k|H_1)}{\sigma^2(k\,H_1)}$$

$$= \frac{1}{2}\sum_{k=0}^{N-1} \ln\left(\frac{\sigma_w^{2}}{\sigma_w^{2}}\right) + \frac{x^2(k)}{\sigma_w^{2}} - \frac{(x(k)-s(k))^2}{\sigma_w^{2}}$$

$$= \frac{1}{2}\sum_{k=0}^{N-1} -\frac{s^2(k)}{\sigma_w^{2}} + \frac{2x(k)s(k)}{\sigma_w^{2}}$$

The resulting LLR test is written, when λ is some threshold, as:

$$\ln\Lambda_{H_2,H_1} \underset{H_o}{\overset{H_1}{\gtrless}} \lambda$$

or

$$\frac{1}{2}\sum_{k=0}^{N-1} -\frac{s^2(k)}{\sigma_w^{2}} + \frac{2x(k)s(k)}{\sigma_w^{2}} \underset{H_o}{\overset{H_1}{\gtrless}} \lambda$$

Hence

$$\sum_{k=0}^{N-1} x(k)s(k) \underset{H_o}{\overset{H_1}{\gtrless}} 2\sigma_w^{2}\lambda + \sum_{k=0}^{N-1} s^2(k)$$

This is the matched filter (correlator) which is optimal for detecting a deterministic signal in AWGN (Whalen [66]). It is noted that the matched filter is usually derived within the framework of the maximisation of the output signal to noise ratio (Whalen [66]).

3.3.2 The Detection of a Coloured Gaussian Signal Process in Additive White Gaussian Noise

This detection problem is formulated as:

$$\text{Hypothesis } H_1: x(n) = z(n)+w(n), \; n=0,\ldots,N-1 \tag{3.33a}$$

$$\text{Hypothesis } H_o: x(n) = \quad w(n), \; n=0,\ldots,N-1 \tag{3.33b}$$

where $\{z(n)\}$ is a coloured Gaussian stochastic process.

It is a well-known property of Gaussian processes that the optimal (in the MMSE sense) prediction is a linear function of the past data (Doob [9]). Hence:

$$\hat{x}(k|H_1) = \sum_{\ell=0}^{k} a_k(\ell)x(k-\ell) \tag{3.34}$$

where the set of forward linear prediction coefficients $\{a_r,(\ell),\ \ell=1,\ldots,k;\ k=0,\ldots,N-1\}$ is such that the mean square error $E[(x(k)-\hat{x}(k|H_1))^2|H_1]$ is a minimum.

The adjective "forward" is used to describe the above set of linear prediction coefficients $\{a_k(\ell)\}$ so as to distinguish them from the set of backward linear prediction coefficients $\{b_k(\ell), \ell=1,\ldots,k; k=0,\ldots,N\}$ where the optimal backward prediction (in a MMSE sense) is a linear function of the future data [Haykin (19)].

The solution for the forward linear prediction coefficients $\{a_k(\ell)\}$ is usually written in terms of the set of normal equations, which may be expressed in matrix form (Box and Jenkins [6]) as

$$\underline{R}_x(k)\underline{a}_k = \underline{\sigma}^2(k|H_1) \tag{3.35}$$

where $\underline{R}_x(k)$ is the $(k+1)x(k+1)$ autocorrelation matrix of the data process $\{x(n) = z(n)+w(n),\ n=0,\ldots,k\}$, $\underline{a}_k$ is the $(k+1)$ vector of the k-th order forward linear prediction error coefficients, i.e. $\underline{a}_k = (1, -a_k(1),\ldots,-a_k(k))^T$, and $\underline{\sigma}^2(k|H_1)$ is the $(k+1)$ vector containing the k-th a priori forward prediction-error power :

$$\underline{\sigma}^2(k|H_1^T) = (\sigma^2(k|H_1),0,\ldots,0)^T$$

From (Eqn. 3.19) the LLR for the detection problem (Eqn. 3.33) may be written as:

$$\ln \Lambda_{H_1,H_o} = \frac{1}{2}\sum_{k=0}^{N-1}[\ln(\frac{\sigma_w^2}{\sigma^2(k|H_1)}) + \frac{x^2(k)}{\sigma_w^2} - \frac{\tilde{x}^2(k|H_1)}{\sigma^2(k|H_1)}] \tag{3.37}$$

where the k-th prediction error $\tilde{x}(k|H_1) = \underline{a}_k^T\underline{x}_k'$. Denoting the k-th data vector $\underline{x}_k^T=(x(o),\ldots,x(k))^T$, then the time reversed k-th data vector $\underline{x}_k'^T=(x(k),\ldots,x(o))^T$. As discussed above, the set of forward linear prediction error filter coefficients $\{\underline{a}_k, k=0,\ldots,N-1\}$ and the forward prediction error powers $\{\sigma^2(k|H_1)\}$ are obtained through a solution of the set of normal equations (Eqn. 3.35).

If the statistics of the Gaussian signal process $\{z(n)\}$ and of the AWGN process $\{w(n)\}$ are exactly known, then the solution of the forward linear prediction parameters $\{\underline{a}_k\}$, $\{\sigma^2(k|H_1)\}$ is trivial. However, if these statistics are not known, then the calculation of the forward (and the backward) linear prediction coefficients is performed in an adaptive manner (Friedlander [13]).

It should be noted that the linear prediction (Eqn. 3.34) which is used in the LLR (Eqn. 3.37) explicitly demands the use of the adaptively updated linear prediction coefficients. This follows from the definition of the IR where all past data must be taken into account in the calculation of the prediction error process (the innovations process).

This form of the LLR (Eqn. 3.37) agrees with the form of the LLR's as derived by both Scholtz et al [53] and by Porat and Friedlander [51].

3.3.3 The Detection of a Deterministic Signal in Additive Coloured Gaussian Noise

Consider the detection problem:

$$\text{Hypothesis } H_2: \; x(n) = s(n)+c(n)+w(n), \quad n=0,\ldots,N-1 \qquad (3.38a)$$

$$\text{Hypothesis } H_1: \; x(n) = \qquad c(n)+w(n), \quad n=0,\ldots,N-1 \qquad (3.38b)$$

where the signal process $\{s(n)\}$ is a known deterministic signal, $\{c(n)\}$ is a coloured Gaussian process, and $\{w(n)\}$ is the usual WGN process.

From (Eqn. 3.32) the LLR for this detection problem (Eqn. 3.38) has the form:

$$\ln \Lambda_{H_2,H_1} = \frac{1}{2} \sum_{k=0}^{N-1} \left[\ln\left(\frac{\sigma^2(k|H_1)}{\sigma^2(k|H_2)}\right) + \frac{(x(k)-\hat{x}(k|H_1))^2}{\sigma^2(k|H_1)} - \frac{(x(k)-\hat{x}(k|H_2))^2}{\sigma^2(k|H_2)} \right] \qquad (3.39)$$

where $\hat{x}(k|H_i)$ is the optimal (in the MMSE sense) prediction of $x(k)$ assuming Hypothesis H_i, $i=1,2$ is correct. Similarly, $\sigma^2(k|H_i) = E[(\hat{x}(k)-x(k|H_i))^2 | H_i]$ is the k-th prediction error power under Hypothesis H_i, $i=1,2$.

Since, under Hypothesis H_1, $\{x(n)\}$ is a Gaussian process, then:

$$s(k|H_1) = \sum_{\ell=1}^{k} a_k(\ell)\, x(k-\ell), \qquad k=0,\ldots,N-1$$

where the set of f[illegible] linear prediction coefficients $\{a_k(\ell)\}$ are the solution to the set of normal equations

$$\underline{R}_x(k)\underline{a}_k = \sigma^2(k|H_1) \qquad (3.40)$$

However, it should be noted that the process $\{x(k)-s(k)|H_2\}$ has exactly the same Gaussian statistics as does the process $\{x(k)|H_1\}$. Hence,

$$\hat{x}(k|H_2)-s(k) = \sum_{\ell=1}^{k} a_k(\ell)(x(k)-s(k)), \quad k=0,\ldots,N-1 \qquad (3.41)$$

and it is trivial to demonstrate that:

$$\sigma^2(k|H_1) = \sigma^2(k|H_2), \quad k=0,\ldots,N-1 \qquad (3.42)$$

Thus using matrix notation, the LLR (Eqn. 3.39) is rewritten as:

$$\ln\Lambda_{H_2,H_1} = \frac{1}{2}\sum_{k=0}^{N-1}\frac{1}{\sigma^2(k|H_1)}[(\underline{a}^T(k)\underline{x}'(k))^2-(\underline{a}^T(k)(\underline{x}'(k)-\underline{s}'(k)))^2] \qquad (3.43)$$

where the k-th signal vector is denoted by $\underline{s}^T(k) = (s(o), \ldots, s(k))^T$, and the time reversed signal vector is denoted by $\underline{s}'^T(k) = (s(k), \ldots, s(o))^T$.

If the statistics of the Gaussian process are not known a priori, then the linear prediction coefficients and powers are calculated in an adaptive manner (Friedlander

[13]).

This LLR (Eqn. 3.43) may be compared with the result obtained by Kay [38]. In this paper, concerned with the detection of a deterministic signal in AWGN of unknown variance, the result may be written in terms of the notation used in this thesis as follows:

$$\ln\Lambda^{(KAY)}_{H_2,H_1} = \frac{N}{2}[\ln \sum_{k=0}^{N-1} (\underline{a}^T(k)\underline{x}'(k))^2 - \ln \sum_{k=0}^{N-1} (\underline{a}^T(k)(\underline{x}'(k)-\underline{s}'(k)))^2] \quad (3.44)$$

Apart for the normalising quantity $\overline{\sigma^2(k|H_1)}$ in (Eqn. 3.43) and the use of the logarithm in (Eqn. 3.44) the LLRs are the same. Therefore, the performance of these two LLR tests will be identical.

Consider next the detection problem (Eqn. 3.38) where the statistics of the Gaussian process $\{c(n)+w(n)\}$ are exactly known. Although this makes not the slightest difference in the formulation of the LLR (Eqn. 3.43), it is a detection problem whose solution is well-known (deLong and Hofstetter [8], Taylor [62]) and is written as:

$$\ln\Lambda^{(MTD)}_{H_2,H_1} = (\underline{R}_x^{-1}(N-1)\underline{s}(N-1))^T\underline{x}(N-1) \quad (3.45)$$

The superscript (MTD) is employed in (Eqn. 3.45) to emphasise that this is the basis of the test statistic used in the MTD algorithms. This LLR may be derived in a number of ways (Brooks and Reed [7]) including the maximisation of the output signal to noise ratio (de Long and Hofstetter [6]). It is pertinent to this thesis to demonstrate that this LLR (Eqn. 3.45) is identical to the LLR (Eqn. 3.43) which results from an application of the IBDA to the same detection problem (Eqn. 3.38).

The Cholesky decomposition (Friedlander [13]), as applied to the autocorrelation matrix $\underline{R}_x(N|H_1)$ is written as:

$$\underline{R}_x^{-1}(N|H_1) = \underline{U}\ \underline{D}\ \underline{U}^T \tag{3.46}$$

where $\underline{U}$ is an upper diagonal matrix defined by various sets of backward linear prediction-error coefficients:

$$\underline{U} = \begin{bmatrix} 1 & -b_1(1) & \cdots & -b_{N-1}(N-1) \\ 0 & 1 & \cdots & -b_{N-1}(N-2) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

and $\underline{D}$ is a diagonal matrix defined by the inverses of the backward prediction-error powers, given that hypothesis H_1 is correct:

$$\underline{D} = \mathrm{diag}\left(\frac{1}{\sigma_b^2(0|H_1)}, \frac{1}{\sigma_b^2(1|H_1)}, \ldots, \frac{1}{\sigma_b^2(N-1|H_1)}\right).$$

The backward linear prediction error coefficients $\{\underline{b}_k^T = (-b_k(k),\ldots,1),\ k=0,\ldots,N-1\}$ and the backward prediction error powers $\{\sigma_b^2(k|H_1),\ k=0,\ldots,N-1\}$ are the solution to the set of normal equations:

$$\underline{R}_x(k)\underline{b}_k = \underline{\sigma}_b^2(k), \quad k=0,\ldots,N-1$$

where $\underline{\sigma}_b^2(k)^T = (0,\ldots,0,\ \sigma_b^2(k))^T$.

The above Cholesky decomposition may now be rewritten:

$$\underline{R}_x^{-1}(N-1) = \begin{bmatrix} 1 & -b_1(1) & \ldots & -b_{N-1}(N-1) \\ 0 & 1 & & -b_{N-1}(N-2) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 0 \end{bmatrix} \times \left(\frac{1}{\sigma_b^2(0|H_1)} \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & & 0 \end{bmatrix} + \right.$$

$$\left. + \frac{1}{\sigma_b^2(1|H_1)} \begin{bmatrix} 0 & 0 & \ldots & 0 \\ -b_1(1) & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 0 \end{bmatrix} + \ldots + \frac{1}{\sigma_b^2(N-1|H_1)} \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & & \\ -b_{N-1}(N-1) & -b_{N-1}(N-2) & \ldots & 1 \end{bmatrix} \right)$$

Hence,

$$\underset{\rightarrow}{R}_x^{-1}(N-1)=\frac{1}{\sigma_b^2(0|H_1)}\begin{bmatrix}1 & 0\ldots0\\ 0 & 0\ldots0\\ \vdots & \vdots & \vdots\\ 0 & 0 & 0\end{bmatrix}+\frac{1}{\sigma_b^2(1|H_1)}\begin{bmatrix}0 & -b_1(1)\ldots0\\ 0 & 1 & \ldots0\\ \vdots & \vdots & \vdots\\ 0 & 0 & \ldots0\end{bmatrix}\begin{bmatrix}0 & 0\ldots0\\ -b_1(1) & 1\ldots0\\ \vdots & \vdots & \vdots\\ 0 & 0 & 0\end{bmatrix}+\ldots$$

$$+\frac{1}{\sigma_b^2(N-1|H_1)}\begin{bmatrix}0 & 0\ldots-b_{N-1}(N-1)\\ 0 & 0 & -b_{N-1}(N-2)\\ \vdots & \vdots & \vdots\\ 0 & 0\ldots & 1\end{bmatrix}\begin{bmatrix}0 & 0 & \ldots & 0\\ 0 & 0 & \ldots & 0\\ \vdots & \vdots & & \vdots\\ -b_{N-1}(N-1) & -b_{N-1}(N-2) & & 1\end{bmatrix}$$

Thus, upon further expansion, then:

$$\underset{\rightarrow}{R}_x^{-1}(N-1)=\frac{1}{\sigma_b^2(0|H_1)}\begin{bmatrix}1\\ 0\\ \vdots\\ 0\end{bmatrix}\begin{bmatrix}1 & 0\ldots0\end{bmatrix}+\frac{1}{\sigma_b^2(1|H_1)}\begin{bmatrix}-b_1(1)\\ 1\\ \vdots\\ 0\end{bmatrix}\begin{bmatrix}-b_1(1) & 1\ldots0\end{bmatrix}+$$

$$+\ldots+\frac{1}{\sigma_b^2(N-1|H_1)}\begin{bmatrix}-b_{N-1}(N-1)\\ -b_{N-1}(N-2)\\ \vdots\\ 1\end{bmatrix}\begin{bmatrix}-b_{N-1}(N-1) & -b_{N-1}(N-1)\ldots1\end{bmatrix} \qquad (3.47)$$

Now, using matrix notation, (Eqn. 3.47) is rewritten as:

$$\underline{R}_x^{-1}(N-1) = \frac{1}{\sigma_b^2(0|H_1)}\,\underline{b}_o\underline{b}_o^T + \frac{1}{\sigma_b^2(1|H_1)}\,\underline{b}_1\underline{b}_1^T + \ldots + \frac{1}{\sigma_b^2(N-1|H_1)}\,\underline{b}_{N-1}\underline{b}_{N-1}^T$$

$$= \sum_{k=0}^{N-1} \frac{1}{\sigma_b^2(k|H_1)}\,\underline{b}_k\underline{b}_k^T \qquad (3.48)$$

Hence, from (Eqn. 3.45):

$$\ln\Lambda_{H_2,H_1}^{(MTD)} = (\underline{R}_x^{-1}(N-1)\underline{s}(N-1))^T\underline{x}(N-1)$$

$$= \underline{x}^T(N-1)(\underline{R}^{-1}(N-1)\underline{s}(N-1))$$

$$= \underline{x}^T(N-1)\Big(\sum_{k=0}^{N-1} \frac{1}{\sigma_b^2(k|H_1)}\,\underline{b}_k\underline{b}_k^T\Big)\underline{s}(N-1)$$

$$= \sum_{k=0}^{N-1} \frac{1}{\sigma_b^2(k|H_1)}\,\underline{x}^T(k)\underline{b}_k\underline{b}_k^T\underline{s}(k) \qquad (3.49)$$

Since for stationary processes, the forward and backward prediction parameters are related by $\underline{b}_k = \underline{a}_k'$, then

$$\ln\Lambda_{H_2,H_1}^{(MTD)} = \sum_{k=0}^{N-1} \frac{1}{\sigma^2(k|H_1)}\,\underline{x}^T(k)\underline{a}_k'\underline{a}_k'^T\underline{s}(k) \qquad (3.50)$$

Returning now to the IBDA derived LLR (Eqn. 3.43) where:

$$\ln\Lambda_{H_2,H_1} = \frac{1}{2}\sum_{k=0}^{N-1} \frac{1}{\sigma^2(k|H_1)}[(\underline{a}^T(k)\underline{x}'(k))^2-(\underline{a}^T(k)(\underline{x}'(k)-\underline{s}'(k)))^2]$$

then

$$\ln\Lambda_{H_2,H_1} = \frac{1}{2}\sum_{k=0}^{N-1} \frac{1}{\sigma^2(k|H_1)}[-(\underline{a}^T(k)\underline{s}'(k))^2+2\underline{a}^T(k)\underline{x}'(k)\underline{a}^T(k)\underline{s}'(k)]$$

or

$$2\ln\Lambda_{H_2,H_1} + \sum_{k=0}^{N-1} \frac{1}{\sigma^2(k|H_1)}\underline{a}^T(k)\underline{x}'(k)\underline{a}^T(k)\underline{s}'(k)$$

$$= \sum_{k=0}^{N-1} \frac{1}{\sigma^2(k|H_1)}\underline{a}^T(k)\underline{x}'(k)\underline{a}^T(k)\underline{s}'(k)$$

$$= \sum_{k=0}^{N-1} \frac{1}{\sigma^2(k|H_1)}\underline{x}'^T(k)\underline{a}(k)\underline{a}^T(k)\underline{s}'(k)$$

$$= \sum_{k=0}^{N-1} \frac{1}{\sigma^2(k|H_1)}\underline{x}^T(k)\underline{a}'(k)\underline{a}'^T(k)\underline{s}(k) \qquad (3.51)$$

where, as before, the prime devotes time reversal.

Since the second term on the LHS of (Eqn. 3.51) is a constant, the quantity on the RHS of (Eqn. 3.51) is the test statistic for the LLR test resulting from an application of the IBDA to the detection problem (Eqn. 3.38). This test statistic is the same as that on the RHS of (Eqn. 3.49) which is the test statistic for the original MTD algorithm (Eqn. 3.45). This clearly shows that the MTD algorithm is a special case of the IBDA.

3.3.4 The Detection of a Coloured Gaussian Signal Process in Additive Coloured Gaussian Noise

This detection problem is formulated as:

$$\text{Hypothesis } H_2:\ x(n) = z_2(n)+z_1(n)+w(n), \quad n=0,\ldots,N-1 \tag{3.52a}$$

$$\text{Hypothesis } H_1:\ x(n) = z_1(n)+w(n), \quad n=0,\ldots,N-1 \tag{3.52b}$$

where both the processes $\{z_1(n)\}$ and $\{z_2(n)\}$ are coloured Gaussian processes.

Since the observation process under either the two hypotheses is Gaussian, then linear prediction is optimal and the LLR for this detection problem is, from (Eqn. 3.32):

$$\ln\Lambda_{H_2,H_1} = \frac{1}{2}\sum_{k=0}^{N-1}\left[\ln\left(\frac{\sigma^2(k|H_1)}{\sigma^2(k|H_2)}\right) + \frac{\tilde{x}^2(k|H_1)}{\sigma^2(k|H_1)} - \frac{\tilde{x}^2(k|H_2)}{\sigma^2(k|H_2)}\right] \tag{3.53}$$

where $\tilde{x}(k|H_i) = \underline{a}^T(k|H_i)\underline{x}(k)$, $i=1,2$. The sets of prediction error coefficients $\{\underline{a}(k|H_i),\ k=0,\ldots,N-1;\ i=1,2\}$ and the a priori prediction error powers $\{\sigma^2(k|H_i),\ k=0,\ldots N-1;\ i=1,2\}$ are the solutions to the sets of normal equation:

$$\underline{R}_x(k|H_i)\underline{a}(k|H_i) = \underline{\sigma}^2(k|H_i),\quad k=0,\ldots,N-1,\ i=1,2$$

The dependence of the LLR on the number of data samples N is explicitly illustrated by writing:

$$\ln\Lambda_{H_2,H_1} = \ln\Lambda_{H_2,H_1}(N)$$

where $\ln\Lambda_{H_2,H_1}$ is the LHS of [Eqn. 3.53).

This LLR may be trivially rewritten so that it may be utilised as a sequential detection algorithm by noting that from (Eqn. 3.53):

$$\ln\Lambda_{H_2,H_1}(N)=\ln\Lambda_{H_2,H_1}(N-1)+\ln\left(\frac{\sigma^2(N-1|H_1)}{\sigma^2(N-1|H_2)}\right)+\frac{\tilde{x}^2(N-1|H_1)}{\sigma^2(N-1|H_1)}-\frac{\tilde{x}^2(N-1|H_2)}{\sigma^2(N-1|H_2)}$$

This is exactly the sequential detection algorithm proposed by Therrien [63].

3.4 Summary

In this chapter, a very rapidly convergent[1] solution to the problem of detecting a stationary discrete-time stochastic process is AWGN is derived (Eqn. 3.26). This solution, which is referred to as the Innovations-Based Detection Algorithm (IBDA) is then applied, in section 3.2, to the detection of a stochastic process in non-Gaussian noise.

In order to illustrate the usefulness of the IBDA as a detection problem solving tool, in section 3.3 the IBDA is applied to several specific Gaussian detection problems.

It should be noted that two of these specific problems involved the detection of deterministic signal processes. Hence, the above requirement of stationarity of the received data process is violated, since deterministic signals when considered in the context of stochastic processes are nonstationary. Yet, as demonstrated in section 3.3.1 and section 3.3.3 the IBDA yields the correct test statistic as confirmed by comparison with the results obtained through different methods such as the maximisation of the output signal to noise ratio (Whalen [66], deLong and Hofstetter [8]). Hence, it can be concluded that this requirement of strict

[1]This rate of convergence is rapid in the context of asymptotic normality.

sense stationarity is only a sufficient condition, and is not a necessary condition.

CHAPTER 4

THE RADAR TARGET DETECTION PROBLEM

In Chapter 3, the LLR (Eqn. 3.21) for the detection of a discrete-time stochastic process in AWGN is derived. This specific problem is chosen since it may be used as a basic building block in developing detection algorithms for many radar, communication and sonar systems.

In this chapter, the IBDA is applied to a much more difficult detection problem, the detection of targets by airport surveillance radar (ASR) systems (Metford and Haykin [45]). As discussed in section 4.1, the target, which is characterised in a statistical manner must be detected against a background of both Gaussian and non-Gaussian stochastic processes. In addition, all of these processes may be extremely dynamic. Hence, the simple analytic comparisons of the form employed in section 3.3 are not as easily made since any detection scheme designed for an ASR system will be a solution to some particular model of the radar target detection problem. Due to the theoretical ability of the IBDA to be applied to the detection of a stochastic signal in non-Gaussian noise (section 3.2), a much more robust model of the radar detection problem can be employed than is possible with the classical detection problem solving techniques.

To facilitate performance comparisons, high performance detection algorithms for the radar target detection problem were made available to the author of this thesis. These algorithms, the MTD-1 (Drury [10]), the MTD-2 (Karp and Anderson [37]) and the MTD-3 (Taylor [62]) have all been successfully implemented on modern ASR systems and represent the application of classical detection theory to the radar target detection problem.

Performance evaluations and comparisons are made using real radar data recorded on a digital tape recorder from a modern ASR system and, later transferred into a signal processing computer. The use of real radar data, as opposed to simulated radar data, strengthens the conclusions and recommendations of this thesis.

In section 4.1 a general description of ASR systems is presented, along with a discussion of the statistical properties of the radar hypothesis testing problem. The IBDA is applied to this radar target detection problem in section 4.2. The details of the implementation of the IBDA are discussed in section 4.3. The MTD algorithms, which provide the performance standards against which the IBDA is to ~~be~~ tested, are outlined in section 4.4. A summary of this chapter is given in section 4.5.

Performance evaluations and comparisons are made in Chapter 5.

4.1 Airport Surveillance Radar Systems

One of the most important applications of radar is as a sensor in the air traffic control network. These high power, long range radars typically operate in the L-band (1GHz) region of the radio frequency spectrum. The radar is ground-based, with the slowly rotating antenna emitting uniform pulses of energy. Both the transmitter and the receiver are connected via a duplexer to the common antenna. Modern ASR systems are fully coherent and digitally process the received video inphase and quadrature channels for each range-azimuth cell (Skolnik [59]).

The present policy of the Federal Aviation Administration is to completely automate the air traffic control network (Muehe et al. [48]). Although most ASR systems give good performance when manually operated by skilled airtraffic controllers, the automatic systems perform poorly due to an excessive number of missed detections and false alarms.

This unsatisfactory performance is due to the corruption of the target echos by clutter. Clutter, in this context, is an aggregate term for the echos from everything but the target such as ground and weather returns. Also, implicit in any discussion of the radar target detection problem is the receiver noise which corrupts the incoming reflections from the radar's environment. This corruption is invariably modelled as additive white Gaussian noise (Van Trees [65]).

The radar target detection problem is formalised in terms of the hypothesis testing problem:

$$\text{Hypothesis } H_2\text{: } x(n) = s(n)+c(n)+w(n), \quad n=\ell-N+1,\ldots\ell \qquad (4.1a)$$

$$\text{Hypothesis } H_1\text{: } x(n) = c(n)+w(n), \quad n=\ell-N+1,\ldots\ell \qquad (4.1b)$$

where $\{x(n)\}$ is the complex baseband signal from the n-th range azimuth cell, $\{s(n)\}$ is the target process, $\{c(n)\}$ is the usually non-Gaussian clutter process and $\{w(n)\}$ is the WGN process.

The data process $\{x(n), n=\ell-N+1,\ldots,\ell\}$ is acquired by sampling the N consecutive radar returns from a specific range ring as the radar scans across the ℓ-th azimuth cell. Hence N is the number of pulses illuminating the range-azimuth cell of interest, and is directly related to the antenna beam width and the scanning speed of the antenna.

The WGN process is assumed to be statistically independent of both the target process and the clutter process, (Van Trees [65]). However, the target process and the clutter process may not be mutually statistically independent since both are composed of echos generated by the same pulse of transmitted energy. The received signal process $\{x(n),\ n=\ell-N+1,\ldots,\ \ell\}$ may be assumed to be stationary since it consists of the echos from a single range-azimuth cell from a single antenna scan (Van Trees [65]).

Clutter is usually classified into one of three groups. These are ground clutter, weather clutter and sea clutter (Barton [4]). Sea clutter is not normally observed by ASR systems and will not be included in the following discussions.

Due to the macroscopic complexity of the natural phenomena which generate ground clutter and weather clutter, statistical modelling of a given clutter process $\{c(n)\}$ is parametrised in terms of the first order pdf (i.e. the pdf of the random variable $c(n)$) and the second order statistics (i.e. the autocorrelation function or equivalently, the power spectrum (Haykin [20]) of the stochastic process $\{c(n)\}$.

Weather clutter is generated by the superposition of the reflections from the many identical point scatterers (raindrops, etc.) within the illuminated weather pattern. As such, due to the central limit theorem (Loeve [43]), it might be expected that the random variable c(n) is Gaussian distributed. This has been confirmed experimentally (Barton [47]).

It has been observed that the power spectrum of weather clutter can be approximated by a smoothly varying function of frequency. (Haykin ef al. [21]). The centroid of the power spectrum is situated at the overall net doppler frequency of the weather pattern. The width of the power spectrum is directly related to the internal turbulence of the weather pattern (Barton [4]). Hence weather clutter may be characterized as a correlated Gaussian process with a smoothly varying power spectrum.

Due to the widely varying nature of the reflectors which generate ground clutter, the amplitude statistics (i.e. the pdf of the random variable c(n)) for ground clutter processes are not as easily characterised as are the amplitude statistics of weather clutter. According to Sekine et al [56], for ground clutter, the complex envelope $r(n) = (Re^2[c(n)]+Im^2[c(n)])^{1/2}$ is experimentally observed to be Weibull distributed. The Weibull distribution:

$$p(r;a) = \frac{a}{b}\left(\frac{r}{b}\right)^{a-1} \exp-\left(\frac{r}{b}\right)^{a}, \qquad r,\ a,\ b > 0$$

is parametrised by the constant a. The other constant, b may be regarded as a scaling factor.

Sekine et al. [56] demonstrated that for an L band radar, this shape parameter a varies from 1.507 to 2.0 for the clutter from cultivated farm land. However, for clutter generated by low rolling wooded hills and grassland, a = 0.626. The critical value is for a = 2, where the Weibull distribution possesses the functional form of the Rayleigh distribution (Sekine et al. [56]). The Rayleigh distribution is of particular interest since if the inphase and quadrature components of a complex random variable are independent and identically distributed Gaussian random variables, then the resultant complex envelope is Rayleigh distributed (Papoulis [50]). As a result, it is seen that unless the shape parameter a = 2.0 then the clutter is non-Gaussian. Hence, it is concluded that ground clutter is best characterised as non-Gaussian.

Since the natural phenomena which generate ground clutter are usually stationary, the centroid of the power spectrum is centred at zero doppler. The shape of this function is dominated by the antenna beam pattern which imparts a modulation on the reflectivity of a specific ground clutter cell (Barton [4]). Since antenna patterns are adequately modelled (for analytic purposes) as having a Gaussian shape, then the power spectrum for ground clutter will also posess this same functional form.

Hence, it is assumed that the ground clutter process is non-Gaussian, with a Gaussian shaped power spectrum.

As mentioned in the introduction to this chapter, the target process {s(n)} may be characterised in a statistical manner. An arbitrary target is essentially a point reflector with a random amplitude and doppler response (Van Trees [65]). Hence, due to the modulation imposed upon the target response by the Gaussian shaped antenna pattern, the power spectrum of the target process may also be modelled as a Gaussian shaped function, centred at the doppler frequency of the target.

4.2 The Application of the Innovations Based Detection Algorithm to the Radar Target Detection Problem

As derived in section 3.1, the IBDA is the LLR for the detection problem:

Hypothesis H_1: $x(n) = z(n)+w(n)$, $\quad n=0,\ldots N-1$ (4.2a)

Hypothesis H_o: $x(n) = \quad w(n)$, $\quad n=0,\ldots N-1$ (4.2b)

where all of the pertinent processes are real valued. The LLR has the form:

$$\ln \Lambda_{H_1,H_o} = \frac{1}{2}\sum_{k=0}^{N-1}[\ln(\frac{\sigma_w^2}{\sigma^2(k|H_1)})+\frac{x^2(k)}{\sigma_w^2}-\frac{\tilde{x}^2(k|H_1)}{\sigma^2(k|H_1)}] \qquad (4.3)$$

From Woodward [68] it is easily seen that the LLR (Eqn. 4.3) for the above real detection problem (Eqn. 4.2) may easily be generalised such that the LLR for the complex detection problem:

Hypothesis H_1: $x(n) = z(n)+w(n)$, $\quad n=0,\ldots,N-1$ (4.4a)

Hypothesis H_o: $x(n) = \quad w(n)$, $\quad n=0,\ldots,N-1$ (4.4b)

where all of the pertinent processes are complex has the form:

$$\ln \Lambda_{H_2,H_1} = \frac{1}{2}\sum_{k=0}^{N-1}[\ln(\frac{\sigma_w^2}{\sigma^2(k|H_1)})+\frac{|x(k)|^2}{\sigma_w^2}-\frac{|\vec{x}(k|H_1)|^2}{\sigma^2(k|H_1)} \qquad (4.5)$$

where the variance σ_w^2 of the WGN process $\{w(n)\}$ is defined:

$$\sigma_w^2 \triangleq E[|w(n)|^2], \qquad n=0,\ldots,N-1$$

and the k-th apriori prediction error power $\sigma^2(k|H_1)$ is defined as:

$$\sigma^2(k|H_1) \triangleq E[|\tilde{x}(k(H_1)|^2|H_1], \quad k=0,\ldots,N-1 \qquad (4.6)$$

The radar target detection problem (Eqn. 4.1) may now be solved in terms of the solution (Eqn. 4.5) to the complex detection problem (Eqn. 4.4) in exactly the same manner as is used in section 3.2. Specifically, since the data process under either of the two hypotheses is a finite energy process (Abbot and Blum [2]) then through the use of the chain rule of LR's, the LLR for the radar target detection problem (Eqn. 4.1) is written:

$$\ln \Lambda_{H_2,H_1}(\ell) = \sum_{k=\ell-N+1}^{\ell} [\ln(\frac{\sigma^2(k|H_1)}{\sigma^2(k|H_2)}) + \frac{|\tilde{x}(k|H_1)|^2}{\sigma^2(k|H_1)} - \frac{|\tilde{x}(k|H_2)|^2}{\sigma^2(k|H_2)}] \qquad (4.7)$$

where the index ℓ has been explicitly introduced to emphasise that this is the LLR for the ℓ-th azimuth cell and where the k-th a priori prediction error power assuming Hypothesis H_i:

$$\sigma^2(k|H_i) \triangleq E[|\tilde{x}(k|H_i)|^2|H_i] \qquad (4.8)$$

and $\tilde{x}(k|H_i)$ is the k-th prediction error assuming Hypothesis H_i, i=1,2. Thus knowledge of the statistical properties of x(k) under either of the two hypotheses, Hypothesis H_1 and Hypothesis H_2, is required in order to construct the two needed sets of prediction errors and a priori prediction error powers.

This LLR (Eqn. 4.7) may, in the context of radar signal processing, be referred to as a block processing algorithm in that the block of data $\{x(\ell-N+1),\ldots, x(\ell)\}$ can be considered to have been obtained by viewing the incoming data stream $\{\ldots, x(-1), x(0), x(1),\ldots\}$ from the range ring of interest through a window of length N. But detection decisions must be made not only on the ℓ-th range-azimuth cell represented by the data set $\{x(\ell-N+1), \ldots,x(\ell)\}$ but also upon all of the remaining range-azimuth cells. These include the preceding (along lines of constant range) range-azimuth cell represented by the data set $\{x(\ell-N),\ldots, x(\ell-1)\}$ and the following range-azimuth cell, represented by the data set $\{x(\ell-N+2),\ldots,x(\ell+1)\}$. Thus each detection problem is highly correlated with the preceding and the following detection problems.

This high degree of data set congruence may be taken advantage of by modifying the definition of the prediction errors. The k-th prediction error under Hypothesis H_i is calculated upon the basis of the past k data samples. The needed modification is to calculate the prediction errors upon the basis of the preceding

M data samples, where M is fixed and to be determined empirically. If M is sufficiently large, then the resulting prediction error processes will not be significantly statistically different then if all past data were used as a basis.

The LLR for the ℓ-th range-azimuth cell is now written as:

$$\ln \Lambda_{H_2,H_1}(\ell) = \frac{1}{2}\sum_{k=\ell-N+1}^{\ell}\left[\ln\left(\frac{\sigma^2(k|H_1)}{\sigma^2(k|H_2)}\right)+\frac{|\tilde{x}(k|H_1)|^2}{\sigma^2(k|H_1)}-\frac{|\tilde{x}(k|H_2)|^2}{\sigma^2(k|H_2)}\right] \quad (4.9)$$

where the k-th prediction error, $\tilde{x}(k|H_i)$, assuming Hypothesis H_i, is calculated upon the basis of the past M data samples $\{x(k-1),\ldots,x(k-M)\}$.

It is easily seen that:

$$\ln \Lambda_{H_2,H_1}(\ell) = \frac{1}{2}\left[\ln\left(\frac{\sigma^2(\ell|H_1)}{\sigma^2(\ell|H_2)}+\frac{|\tilde{x}(\ell|H_1)|^2}{\sigma^2(\ell|H_1)}-\frac{|\tilde{x}^2(\ell|H_2)|^2}{\sigma^2(\ell|H_2)}\right)\right] +$$

$$+ \ln \Lambda_{H_2,H_1}(\ell-1)$$

$$-\frac{1}{2}\left[\ln\left(\frac{\sigma^2(\ell-N|H_1)}{\sigma^2(\ell-N|H_2)}\right)+\frac{|\tilde{x}(\ell-N|H_1)|^2}{\sigma^2(\ell-N|H_1)}-\frac{|\tilde{x}(\ell-N|H_2)|^2}{\sigma^2(\ell-N|H_2)} \quad (4.10)$$

This structure is efficiently implemented by loading the

ℓ-th incremental term:

$$\frac{1}{2}[\ln(\frac{\sigma^2(\ell|H_1)}{\sigma^2(\ell|H_2)} + \frac{|\tilde{x}(\ell|H_1)|^2}{\sigma^2(\ell|H_1)} - \frac{|\tilde{x}^2(\ell|H_2)|^2}{\sigma^2(\ell|H_2)}] \qquad (4.11)$$

into a shift register of length N and summing the contents.

This LLR (Eqn. 4.10) is now compared with a threshold $\lambda(\ell)$ which is preset for the desired probability of false alarm. Statisticians refer to this method of threshold setting as the Neyman-Pearson criterion (Van Trees [64]). In radar, this is referred to as a constant false alarm rate (CFAR) threshold (Van Trees [64]).

The LLR test for this ℓ-th range azimuth cell is written as:

$$\ln \Lambda_{H_2,H_1}(\ell) \underset{H_1}{\overset{H_2}{\gtrless}} \lambda(\ell) \qquad (4.12)$$

A block diagram of this LLR test is shown in Figure 4.1.

4.3 Two Adaptive Implementations of the Innovations Based Detection Algorithm

The general structure of the LLR test as obtained from the application of the IBDA to ASR systems is derived in the previous section. Consideration is now given to the calculation of the prediction errors, the a priori

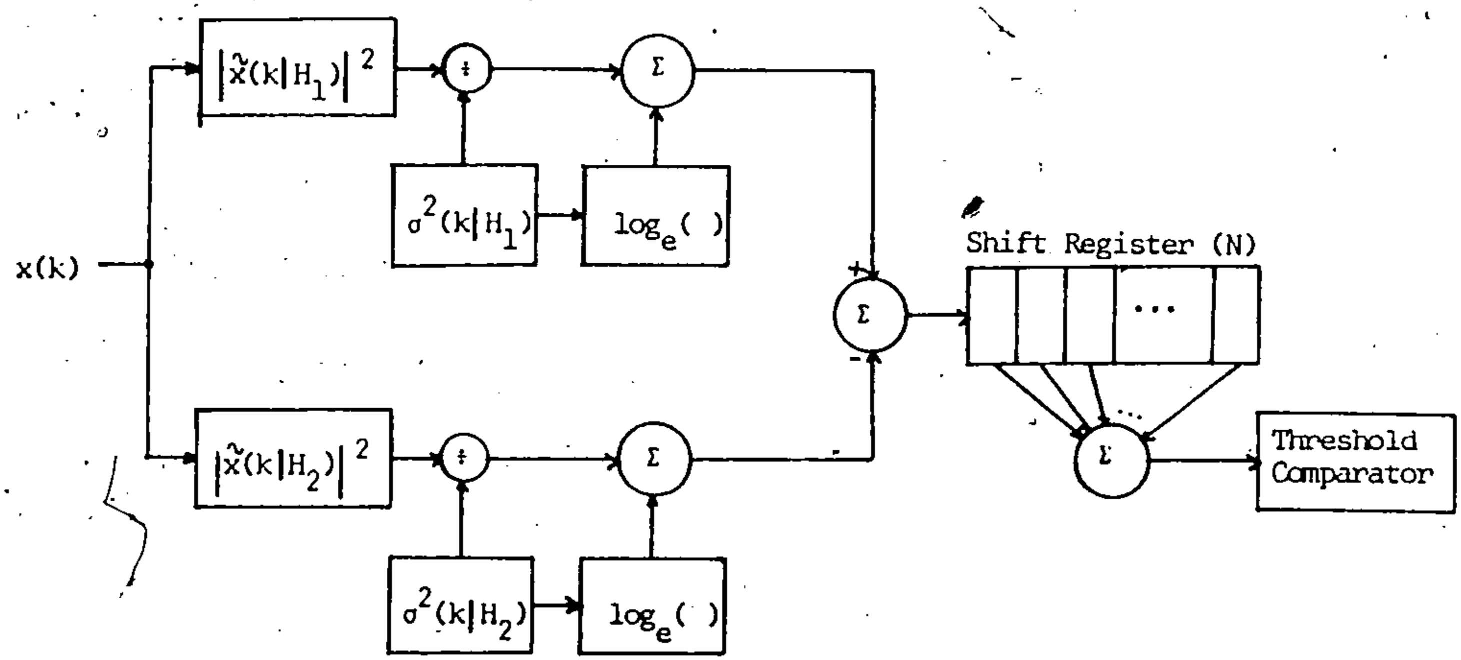


Fig. 4.1 A block diagram of the LLR test resulting from the application of the IBDA to the radar target detection problem

prediction error powers and the CFAR threshold.

As mentioned in section 4.2, knowledge of the statistical properties of the received data under either of the two hypotheses (Eqn. 4.1a, Eqn. 4.1b) is required in order to construct the two needed sets of prediction errors and the a priori prediction error powers. The experimentally observed statistical properties of ground clutter and weather clutter as discussed in section 4.1 are now utilised in the design of the two different implementations of the LLR (Eqn. 4.10).

These two implementations are referred to as the IBDA (LSL) and the IBDA (KALMAN) and differ only in the method by which the prediction errors are calculated. The two algorithms used in the prediction error calculations are developed in section 4.3.1. The methods used in the calculation of the a priori prediction error powers and the CFAR threshold are discussed in section 4.3.2 and section 4.3.3 respectively.

4.3.1 The Adaptive Implementation of the Prediction Error Structure

Although linear prediction is only optimal for Gaussian processes (Doob [9]), there is usually little choice. Non-linear prediction in general is not as highly developed, and requires an extensive knowledge of the statistical properties of the data. However, linear prediction depends solely upon the second order statistics of the data. In addition, adaptive linear prediction algorithms (Gibson [16], Shichor [58], Zhang and Haykin [69], Friedlander [13]) are simple and efficient. This adaptivity is pertinent to the above sequential method of calculating the LLR (Eqn. 4.10) where the prediction errors are calculated on the basis of the past M data samples.

The LLR (Eqn. 4.10) requires two parallel prediction error filters (Figure 4.1). One of the prediction error filters is designed to be optimal for Hypothesis H_2 (target plus clutter plus noise), while the other prediction error filter is designed to be optimal for Hypothesis H_1 (clutter plus noise). This is intuitively satisfying. Since the power spectra for these two hypotheses are not exactly known a priori, in every range-azimuth cell of interest, simple robust models of the power spectrum must be constructed and used in conjunction with adaptive linear prediction error filters designed on the basis of the models.

The first model uses the observation that the width of the power spectra of ground and weather clutter processes is greater than the width of the power spectrum for target possesses (Gibson [16], Kesler [41]). Equivalently, it may be said that the decorrelation time under Hypothesis H_1 is larger than that for Hypothesis H_2.

This difference in decorrelation times may be accounted for in a transparent fashion in the prediction error lattice filter structure (Figure 4.2). The algorithm for this adaptive structure is implemented with a damping constant μ which controls the relaxation rate of the lattice algorithm (Gibson [16]). Thus the prediction error lattice filter for Hypothesis H_1 is designed with a larger damping factor than that used in the lattice filter for Hypothesis H_2.

Many different implementations of the adaptive linear prediction error lattice filter have been discussed in the literature (Friedlander [13]). These adaptive algorithms are classed as either a gradient search algorithm or the exact least squares technique. As pointed out by Friedlander [13], the gradient techniques are inferior in performance as compared with the true least squares algorithm. As a result, the least squares lattice (LSL) form (Schichor [58] is employed in this implementation of the IBDA. The necessary recursive equations are presented in Appendix 2.

The order M of the LSL (the number of lattice

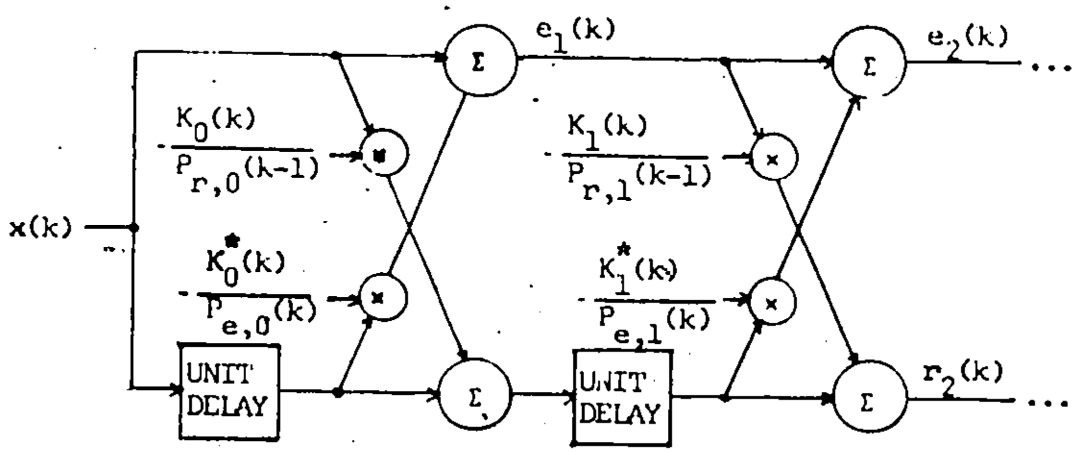


Fig. 4.2 The LSL prediction error filter

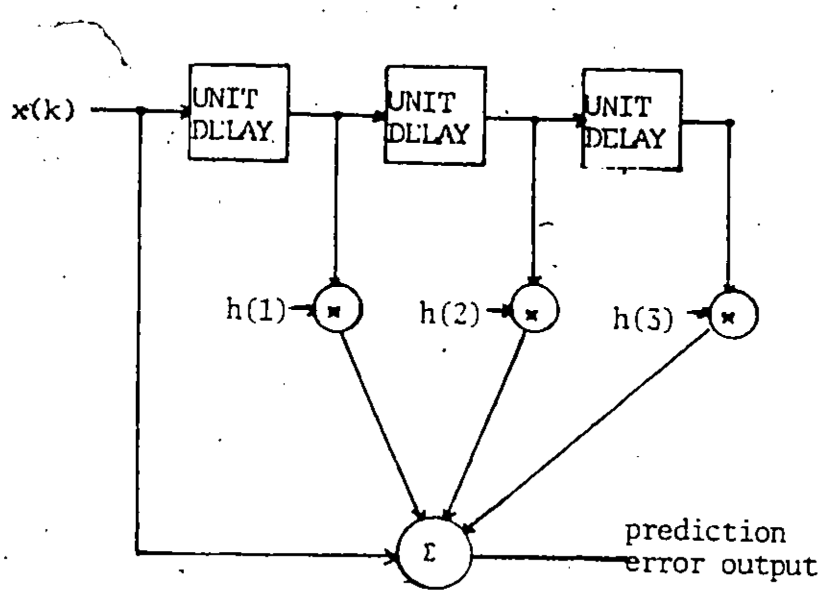


Fig. 4.3 A tapped delay-line filter

stages governed by the number of past data samples used in the calculation), and the precise values of the two damping constants (μ_{H_1}, μ_{H_2}) are determined empirically and then fixed. This implementation is referred to as the IBDA (LSL) (Figure 4.4).

The second model is based upon the following observation. In an air traffic control environment the antenna pattern is typically Gaussian shaped. Hence, the antenna scanning modulation has the effect of making the power spectra of the pertinant random processes assume a form which may be closely approximated by a Gaussian function of frequency. The random processes include the ground clutter and weather clutter and target echoes.

From Haykin et al. [19] a random process possessing a Gaussian shaped power spectrum may be adequately modelled as a low order autoregressive process (Box and Jenkins [6]). An autoregressive process of order M (AR(M)) is a stochastic process {y(n)} such that

$$y(n) = \sum_{k=1}^{M} a_k y(n-k)+w(n) \tag{4.13}$$

where the "shock" process {w(n)} is a white noise process with variance $\sigma_w^2 = E[|w(n)|^2]$. The set of AR coefficients $\{a_k, k=1,...,M\}$ and the variance σ_w^2

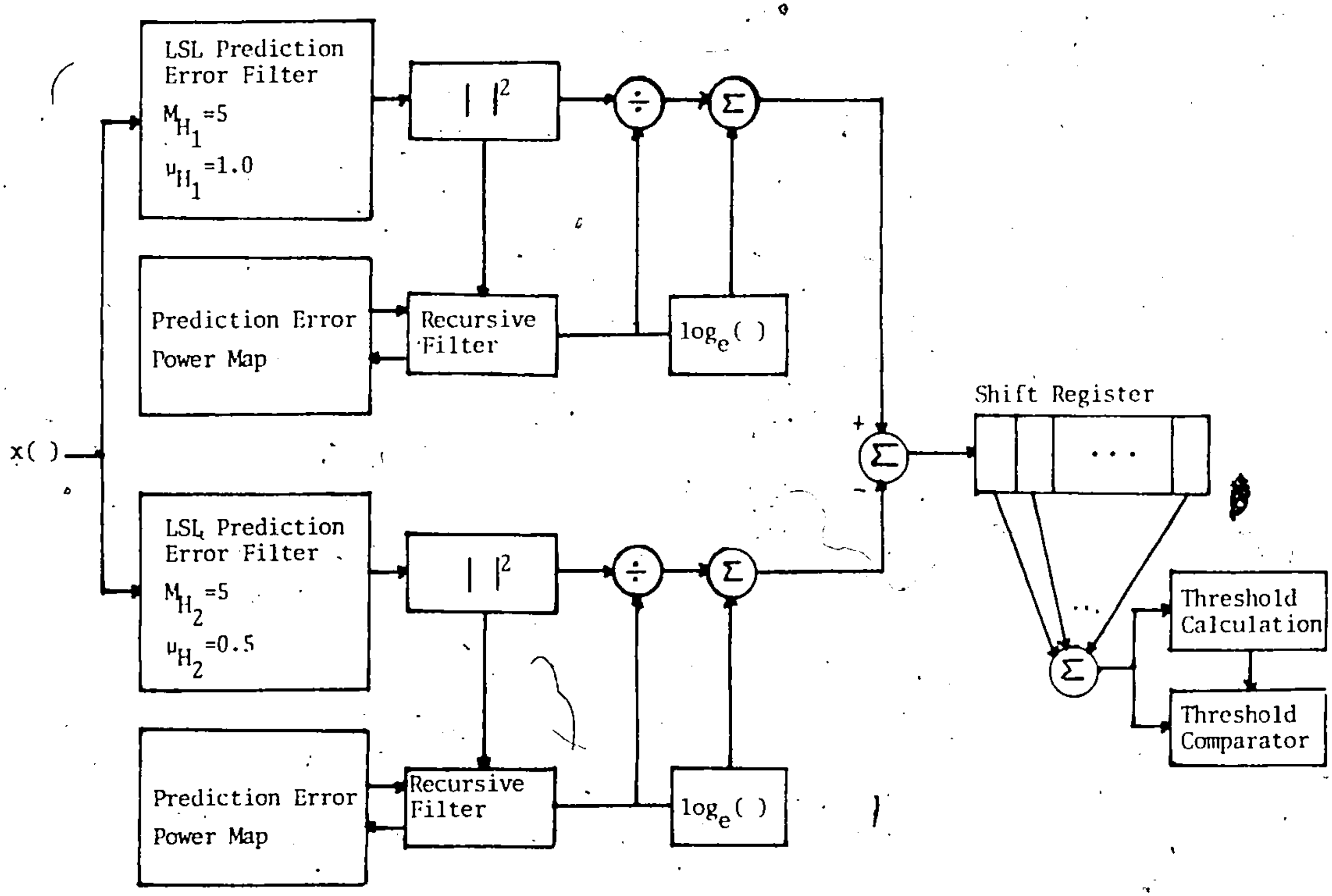


Fig. 4.4 The IBDA(LSL) implementation

of the white noise process are related through the set of normal equations similar to those used in linear prediction theory (Eqn. 3.35).

The power spectrum $S_y(f)$ for an AR(M) process is written as:

$$S_y(f) = \frac{\sigma_w^2}{\left|1+\sum_{k=1}^{M} a_k \exp[-2\pi j f k]\right|^2} \tag{4.14}$$

As demonstrated in Figure 4.5 the AR coefficients for an AR(3) can be adjusted so as to yield an acceptable fit of the power spectrum $S_y(f)$ to a Gaussian function. Hence, since under Hypothesis H_1 (no target present) and assuming a relatively small background white noise power (i.e. a clutter dominated environment) a first or a second order linear prediction will be nearly optimal.

Under Hypothesis H_2 (target present), and again assuming a relatively small background white noise power, the received data is complicated by the presence of the radar return due to the target. Since the target process possesses a Gaussian spectrum, the total received data process may be approximated as the sum of two first or second order AR processes. From Box and Jenkins [6], the sum of two AR processes is an autoregressive moving

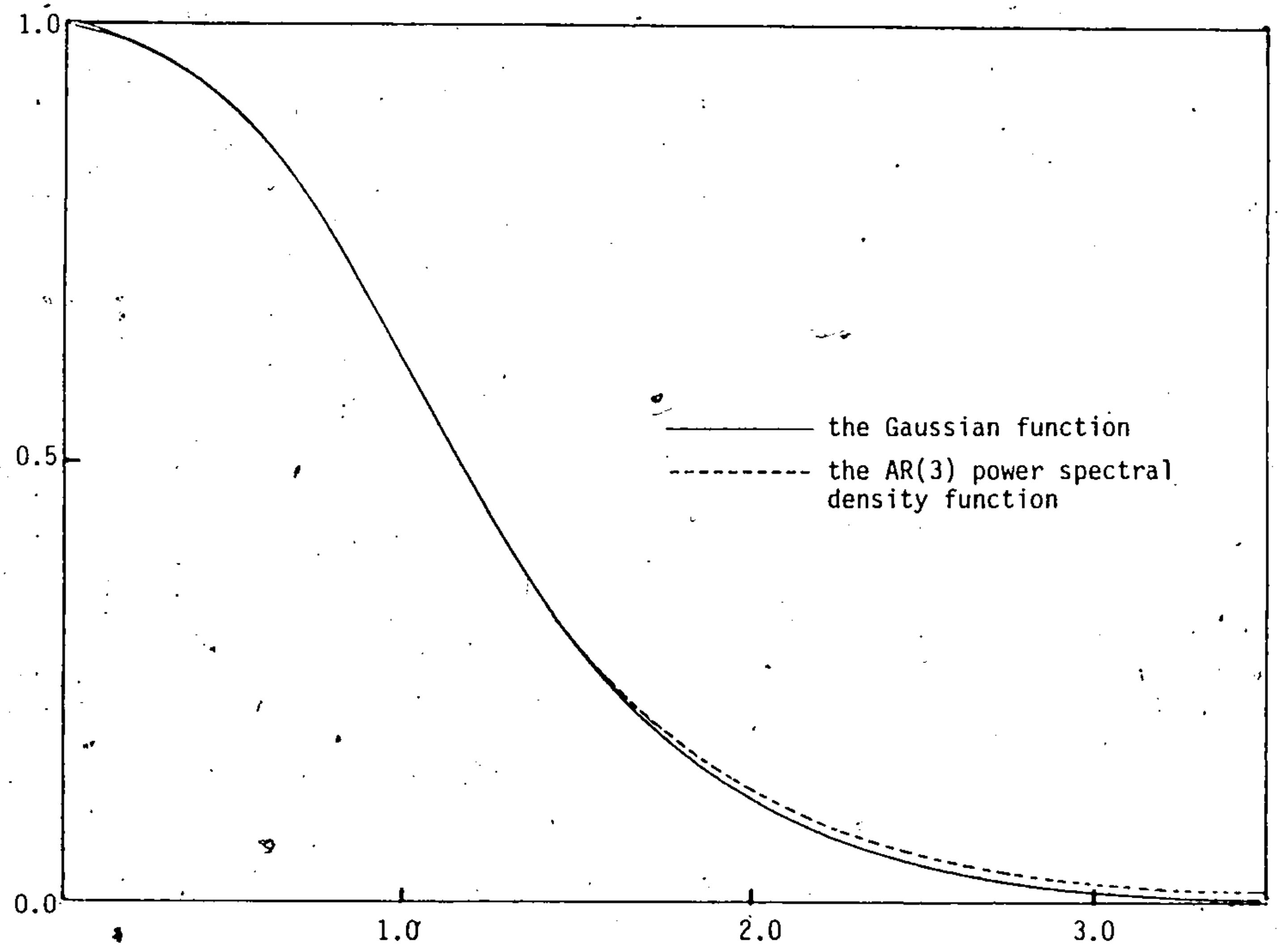


Fig. 4.5 The fit of an AR(3) power spectral density function to a Gaussian function

average (ARMA) process. Since an ARMA process may, in general, be approximated as a higher order AR process (Box and Jenkins [6]), it is concluded that under Hypothesis H_2, the received data may be modelled as a higher (with respect to the data under Hypothesis H_1) order AR process. Hence a fourth or a fifth order linear prediction will be acceptable for Hypothesis H_2

For the implementation of this second model the Kalman tapped delay-line prediction error filter (Figure 4.3) is utilised. In this application of Kalman theory (Kalman [35]) a random walk state model is used to represent nonstationarity (Zhang and Haykin [69]). As dictated by the above model, the prediction error structure is implemented with a first-order tapped delay-line Kalman filter, and the prediction error structure for Hypothesis H_2 is implemented with a fifth order tapped delay-line Kalman filter.

The recursive equations for the Kalman prediction error filter are presented in Appendix 3. This implementation will be referred to as the IBDA (KALMAN) (Figure 4.6).

4.3.2 The Estimation of the a priori Prediction Error Powers

As the radar environment is spatially correlated,

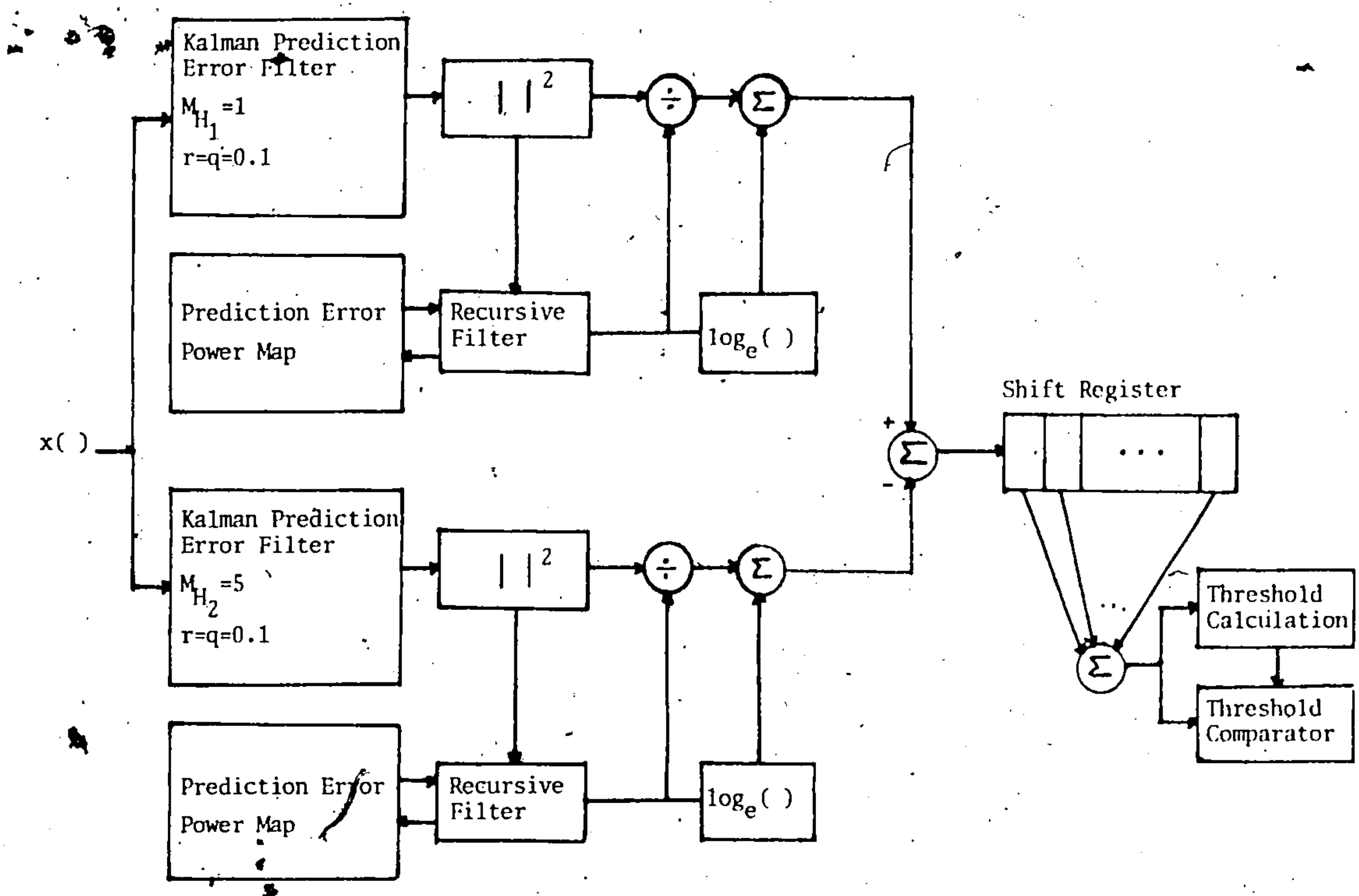


Fig. 4.6 The IBDA(KALMAN) implementation

information concerning any particular range-azimuth cell[1] is contained in surrounding cells. This additional information is used to estimate the a priori prediction error powers $\sigma^2(\ell|H_1)$, $\sigma^2(\ell|H_2)$ (Eqn. 4.10) and the CFAR threshold $\lambda(\ell)$ (Eqn. 4.12). Specifically, the a priori prediction error powers are estimated along lines of constant range, while the CFAR thresholds are estimated along lines of constant azimuth.

The estimates of the a priori prediction error powers are calculated using a prediction error power map. Every resolution cell of interest within the radar environment is represented by a unique map element. Each map element contains two entries which are the time-averaged (on a scan-to-scan basis) outputs $|\tilde{x}(\ell|H_1)|^2$ and $|\tilde{x}(\ell|H_2)|^2$ of the two prediction error filters. The a priori prediction error powers $\sigma^2(\ell|H_1)$ and $\sigma^2(\ell|H_2)$ used in the calculation of the incremental update term (Eqn. 4.11) are estimated by averaging the map element contents for the 26 cells on either side (in azimuth) of the ℓ-th map element.

This time-averaging is performed by recursively updating, on a scan-to-scan basis the contents of the map elements. The update scheme is to add 7/8 of the present contents of the ℓ-th map element to 1/8 of the outputs

[1]Arange-azimuth cell is an area defined by the antenna beamwidth and the range resolution.

$|\tilde{x}(\ell|H_1)|^2$ and $|\tilde{x}(\ell|H_2')|^2$ generated by the prediction error filters for the ℓ-th resolution cell.

The map is initialised by loading it with the outputs from the prediction error filters during the first scan.

The updating scheme described above for the implementation of the IBDA differs from that used for the MTD algorithms in that the latter do not average in azimuth (Drury [10], Karpand Anderson [37]).

4.3.3 The Estimation of the CFAR Threshold

The estimate of the CFAR threshold is obtained from the LLRs generated for the 13 resolution cells on either side (in range) of the cell of interest.

In order to obtain a stable estimate of the correct threshold for the background clutter, the two immediately adjacent resolution cells are exempted from the calculation. This is to eliminate the problem of range-cell splitting where a target is present in two adjacent range cells. A calculation performed including the adjacent cell would unnecessarily raise the threshold value, resulting in decreased detection performance.

A simplistic method of estimating the adaptive threshold is to employ the mean of the LLRs for these range cells. This approach will reduce detection performance due to the possibility of spikes of ground clutter and/or other targets existing in these range-cells, with a subsequent increase in the estimate of the threshold.

This problem is avoided to a certain extent by searching for the largest values of the LLRs from these range azimuth cells and excluding them from the calculation of the mean value. However, the use of this method of estimating the adaptive threshold did not yield satisfactory performance when implemented with the IBOA.

The CFAR threshold for the IBDA implementations is calculated using the median estimate of the LLRs from the range-cells (Ataman et al. [3]). The median estimate of a sample set is the middle sample of the order-ranked set. As such, the median estimate is extremely insensitive (as compared to the mean estimate) to what may be termed outliers. Outliers, in this context, would be the spikes of ground clutter and other target echos.

The CFAR threshold $\lambda(\ell)$ (Eqn. 4.12) is obtained by scaling the median estimate of the LLRs from the surrounding range cells by some fixed constant. The scaling constant is fixed so as to yield the desired probability of false-alarm.

4.4 The Moving Target Detection Algorithms

In most of the earliest attempts to solve the radar target detection problem, moving target indicators (MTI's) were used (Skolnik [60]). These took the form of two-pulse or three-pulse cancellers which are essentially linear filters with a notch at zero frequency so as to eliminate ground clutter. However, they are not able to discriminate against moving weather clutter.

The first generation of moving target detectors (MTD's), the MTD-1 (Drury [10]), is a first approach to a solution of this difficulty. It operates by passing blocks of 10 consecutive returns (called a coherent processing interval (CPI)) from each range ring through a three-pulse cancellor followed by an 8-point Fast Fourier Transform (FFT) (Figure 4.7). A zero velocity filter (ZVF) is also used to increase the probability of detecting targets possessing only a small radial velocity component by forming an average of the raw video data in the CPI.

CFAR performance is maintained through the use of two different methods. The output from the bank of non-zero doppler filters are each separately thresholded. These thresholds are obtained by averaging the outputs from the appropriate filters in CPI's up to 1/2 mile in range on either side of the CPI of interest. As men-

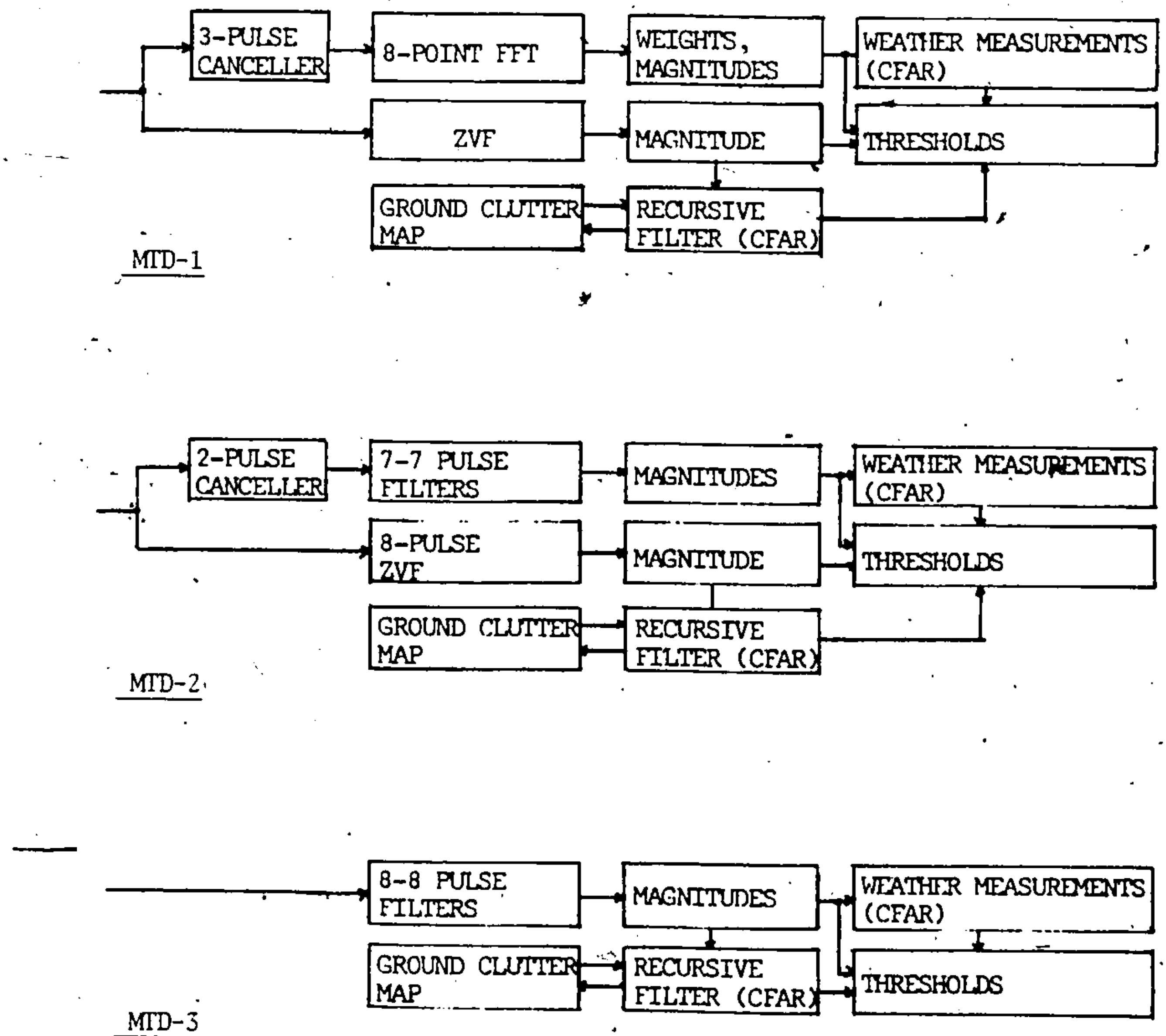


Fig. 4.7 The MTD systems

tioned in section 4.3.3, the largest values in these sets of outputs from the FFT's are excluded from the calculations. This method establishes a stable estimate of any weather clutter which may exist in this CPI. The output from the ZVF is compared with a threshold obtained from a ground clutter map. The ground clutter map contains a stable estimate of the ground clutter in each CPI in every range ring of interest. This estimate is obtained by recursively filtering the outputs of the ZVF for each of the CPIs on a scan to scan basis. This long time constant is possible because of the slow temporal variation of the ground clutter.

The MTD-2 (Karp and Anderson [37]) (Figure 4.7) design philosophy results from a recognition of the limitations of the FFT algorithm when used as a doppler processor. The FFT may be considered to be a bank of matched filters, matched to sinusoidal signals. Hence, the FFT will produce acceptable detection performance only if the background noise corrupting the assumed sinusoidal target returns is white and Gaussian. However, ground clutter and weather clutter are highly correlated processes, and ground clutter is often highly non-Gaussian.

As a result, the non-zero doppler filters for the MTD-2 are designed using the method of deLong and Hofstetter

[8]. This design approach calculates the optimal linear filter for the detection of a specified deterministic signal in additive correlated Gaussian noise.

For a CPI of length (N-1) samples, the tap weights $\underline{h}(N-1)$ for a specific non-zero doppler filter are the solution to the matrix equation (deLong and Hofstetter [8]):

$$\underline{h}(N-1) = [\underline{R}_x^{-1}(N-1)\ \underline{S}(N-1)]^T \tag{4.15}$$

where $\underline{R}_x(N-1)$ is the (N-1)x(N-1) covariance matrix of the assumed Gaussian clutter process plus the WGN process, and $\underline{S}(N-1)$ is the specific target response to which the non-zero doppler filter is to be matched. Hence, the output y of this filter may be written as:

$$y = \underline{h}^T(N-1)\underline{x}(N-1)$$

$$= [\underline{R}_x^{-1}(N-1)\underline{S}(N-1)]^T\underline{x}(N-1) \tag{4.16}$$

The real version of this generalised matched filter (Eqn. 4.16) is discussed in section 3.3.3 where its equivalence to the LLR, generated by an application of the IBDA to the detection of a deterministic signal in coloured Gaussian noise, is established.

The covariance matrix $\underline{R}_x$ is generated by assuming that the ground clutter power spectrum is Gaussian in shape, centered at zero doppler, with a width determined by the rate of rotation of the antenna. The weather clutter power spectrum is modelled as being uniform across the doppler band except for a rectangular notch centred at the doppler frequency of interest. This specific weather clutter model is chosen so as to reduce the filtered sidelobes which enhances the filters' weather rejection capability, resulting in improved target detection performance.

The received target signal is modelled as a sinusoid of the chosen doppler frequency modulated by the Gaussian-shaped antenna pattern.

In terms of the implementation of the MTD-2, (Karp and Anderson [37]), it was determined that the filters' tap weights word size need only be 3 or 4 bits if the doppler filters are preceded by a two-pulse cancellor. In addition, greater resolution could be achieved if the CPI is reduced from the 10 samples (as in the MTD-1) to 8 samples. As a result, each of the filters in the seven non-zero doppler filter bank is described by ~~seven 4-bit~~ complex coefficients.

The ZVF for the MTD-2 is also implemented as a linear FIR filter (Karp and Anderson [37]). The ZVF is designed so as to maximise the gain in that portion of the doppler spectrum centred about zero doppler. Additional constraints are placed on the filter coefficients in order to reduce the sidelobes. The ZVF is described using eight 4-bit tap weights.

CFAR processing is accomplished in much the same manner as is used in the MTD-1.

The MTD-3 is the third generation in the evolution of the MTD algorithm (Taylor [62]). It was recognised that by increasing the filters' tap weights to 10-bit accuracy, the filter sidelobes are reduced as compared to the MTD-2. In addition, the two-pulse cancellor is no longer necessary. Hence the MTD-3 consists of a bank of eight filters, each of which is described by eight 10-bit complex tap weights (Figure 4.7).

The filters employed by the MTD-3 are again calculated using the method of deLong and Hofstetter [8].

CFAR processing is performed in a similar way as is used in both the MTD-1 and the MTD-2.

The precise details of the implementation of the MTD-1 and the MTD-2 were made available to the author for this research by M.I.T. Lincoln Laboratory. Details for the implementation of the MTD-3 were made available by Westinghouse.

The MTD-1 and the MTD-2 have both been successfully tested and implemented by Lincoln Laboratory (Drury [10], Karp and Anderson [37]). The MTD-3 is part of a design proposed by Westinghouse for a national air traffic control system.

4.5 Summary

In this chapter the IBDA is applied to the radar target detection problem. Two different models of this detection problem are discussed in section 4.3. The first model utilises the experimentally observed difference in the widths of the power spectra of the received data under the two hypotheses. Since the power spectra under either of these two hypotheses are Gaussian shaped, and hence, may be modelled in terms of AR processes, then linear prediction may be utilised. The LSL algorithm based on the exact method of least squares is employed in the calculation of the prediction errors. This implementation of the IBDA is referred to as the IBDA (LSL).

The second implementation is referred to as the IBDA (KALMAN). Again, assuming no target present then the data process may be modelled as an AR process. However, if a target is present, then the data process may now be modelled as an ARMA process, which, in turn, may be modelled as a higher order AR process then is used when assuming no target is present. This implementation employs Kalman prediction error filters.

Neither of these two implementations explicitly assumes that the clutter process is Gaussian. However, the prediction errors in both implementations are generated through the use of linear prediction algorithms. Since linear prediction is optimal only for Gaussian processes, these two implementations will suffer some degradation in performance if the clutter is non-Gaussian.

The MTD algorithms are discussed in section 4.4. Basically, these algorithms are designed to **maximise the** target signal power to the clutter plus noise power ratio, under the assumption that the combination of the clutter and noise is modelled as a correlated Gaussian process.

The IBDA (LSL) and the IBDA (KALMAN) may be compared with the MTD algorithms by first noting that, as is shown in section 3.3.3 the LLR resulting from an application of the IBDA to the detection of a deterministic signal in Gaussian noise is equivalent to the real versions of the MTD detection statistics. Although these comparisons are made in the real domain, they may be easily extended to the complex domain.

Second, the IBDA is implemented using adaptive prediction error filters whereas the MTD algorithms are implemented using fixed FIR filters. Third, the IBDA implementations output a decision for every range-azimuth cell of interest, while the MTD algorithms make decisions based upon the processing of contiguous blocks of data, the coherent processing intervals.

Hence it can be concluded that the MTD algorithms are limiting forms of the IBDA implementations. It should therefore be expected to realise a better detection performance using the IBDA than would be possible with the MTD approach. This is indeed confirmed in the next chapter by using real radar data.

CHAPTER 5

SYSTEMS PERFORMANCE EVALUATIONS AND COMPARISONS

In Chapter 3, the likelihood ratio for the detection of a discrete-time stochastic process in additive WGN is derived. In Chapter 4 the application and implementation of the IBDA to radar target detection is described. The MTD algorithms are discussed in section 4.4.

In this chapter, the performance of these detection algorithms is analysed in terms of their receiver operating characteristics (ROC) (Metford and Haykin [45]. For this application, the ROCs (Van Trees [63]) are described in terms of the probability of false alarm (P_{FA}) as a function of the target signal power to clutter plus receiver noise power ratio for a given probability of detection (P_D). Different ROC curves will result for different target dopplers.

All of the detection algorithms are applied to the same sets of radar data. The performance comparisons are thus internally consistent yielding strong and useful conclusions and recommendations.

In section 5.1, the data acquisition system is described. The methods by which the data is analysed so as to obtain the ROC curves and to obtain the parameters necessary for implementation of the IBDA are discussed in section 5.2. The ROC curves are analysed and performance comparisons are made in section 5.3. The conclusions are stated in section 5.4. Suggestions for future research are made in section 5.5.

5.1 The Data Acquisition System

The radar data for the systems performance analysis are required to satisfy several criteria. Since all of the detection algorithms operate on the complex baseband signal (Whalen [66]), a fully coherent radar system is needed to allow the recording of both the inphase (I) and the quadrature (Q) data channels. As the IBDA requires uniformly sampled data (for ease of implementation) a fixed transmitter pulse repetition frequency (PRF) is mandatory. Following the recommendations of Gibson [16] and Taylor [62]), the sampled video channel must be digitised using a minimum of 10 bits. In addition, since weather clutter-dominated radar data is needed, a linearly polarized transmitted signal is desirable [46].

These requirements are all met by the Westinghouse TRACS radar system installed at the Canadian Forces Base CFB Trenton. The significant operating characteristics of the TRACS are summarised in Table 5.1. Although the TRACS is a fully operational radar system, the air traffic controllers and the technical support group cooperated fully in the reconfiguration of the TRACS to our requirements.

Referring to Figure 5.1, the two channels of 10-bit data, along with the zero range mark, the antenna azimuth change pulse and the antenna azimuth reference pulse comprise 23 parallel digital channels operating at a per channel bit rate of 1.3 MHz. This data stream was recorded on the CRL's digital recorder. The digital recorder consists of an AMPEX HBR-3000 digital processing bay in conjunction with an AMPEX AR-1700 Wideband Airborne/Mobile Tape Transport Unit.

Recordings were initiated when a significant MTI breakthrough occurred on the TRACS plan position indicator (PPI). Two 25 minute tapes were recorded during a medium rate snowfall. A third tape was recorded for comparative purposes after the snowfall had ceased. Photographs of the TRACS PPI taken during these recordings are shown in Figure 5.2.

Several data sets of varying amounts of ground and/or weather clutter are needed for processing by the detection algorithms for performance evaluation. To

Table 5.1 The TRACS Operating Characteristics

Transmitter frequency = 1.23 GHz

Pulse repetition rate = 657.2 Hz

Pulse width = 0.5×10^{-6} sec

Antenna revolution rate = 12.5 rev/min (nominal)

Antenna beam width = 1.35° (nominal)

Sample rate = 1.35 MHz

Sample word size = 10 bits, both inphase and quadrature channels

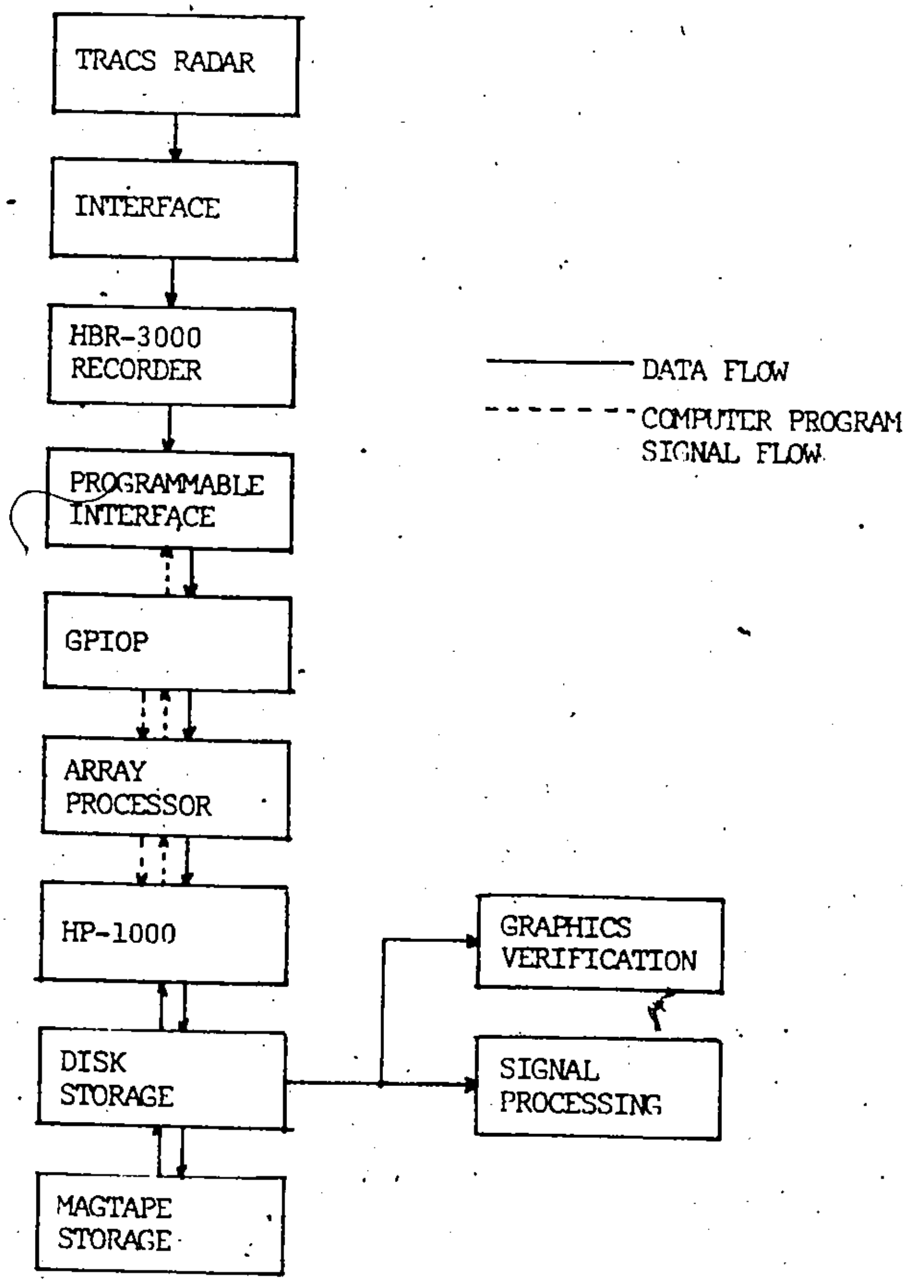
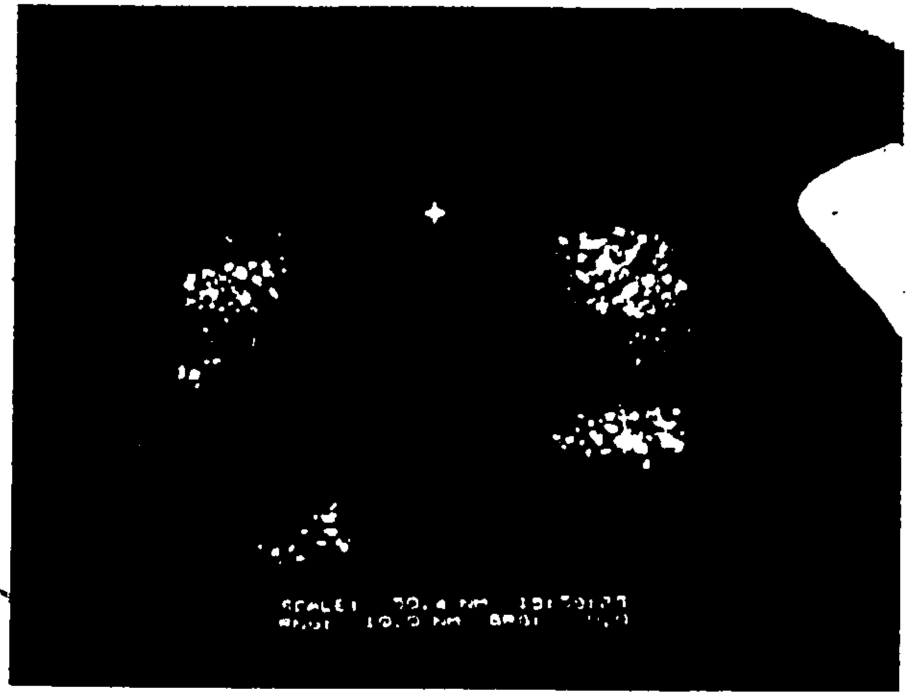
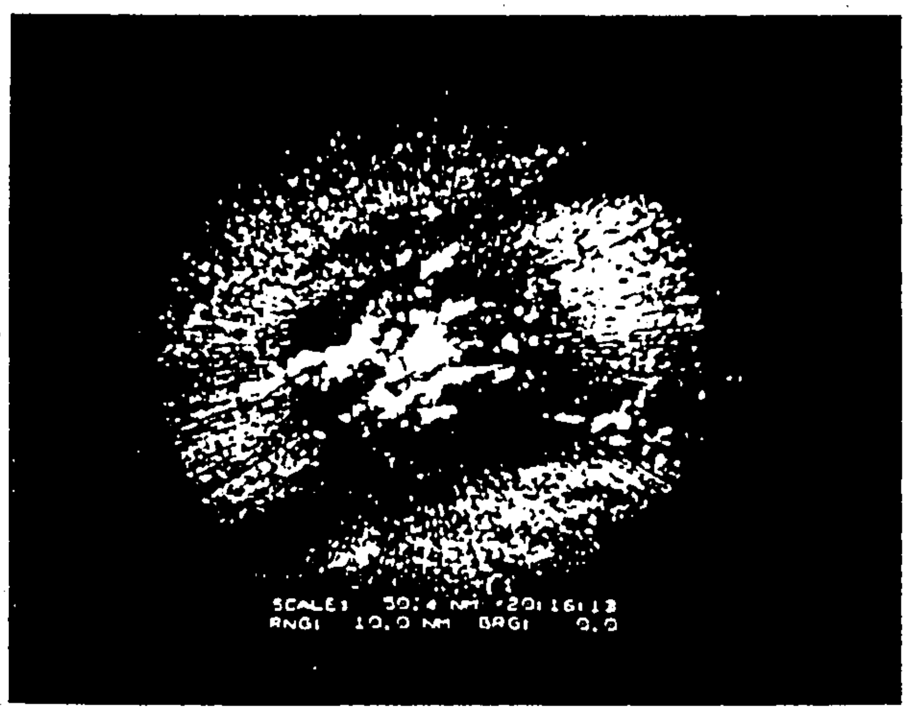


Fig. 5.1 A flowchart of the data acquisition and storage system



Photograph of the MTI Processed I Channel



Photograph of the Log Processed Video ($\sqrt{I^2+Q^2}$)

Fig. 5.2 Photographs of the TRACS PPI display conform with the specifications for implementation of these algorithms, each of the data sets consists of at least 12 consecutive scans of a 2 nautical mile wide swath of radar returns. This large amount of data (approximately 3.1 mega-bytes) necessitates a complex interface between the recorder and the HP-1000 computer. The high recorded data bit rate is not a significant factor since the recorder may be slowed down when it is in the reproduce mode.

A specific data set is acquired by setting the registers within the programmable interface so as to select the swath of interest. The data is then pased through the programmable interface into the General Programmable Input/Output Processor (GPIOP). By using the Array Processor (AP) as a rather expensive buffer memory, the GPIOP writes the data into the AP memory as the HP asynchronously reads data out of AP memory onto disc. This process is continued until at least 12 scans of radar data have been acquired by the HP. This data is then stored on computer magtape for ease of future reference.

Two data sets were acquired in this manner from the first tape. The first data set is positioned in range from 12.75 to 14.75 nmi. so as to contain the weather clutter cells evident in the photographs of the PPI display (Figure 5.2). A plot of the amplitude envelope $\sqrt{I^2+Q^2}$ of the first scan of this data set is shown in Figure 5.3. This data set is characterised as being weather clutter dominated radar data. The second data set is positioned in range from 1.25 to 3.25 nmi. and is characterised as being ground clutter dominated radar data. A plot of the amplitude envelope of the first scan of this data set is shown in Figure 5.4.

5.2 The Systems Performance Evaluation Methodology

Initialisation of the ground clutter maps for the MTD algorithms and the prediction error power maps for the IBDA algorithms achieved by the processing of the first 8 scans of a given data set.

The last four scans are processed for P_{FA} calculations. Since the MTD-1 operates on CPI's of length 10 samples, the MTD-2 and the MTD-3 operate on CPI's of length 8 samples, while the IBDA processes sample by sample, blocks of 80 consecutive sweeps were accessed at a time. Since each scan contains 3296 samples per range ring, 16 samples are left over. These remaining sweeps are ignored by the MTD algorithms. However, as the IBDA requires continuity of data, these remaining samples are

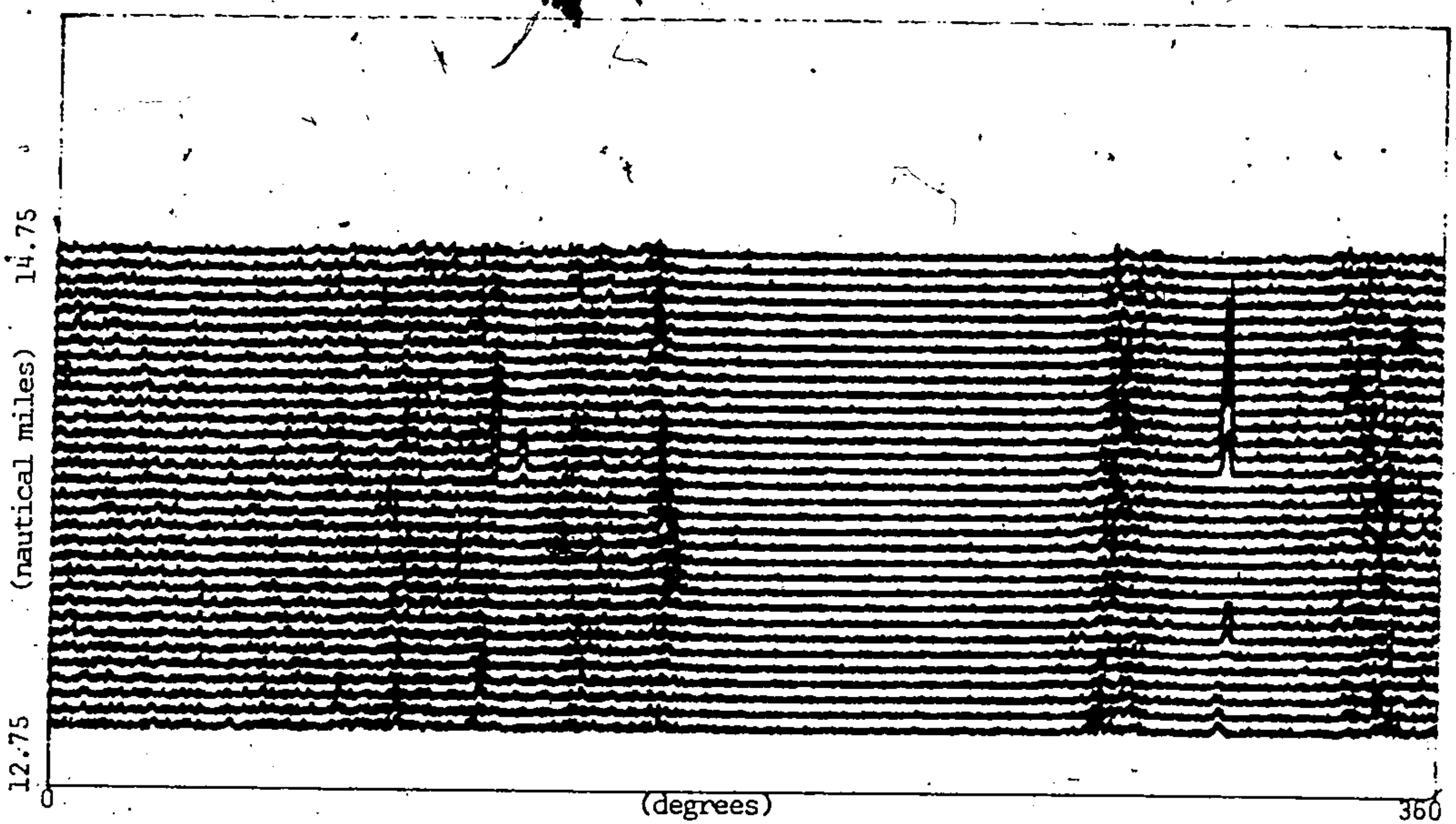


Fig. 5.3 The amplitude envelope of the first scan of the weather clutter dominated radar data

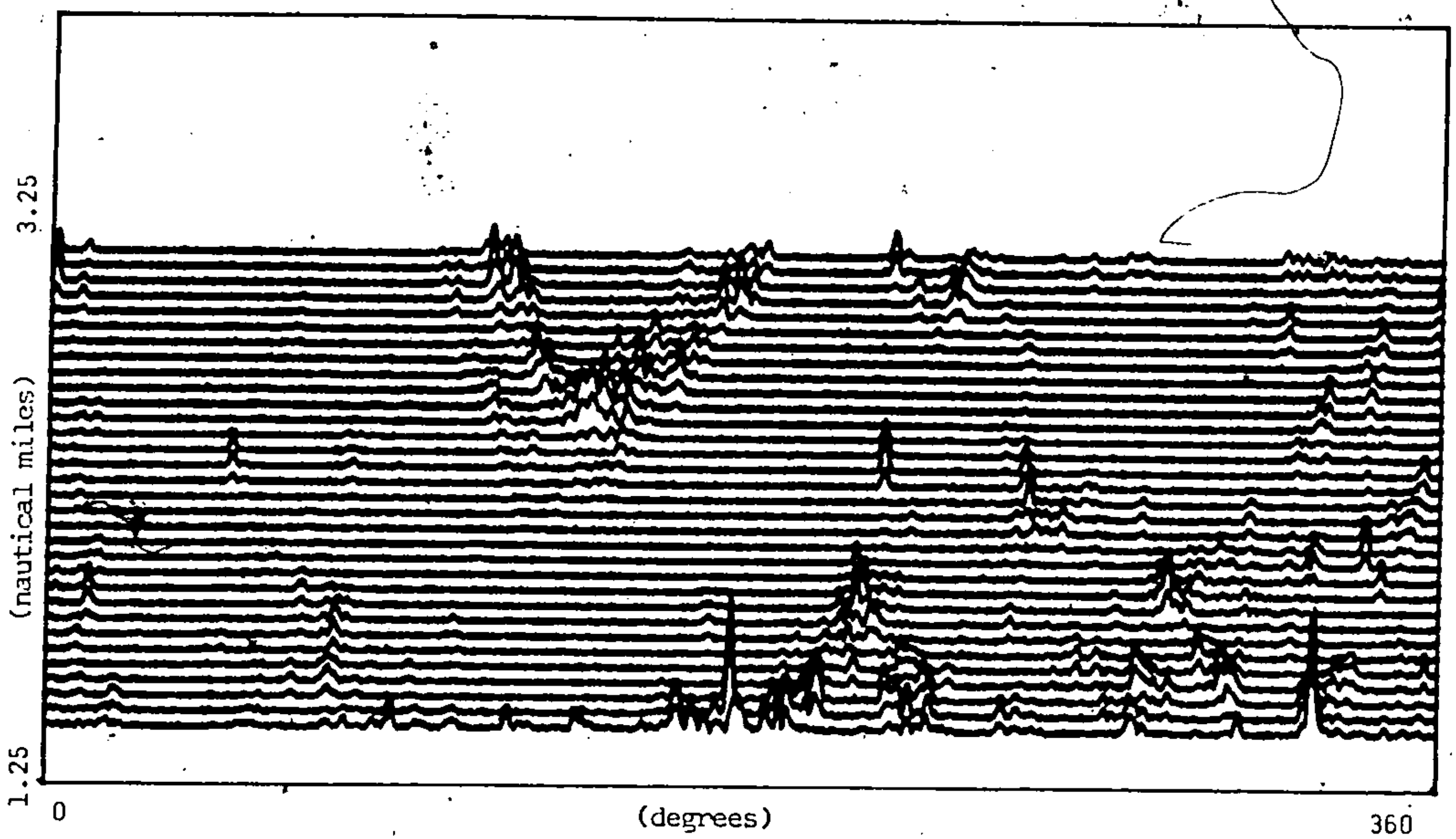


Fig. 5.4 The amplitude envelope of the first scan of the ground clutter dominated radar data

processed by the IBDA but do not contribute to performance calculations.

The processing of these four scans results in 10496 decisions by the MTD-I algorithm and 13120 decisions by the remaining algorithms. Hence, the minimum achievable P_{FA} is approximately 10^{-4}. Although ASR systems usually operate at a P_{FA} of 10^{-5} or less, computer processing-time constraints do not permit this.

The ninth scan is processed for P_D calculations. Since it is necessary to examine the performance of the algorithms for all SCNR's and target dopplers, artificial target responses were constructed. The target response is of the form:

$$s(nT) = \exp[-(n-n_o)(\alpha T/2\sigma)^2]\exp[2\pi j\tilde{f}(n-n_o)T]$$

where the antenna rotation rate α=1.254 rads/sec, the interpulse period T=0.00152 sec, the variance of the modelled Gaussian-shaped two-way voltage pattern σ^2= 0.0000741 and the doppler frequency f is in Hertz.

Using the above-mentioned blocks of 80 consecutive sweeps, the target response is added to the radar data by scaling the target response to the amount of clutter power so as to obtain the desired SCNR.

As discussed in section 4.3.1, various parameters which are needed for the IBDA implementations must be determined empirically. The parameters required for the IBDA (LSL) are the filters order M and the two damping constants (μ_{H_1}, μ_{H_2}) (see Appendix 1). The parameters required for the IBDA (KALMAN) implementation are the filter orders (M_{H_1}, M_{H_2}), the variance r of the innovations process and the variance q of the random walk state model used in the Kalman algorithm (see Appendix 2).

Ideally the optimum set of parameters for each of the two implementations would be determined through comparisons of the ROCs for all possible sets of parameters. However, due to computer processing-time constraints, this is impossible. Instead, performance trends are observed by taking a specific section of radar data and looking for the largest relative difference in the LLRs as a specific target is or is not placed into the data.

In preliminary studies it was observed that if for a specific set of parameters, good detection performance is obtained in weather clutter dominated radar data, then, in general, good detection performance will also be achieved in ground clutter dominated radar data. Hence, the parameter search was performed using weather clutter dominated radar data.

Some plots of the LLRs generated by the IBDA (LSL)

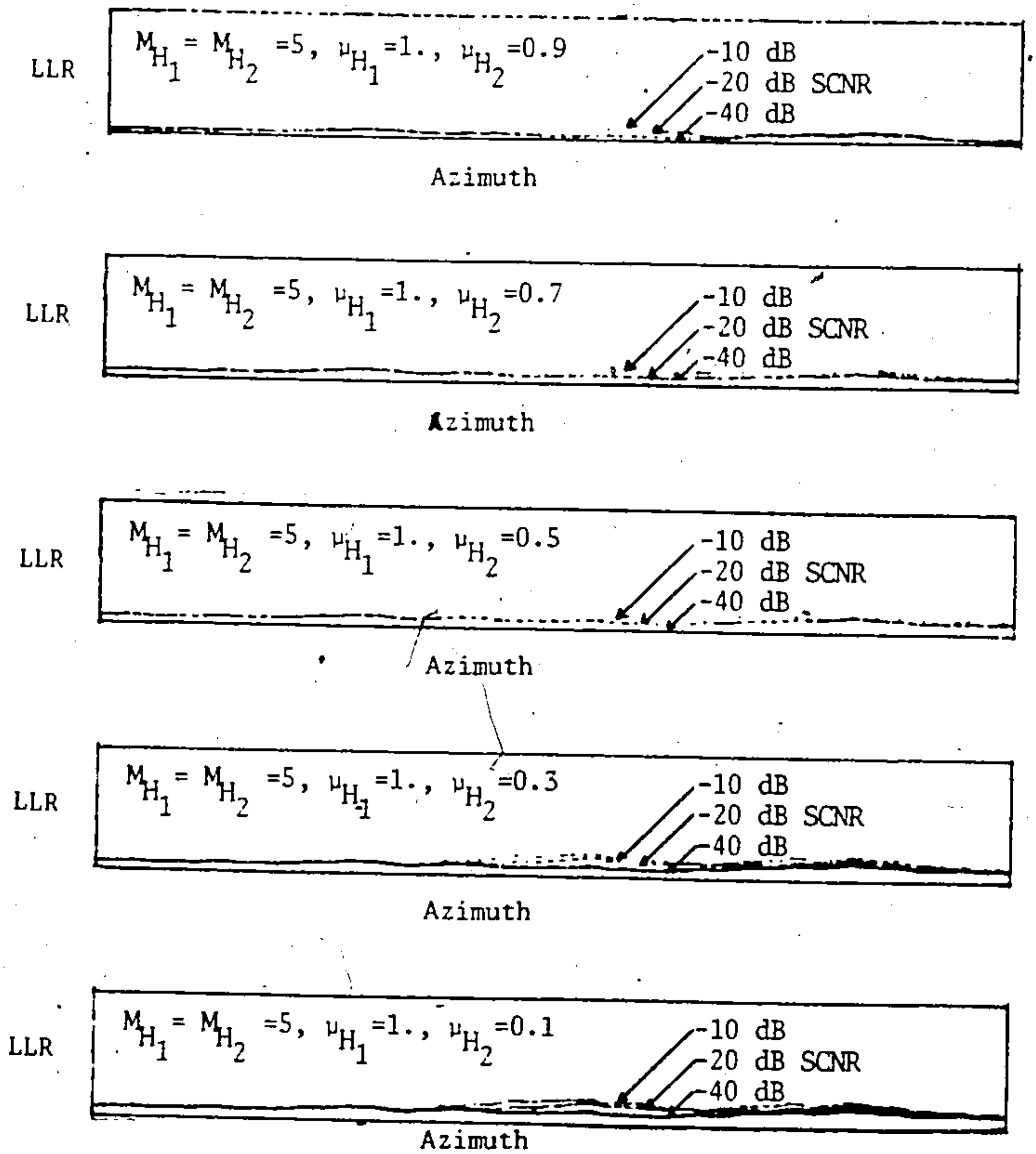


Fig. 5.5 A sample set of the LLRs generated by the IBDA(LSL), (target doppler = 50 Hz)

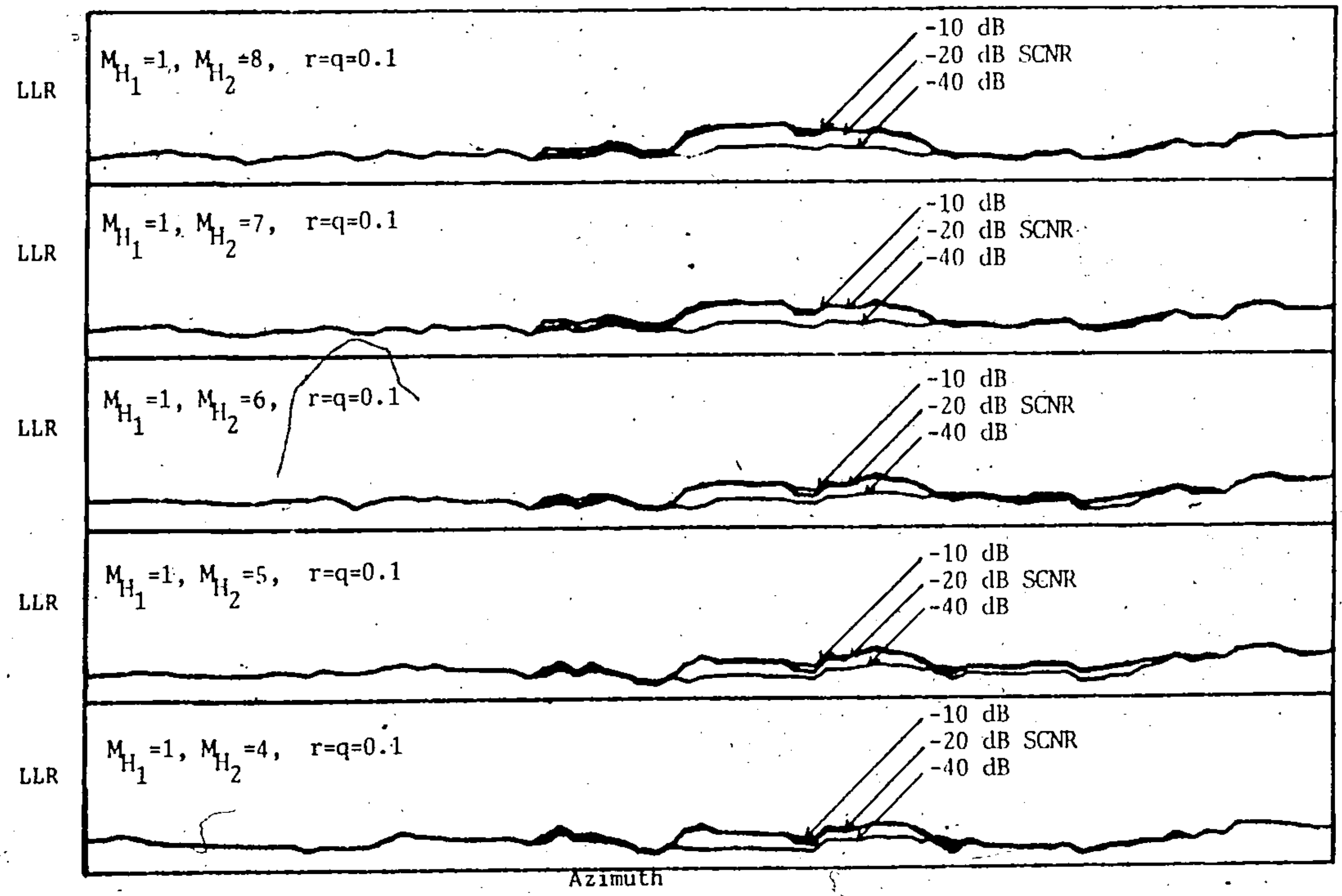


Fig. 5.6 A sample set of the LLRs generated by the IBDA(KALMAN), (target doppler = 50 Hz)

are presented in Figure 5.5. LLRs generated by the IBDA (KALMAN) are presented in Figure 5.6.

5.3 The Systems'Performance Comparisons

The performances of the five detection algorithm, the MTD-1, the MTD-2, the MTD-3, the IBDA (LSL) and the IBDA (KALMAN) are examined in this section.

The five sets of ROC curves resulting from the processing of weather clutter dominated radar data are presented in Figure 5.7, 5.8, 5.9, 5.10, 5.11. The ROC curves from the processing of ground clutter dominated radar data are presented in Figure 5.12, 5.13, 5.14, 5.15, 5.16.

The problem presented by the lack of an SCNR calibration is resolved by choosing the MTD-1 algorithm as a reference. Within this frame of reference, performance comparisons of the other detection algorithms are made. Internal consistentcy is thereby maintained. The MTD-1 algorithm is chosen as the reference since the MTD-2 and the MTD-3 algorithms are referred to as being improvements upon the MTD-1 (Taylor [62]), and also since the two IBDA algorithms are to be compared with the MTD algorithms.

The performance comparisons for weather clutter dominated radar data are presented in Figure 5.17. The performance comparisons in ground clutter dominated radar data are presented in Figure 5.18.

The overall improvement of the performance of the

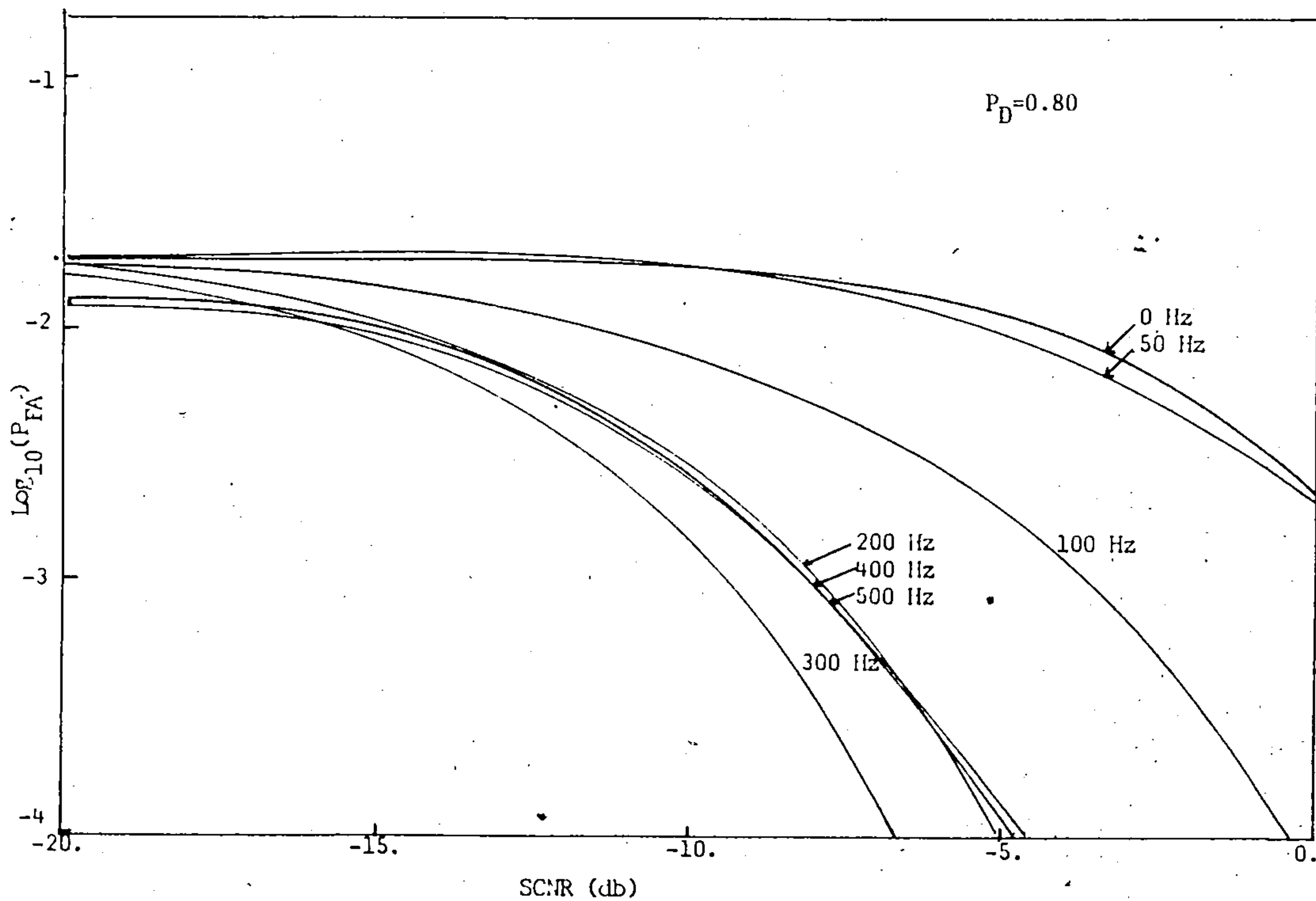


Fig. 5.7 The ROC for the MTD-1 (weather clutter dominated radar data)

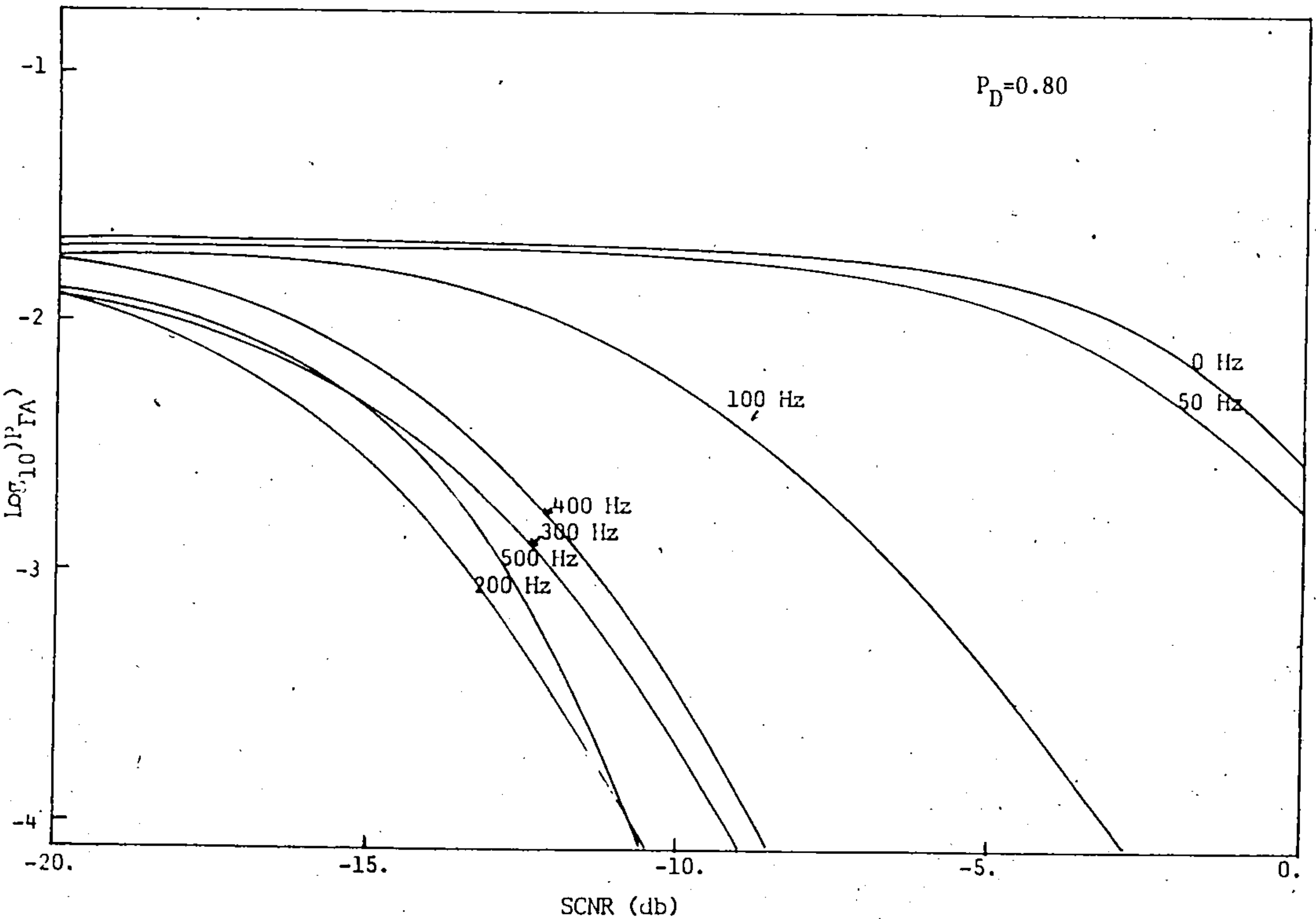


Fig. 5.8 The ROC for the MTD-2 (weather clutter dominated radar data)

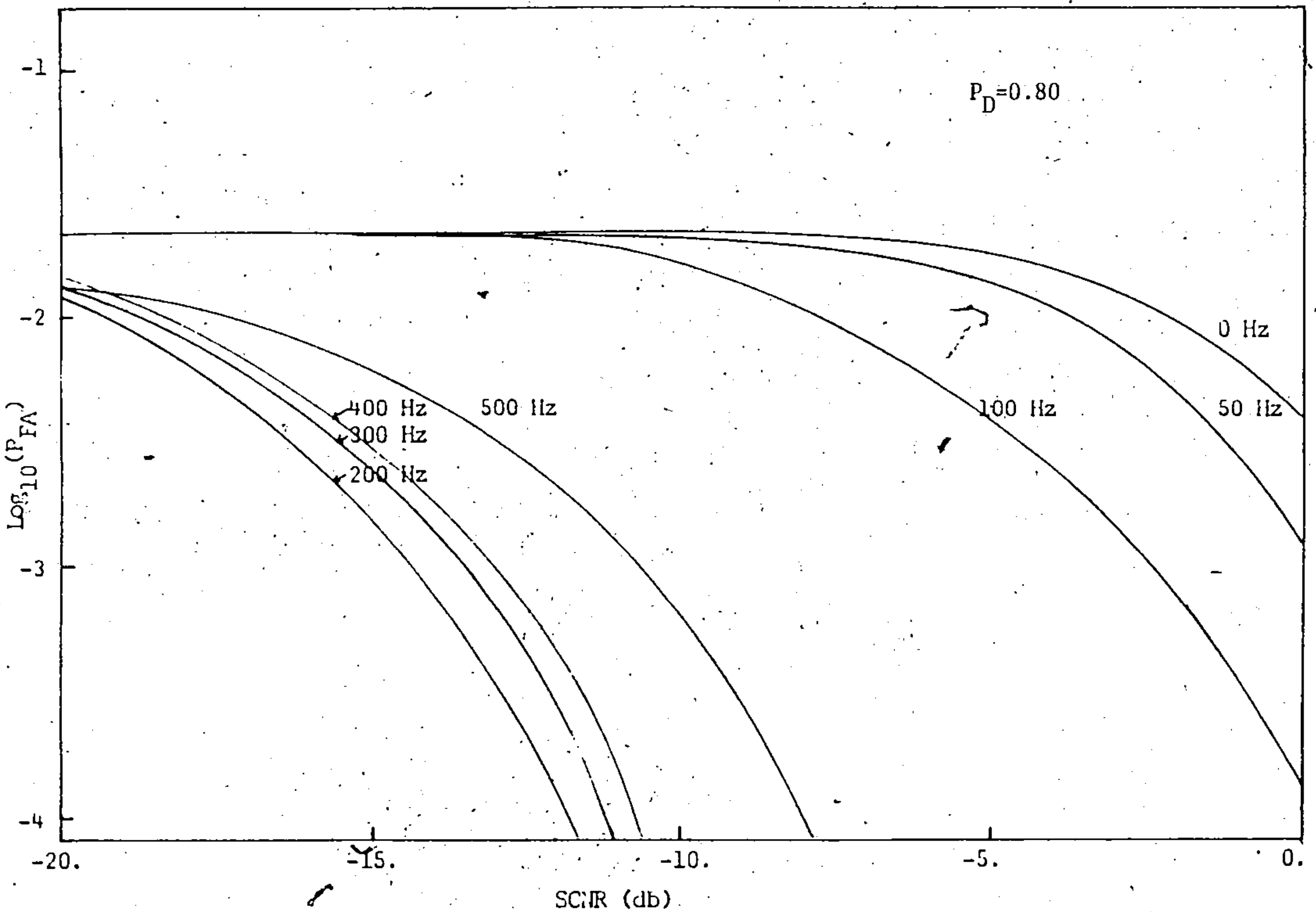


Fig. 5.9 The ROC for the MTD-3 (weather clutter dominated radar data)

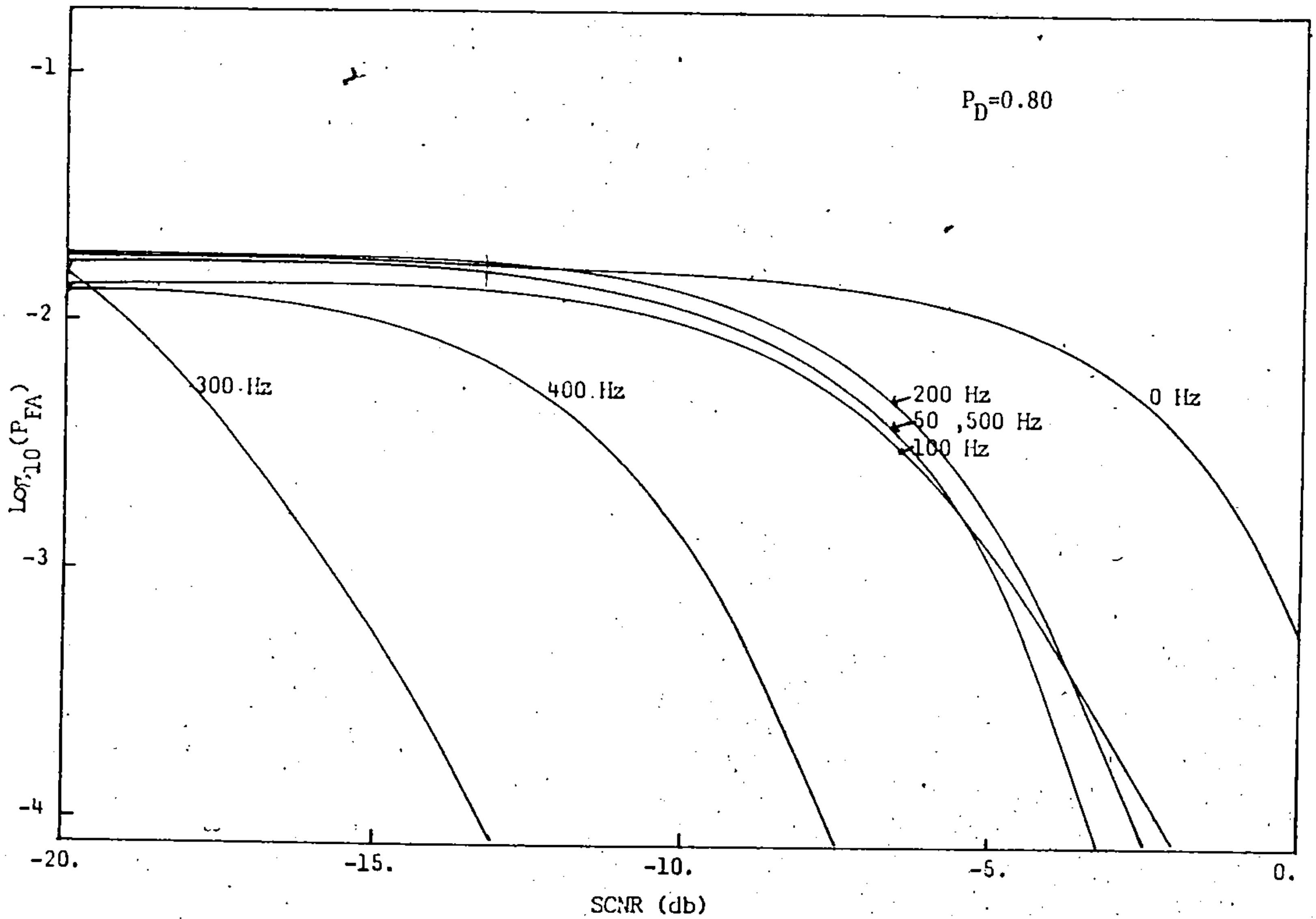


Fig. 5.10 The ROC for the IBDA(LSL) (weather clutter dominated radar data)

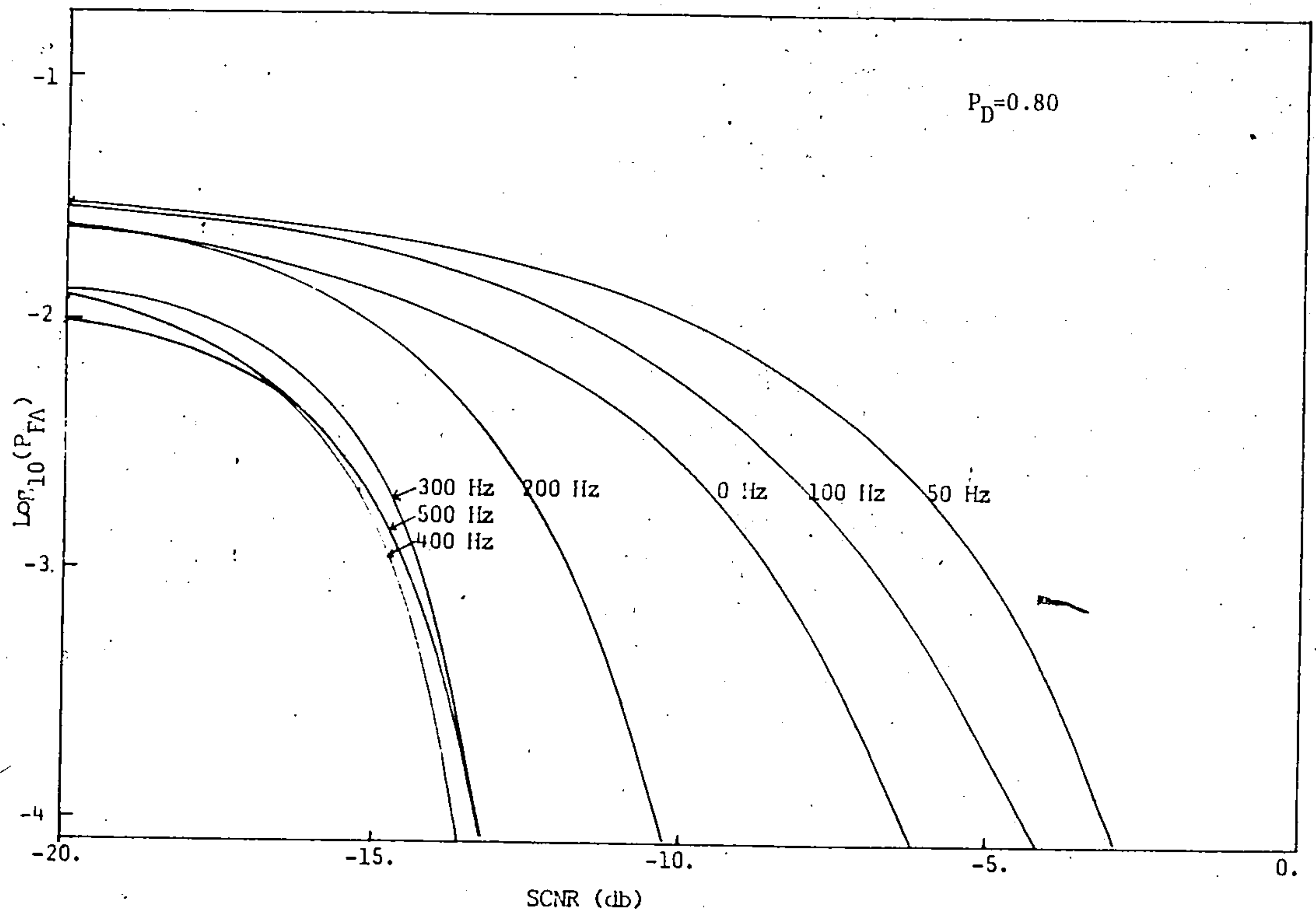


Fig. 5.11 The ROC for the IBDA(KALMAN) (weather clutter dominated radar data)

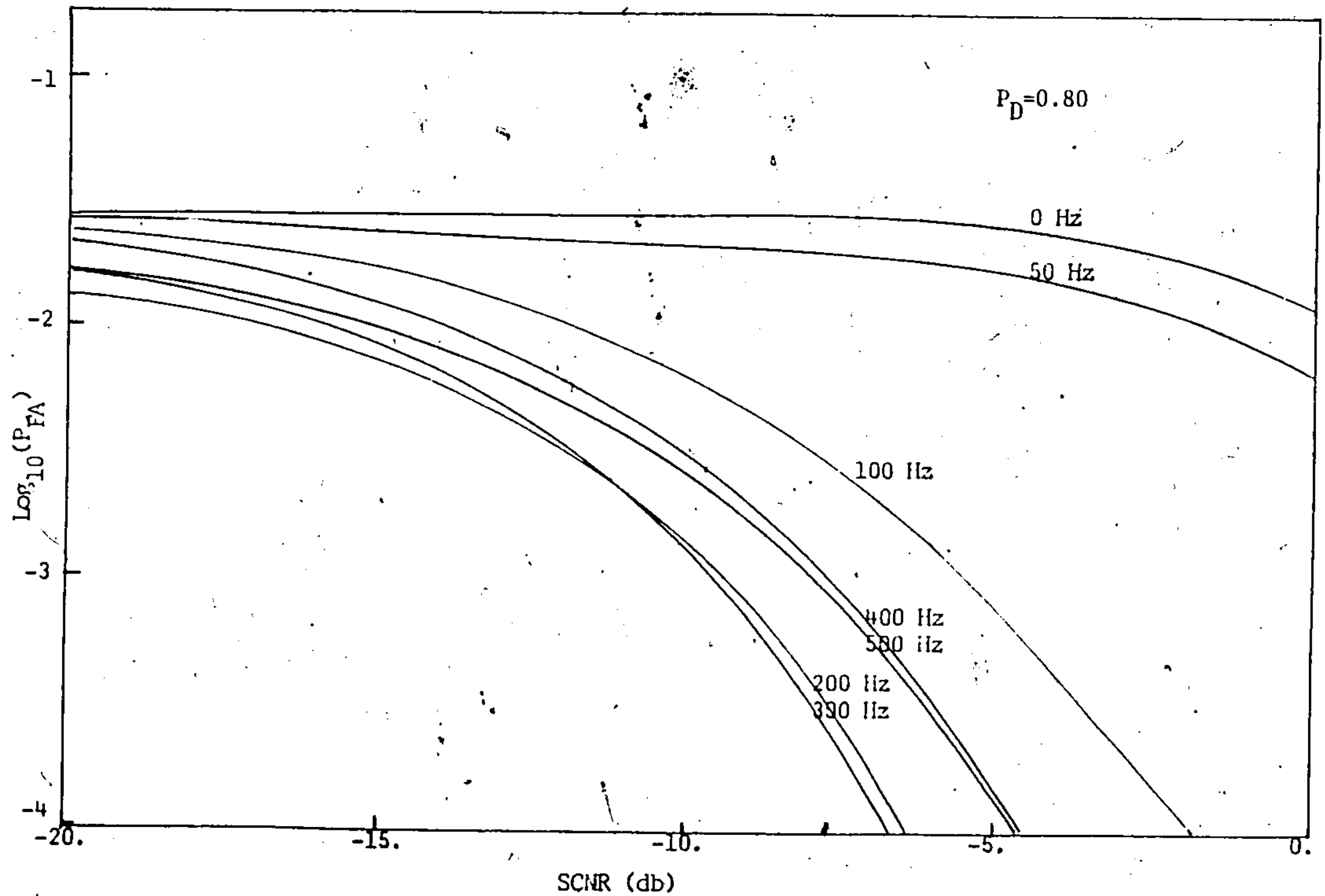


Fig. 5.12 The ROC for the MTD-1 (ground clutter dominated radar data)

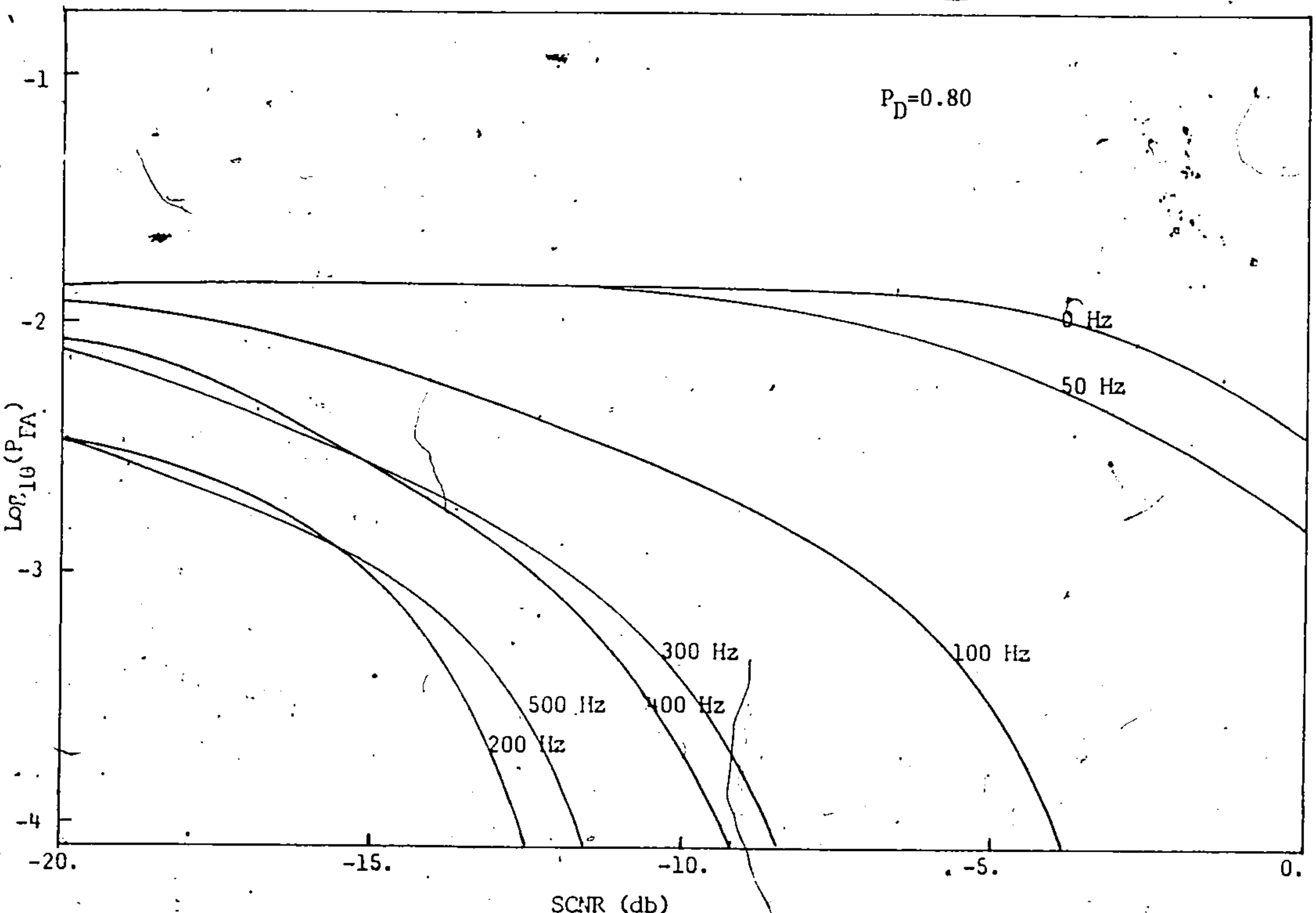


Fig 5.13 The ROC for the MTD-2 (ground clutter dominated radar data)

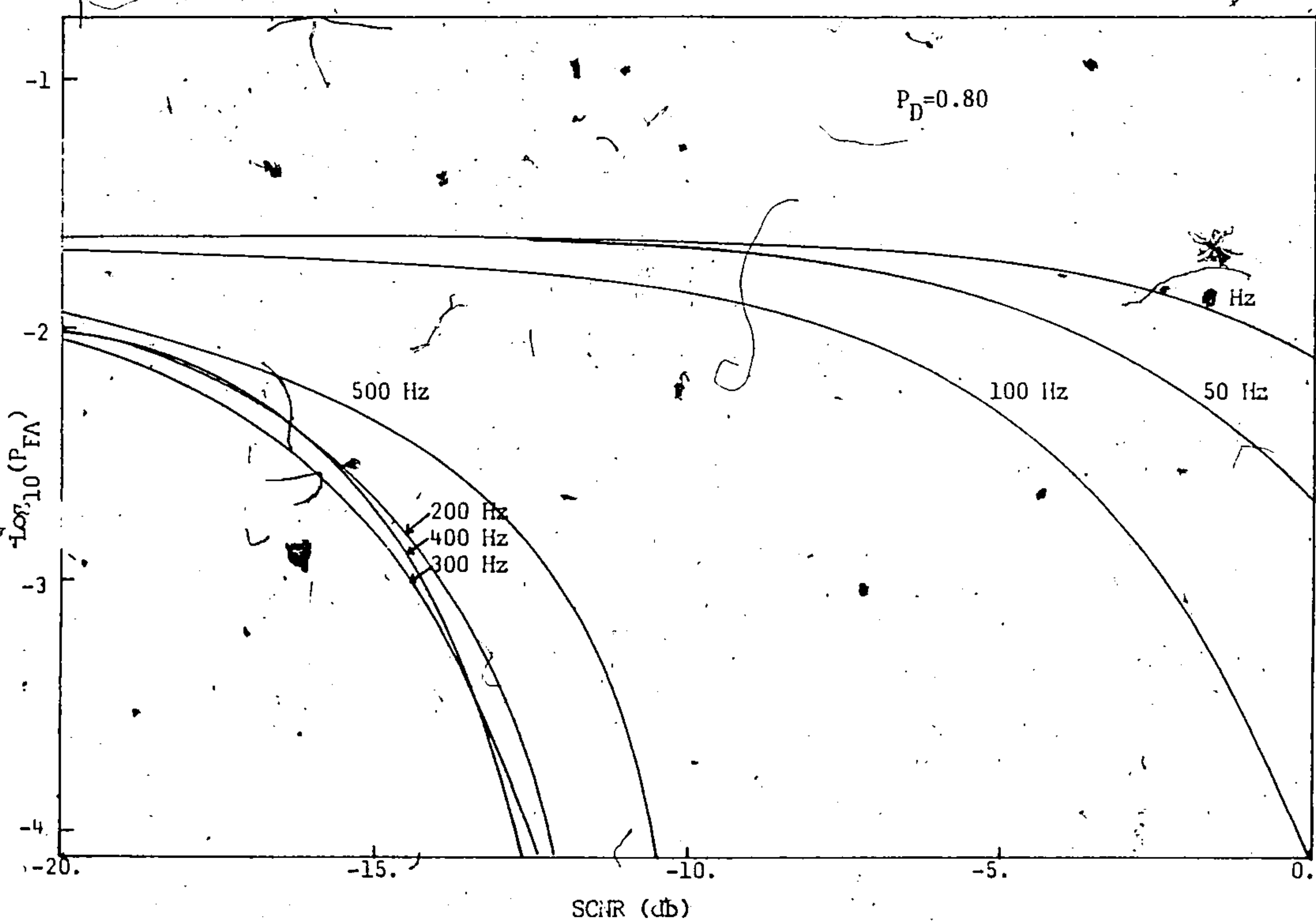


Fig. 5.14 The ROC for the MTD-3 (ground clutter dominated radar data)

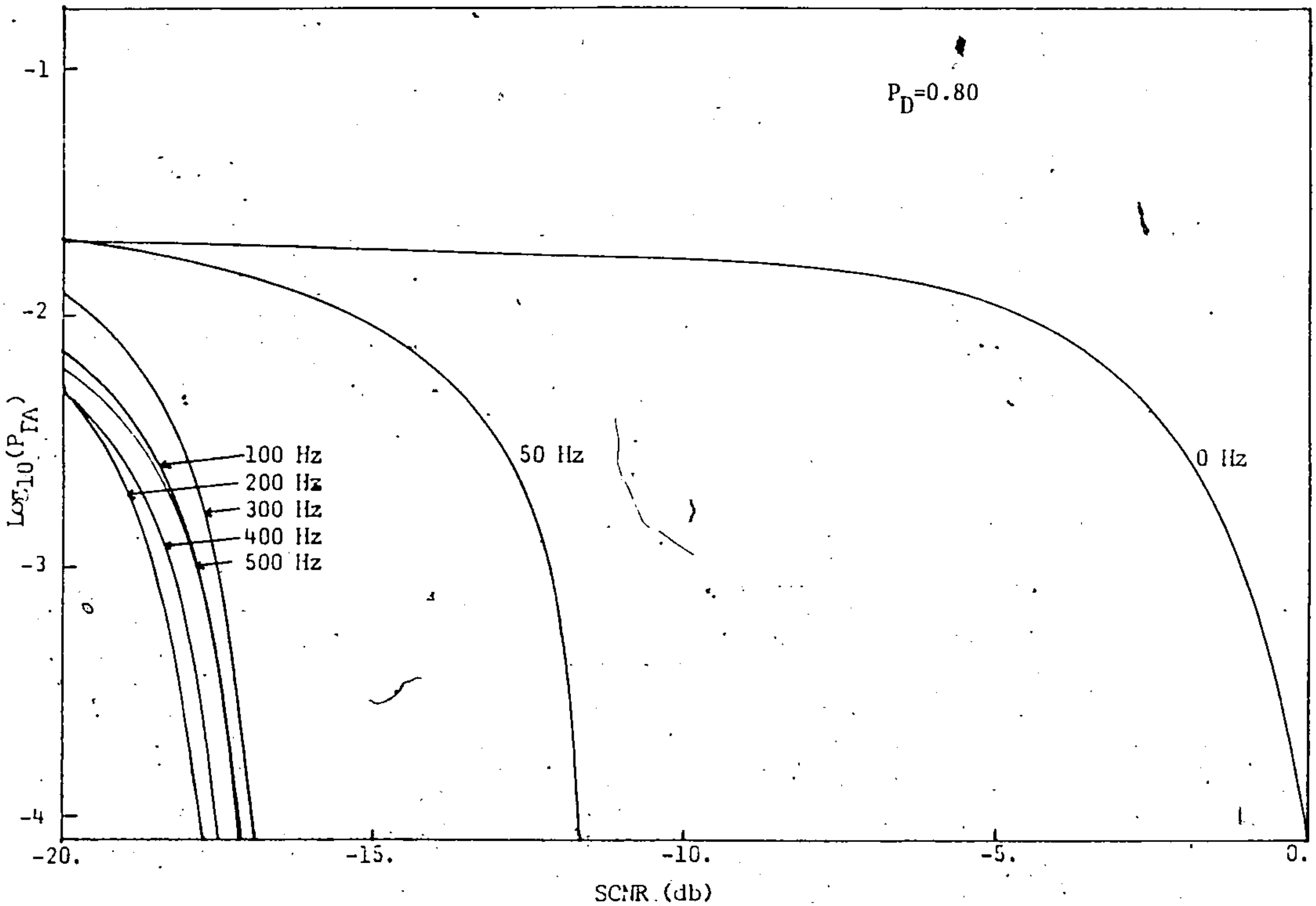


Fig 5.15 The ROC for the IBDA(LSL) (ground clutter dominated radar data)

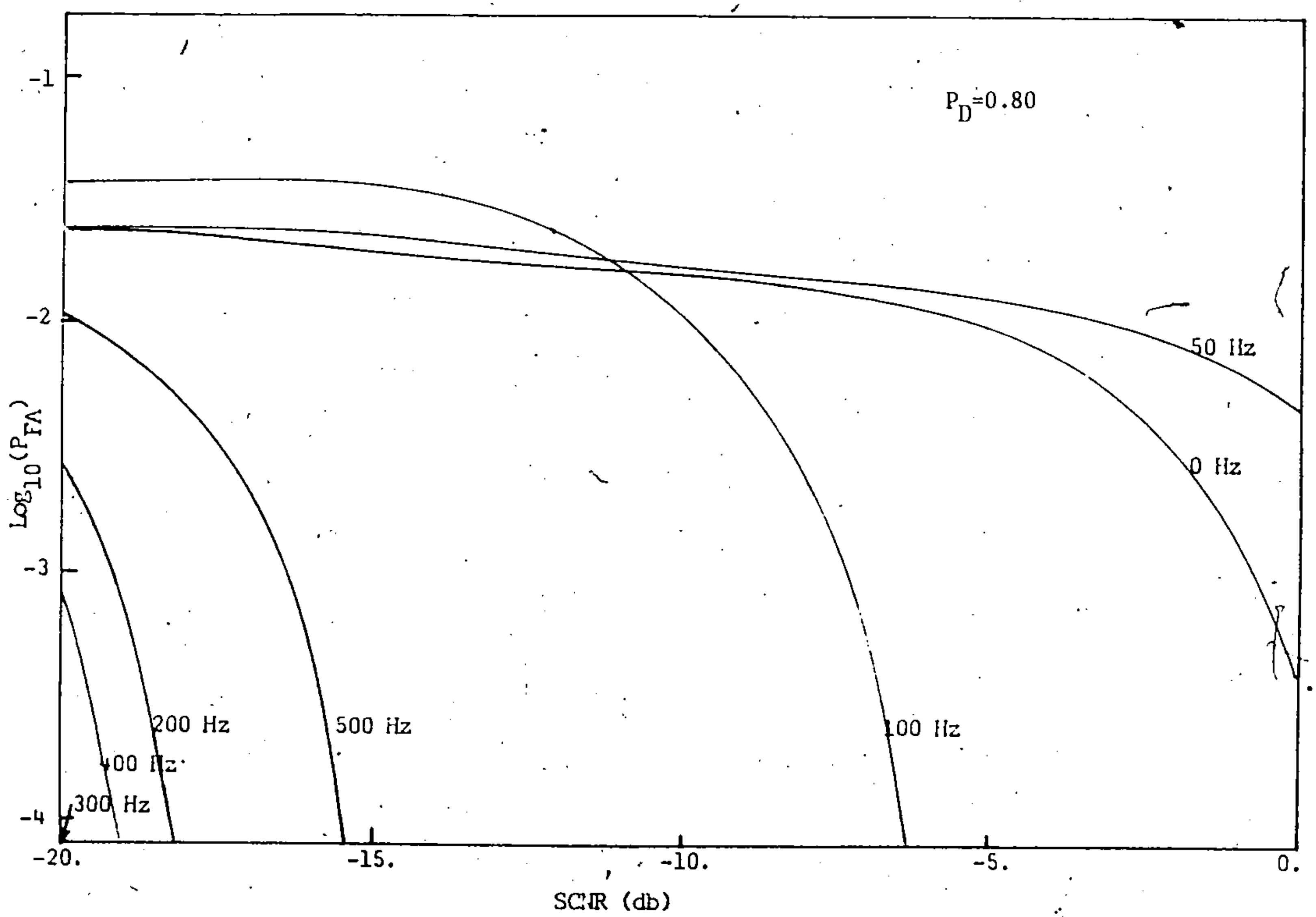


Fig. 5.16 The ROC for the IBDA(KALMAN) (ground clutter dominated radar data)

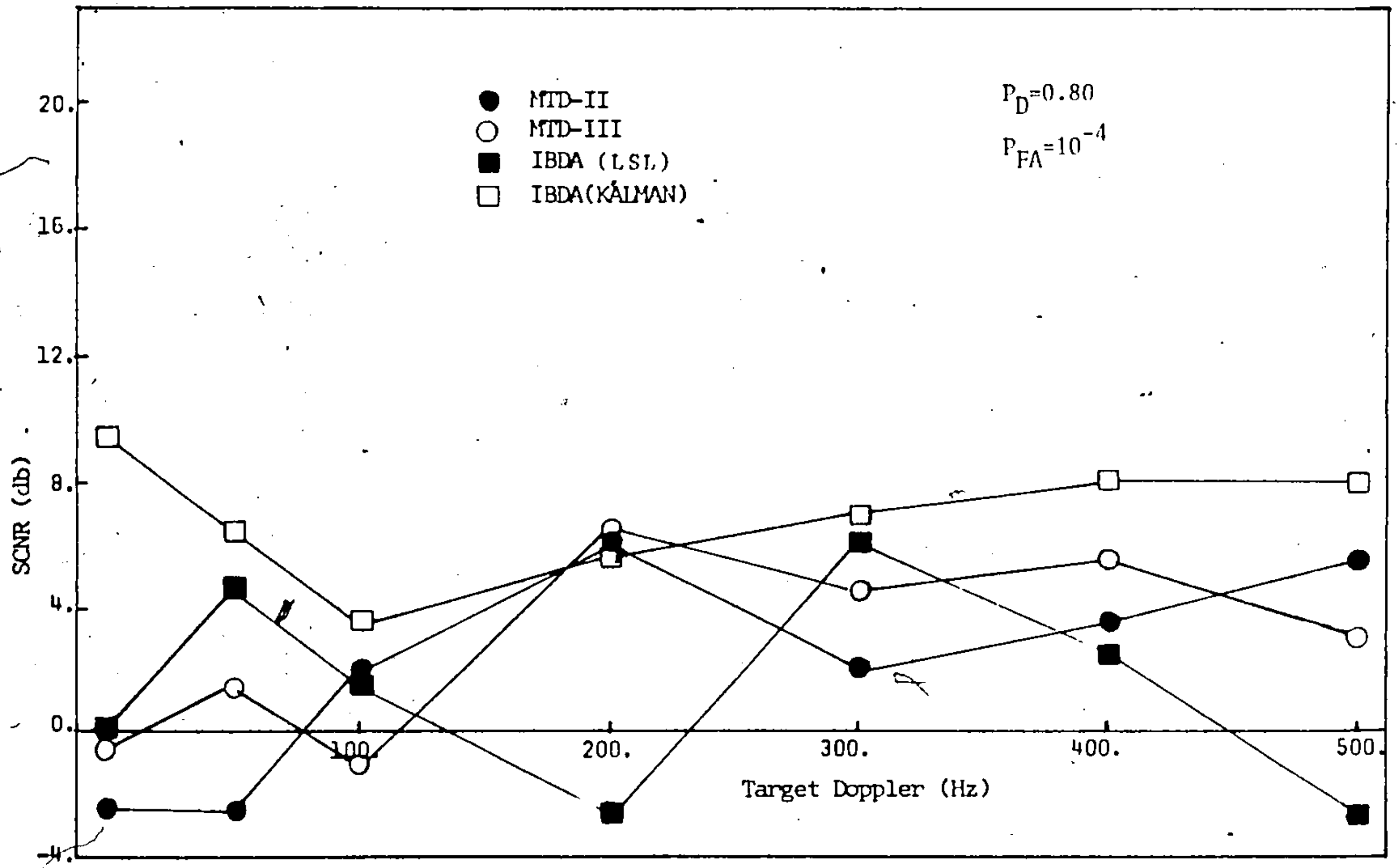


Fig. 5.17 The performance comparisions relative to the MTD-1 (weather clutter dominated radar data)

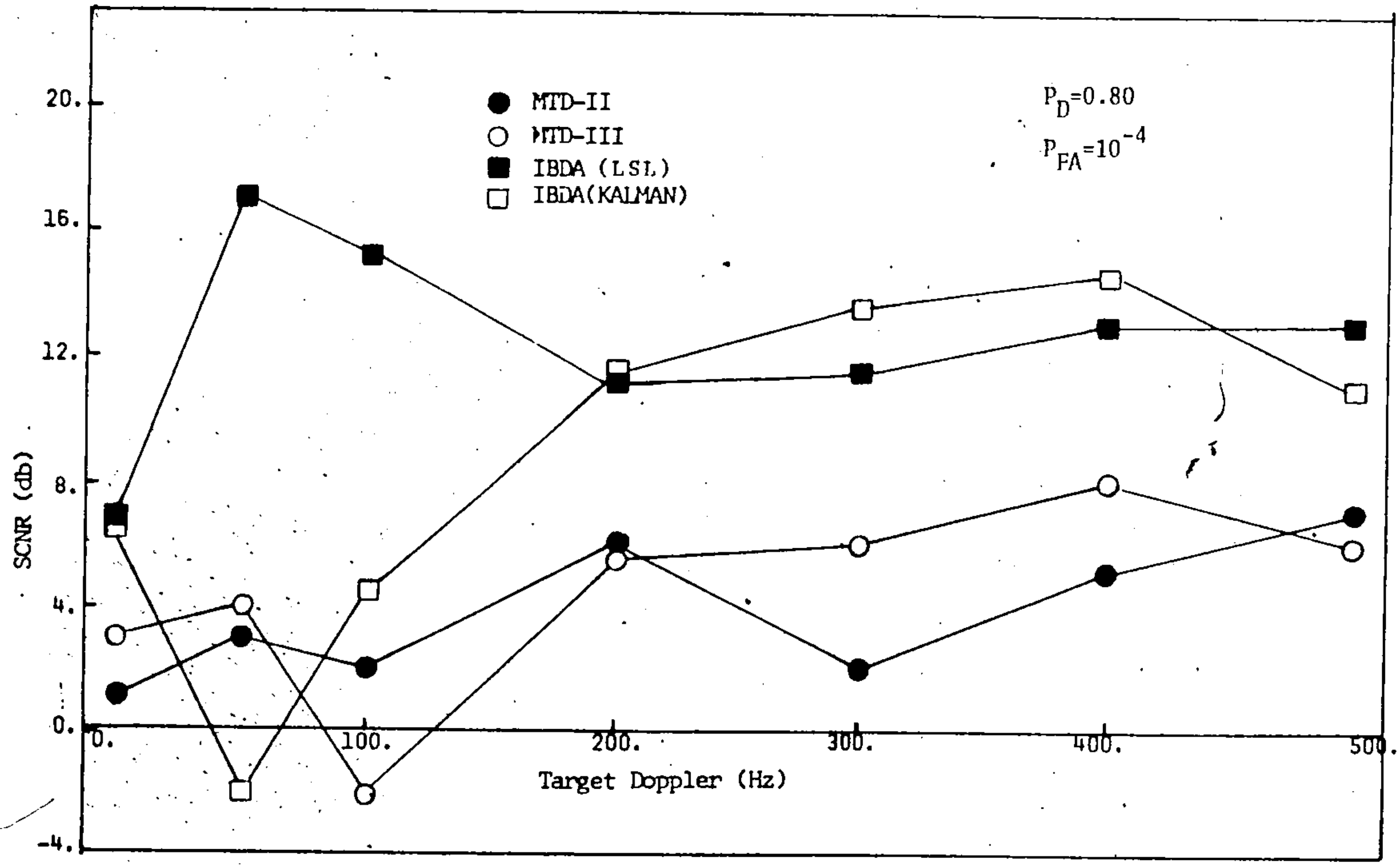


Fig. 5.18 The performance comparisons relative to the MTD-1 (ground clutter dominated radar data)

MTD-2 with respect to the MTD-1 and of the MTD-3 with respect to the MTD-2 is in agreement with the results and predictions made by Karp and Anderson [37] and by Taylor [62]. This is a confirmation of the validity of the methodology used to obtain the sets of ROC curves.

The IBDA (LSL) is implemented using fifth order lattice prediction error filters with $(\mu_{H_1}, \mu_{H_2}) = (1.0, 0.5)$. Good performance relative to the MTD algorithms is achieved for ground clutter dominated radar data (Figure 5.18). However very poor performance is demonstrated for weather clutter dominated radar data (Figure 5.17). Although many other sets of parameters were utilised in an attempt to improve this performance, the results were negative. A possible reason for this erratic behaviour of the IBDA (LSL) implementation is that the adaptive prediction error lattice filter may not be as capable of tracking the spatial variations of the weather clutter process (which is known to be highly nonstationary) as well as is necessary.

The performance of the LSL implementation of the prediction error lattice filter was also examined by allowing the filter orders for the two lattice filters to be different and by making the two damping constants equal to each other. This is the implementation suggested by the model of the radar target detection problem utilised by the IBDA (KALMAN). The performance of this implementation did not improve with respect to that of the IBDA (LSL).

The IBDA (KALMAN) is implemented with $(M_{H_1}, M_{H_2}) = (1, 5)$ and $r = q = 0.1$. For weather clutter dominated radar data (Figure 5.17), the IBDA (KALMAN) is always at least as good as the MTD algorithm, and has approximately a 4 dB average improvement. For ground clutter dominated radar data the IBDA (KALMAN) has an average improvement of 3 dB with respect to the MTD algorithms. It should be pointed out that there is a noticeable degradation in performance at 50 Hz similar to the degradation in performance of the MTD-3 of 100 Hz (Figure 5.18).

The performance of the IBDA (KALMAN) with respect to that of the IBDA (LSL) in weather clutter dominated radar data confirms the poor adaptivity of the LSL algorithms with respect to the adaptivity of the Kalman algorithm in the presence of weather clutter.

5.4 Conclusions

An examination of the performance of the IBDA is made within the context of the detection of aircraft by airport surveillance radar systems. In this application of the IBDA, the LLR is essentially the difference between the outputs of two whitening or linear prediction error filters. The first of these filters is designed so as to whiten the data under the hypothesis that the data is composed of a target response, the clutter and the usual WGN component. The second filter is designed to whiten the data under the alternate hypothesis that the data is composed only of clutter and the WGN.

As a result, the performance of this application of the IBDA depends not only upon the robustness of the necessarily adaptive whitening filters but also upon a proper identification of those spectral properties of the input data, under the two hypotheses, which are pertinent to the design of the whitening filters.

The performances of two different implementations of the IBDA are examined. The IBDA (LSL) uses the LSL form of the exact method of least squares, and is designed to utilise the experimentally observed spectral peak width difference in the data under the two hypotheses. A satisfactory improvement in performance relative to the MTD-3 algorithm is observed for ground clutter dominated radar data. However, there is a significant degradation in performance for weather clutter dominated radar data. This suggests that a simple utilisation of the peak width is not a sufficiently robust characterisation of the spectral properties of radar data.

The second implementation, the IBDA (KALMAN) is designed by making the stronger statement that all of the peaks in the power spectrum have Gaussian shapes. Adaptive prediction error filter structures (based on the Kalman filter, assuming a random walk state model) are utilised. This implementation leads to a 4 dB average improvement for weather clutter dominated radar data and a 3 dB average improvement for ground clutter dominated radar data as compared to the MTD-3 algorithm.

In conclusion, it is observed that when properly implemented, the IBDA is capable of at least a 3 dB average improvement in performance relative to the MTD algorithms which are presently in use in air traffic control radar systems.

This observation is in agreement with the realisation that the MTD algorithms may be regarded as limiting forms of the IBDA as applied to the radar target detection problem.

5.5 Recommendations for Future Research

This thesis has established the improvement in detection performance of the IBDA relative to existing systems. No examination of the resolution properties of this detection algorithm has been made. This is an important factor in the determination of the overall competitiveness of the IBDA as a viable radar system signal processing algorithm.

A second important figure of merit is the performance improvement versus cost increase tradeoff. A full-scale implementation of this system would require an array processor capable of performing on the order of 200 complex additions and 200 complex multiplications per array element every pulse repetition interval. This must be compared to the MTD-3 which performs approximately 100 complex additions and 64 complex multiplications every 8 pulse repetition intervals. Hence, the processor for the IBDA (KALMAN) must operate approximately 64 times faster than the processor in the MTD-3.

To place this needed increase in processing power in perspective, it should be noted that a radar system implemented with the IBDA (KALMAN) and having the same performance as a radar system implemented with the MTD-3 needs only one-half the front end power requirements as that required by the MTD-1 radar system.

APPENDIX 1

SIMPLIFICATION OF SOME MATHEMATICAL TERMS IN THE IBDA

Consider the correlation matrix $\underline{C}(N-1)$ where the (k,k')-th matrix element:

$$C(N-1;k,k') = \min(k+1,k'+1)\frac{\sigma_w^2}{\sigma_s(k)\sigma_s(k')}, \quad k,k' = 0,\ldots,N-1 \tag{A1.1}$$

For $N-1 = 1$

$$\underline{C}(1) = \begin{bmatrix} \dfrac{\sigma_w^2}{\sigma_s^2(o)} & \dfrac{\sigma_w^2}{\sigma_s(o)\sigma_s(1)} \\[2ex] \dfrac{\sigma_w^2}{\sigma_s(o)\sigma_s(1)} & \dfrac{2\sigma_w^2}{\sigma_s^2(1)} \end{bmatrix}$$

Thus $$|\underline{C}(1)| = \frac{\sigma_w^4}{\sigma_s^2(o)\sigma_s^2(1)},$$

and $$\underline{C}^{-1}(1) = \frac{1}{|\underline{C}(1)|}\begin{bmatrix} \dfrac{2\sigma_w^2}{\sigma_s^2(1)} & -\dfrac{\sigma_w^2}{\sigma_s(o)\sigma_s(1)} \\[2ex] -\dfrac{\sigma_w^2}{\sigma_s(o)\sigma_s(1)} & \dfrac{\sigma_w^2}{\sigma_s^2(o)} \end{bmatrix}$$

$$= \frac{1}{\sigma_w^2} \begin{bmatrix} 2\sigma_s^2(o) & -\sigma_s(o)\sigma_s(1) \\ & \\ -\sigma_s(o)\sigma_s(1) & \sigma_s^2(1) \end{bmatrix}$$

Since

$$\underline{y}\ \underline{C}^{-1}(1)\underline{y}^T = \left[y_o,\ y_1\right]\underline{C}^{-1}(1) \begin{bmatrix} y_o \\ y_1 \end{bmatrix}$$

but $$y_k = \frac{1}{\sigma_s(k)} \sum_{\ell=o}^{k} x(\ell)$$

Then

$$\underline{y}\underline{C}^{-1}(1)\underline{y}^T = \begin{bmatrix} \frac{1}{\sigma_w^2} & \frac{x(o)}{\sigma_s(o)}, & \frac{x(o)+x(1)}{\sigma_s(1)} \end{bmatrix} \begin{bmatrix} 2\sigma_s^2(o) & -\sigma_s(o)\sigma_s(1) \\ & \\ -\sigma_s(o)\sigma_s(1) & \sigma_s^2(1) \end{bmatrix} \begin{bmatrix} \frac{x(o)}{\sigma_s(o)} \\ \\ \frac{x(o)+x(1)}{\sigma_s(1)} \end{bmatrix}$$

$$= \frac{1}{\sigma_w^2}\,(x(o)^2 + x(1)^2)$$

This calculation may be repeated for other values of N-1 and it is concluded that

$$|\underline{C}(N-1)| = \prod_{k=o}^{N-1} \frac{\sigma_w^2}{\sigma_s^2(k)}$$

and $\underset{\sim}{y}\ \underset{\sim}{C}(N-1)\underset{\sim}{y}^T = \frac{1}{\sigma_w^2} \sum_{k=o}^{N-1} x^2(k)$

Consider now the correlation matrix $\underset{\sim}{R}_\rho(N-1)$ where the (k,k')-th matrix element

$$R_\rho(N-1;k,k') = \begin{cases} \dfrac{\sigma_s^2(k)}{\sigma_s(k)\sigma_s(k')}, & k < k' \\ 1, & k = k' \\ \dfrac{\sigma_s^2(k')}{\sigma_s(k)\sigma_s(k')}, & k > k' \end{cases}$$

For N-1 = 1 ,

$$\underset{\sim}{R}_\rho(1) = \begin{bmatrix} 1 & \dfrac{\sigma_s(o)}{\sigma_s(1)} \\ \dfrac{\sigma_s(o)}{\sigma_s(1)} & 1 \end{bmatrix}$$

Thus $|\underset{\sim}{R}_\rho(1)| = \dfrac{\sigma_s^2(1)-\sigma_s^2(o)}{\sigma_s^2(1)}$

and $R_\rho^{-1}(1) = \dfrac{\sigma_s^2(1)}{\sigma_s^2(1)-\sigma_s^2(o)} \begin{bmatrix} 1 & -\dfrac{\sigma_s(o)}{\sigma_s(1)} \\ -\dfrac{\sigma_s(o)}{\sigma_s(1)} & 1 \end{bmatrix}$

$$= \begin{bmatrix} \dfrac{\sigma_s^{\,2}(1)}{\sigma_s^{\,2}(1)-\sigma_s^{\,2}(o)} & -\dfrac{\sigma_s(o)\sigma_s(1)}{\sigma_s^{\,2}(1)-\sigma_s^{\,2}(o)} \\ -\dfrac{\sigma_s(o)\sigma_s(1)}{\sigma_s^{\,2}(1)-\sigma_s^{\,2}(o)} & \dfrac{\sigma_s^{\,2}(1)}{\sigma_s^{\,2}(1)-\sigma_s^{\,2}(o)} \end{bmatrix}$$

Since

$$(\underline{y}-\underline{\xi})\underline{R}_\rho^{-1}(1)\ (y-\underline{\xi})^T = \begin{bmatrix} y(o)-\xi(o), & y(1)-\xi(1) \end{bmatrix} \underline{R}_\rho^{-1}(1) \begin{bmatrix} y(o)-\xi(o) \\ y(1)-\xi(1) \end{bmatrix}$$

but $y_k = \dfrac{1}{\sigma_s(k)} \sum\limits_{e=o}^{k} x(k)$ and $\xi_k = \dfrac{1}{\sigma_s(k)} \sum\limits_{e=0}^{k} \hat{x}(k|H_1)$

or $y_k-\xi_k = \dfrac{1}{\sigma_s(k)} \sum\limits_{e=o}^{k} \tilde{x}(k)-\hat{x}(k|H_1)$

$$= \frac{1}{\sigma_s(k)} \sum_{e=o}^{k} \tilde{x}(k|H_1)$$

Then

$$(\underline{y}-\underline{\xi})\underline{R}_\rho^{-1}(1)(\underline{y}-\underline{\xi})^T =$$

$$= \begin{bmatrix} \dfrac{\tilde{x}(o|H_1)}{\sigma_s(o)}, & \dfrac{\tilde{x}(o|H_1)+\tilde{x}(1|H_1)}{\sigma_s(1)} \end{bmatrix} \begin{bmatrix} \dfrac{\sigma_s^{\,2}(1)}{\sigma_s^{\,2}(1)-\sigma_s^{\,2}(o)} & -\dfrac{\sigma_s(o)\sigma_s(1)}{\sigma_s^{\,2}(1)-\sigma_s^{\,2}(o)} \\ -\dfrac{\sigma_s(o)\sigma_s(1)}{\sigma_s^{\,2}(1)-\sigma_s^{\,2}(o)} & \dfrac{\sigma_s^{\,2}(1)}{\sigma_s^{\,2}(1)-\sigma_s^{\,2}(o)} \end{bmatrix} \begin{bmatrix} \dfrac{\tilde{x}(o|H_1)}{\sigma_s(o)} \\ \dfrac{\tilde{x}(o|H_1)+\tilde{x}(1|H_1)}{\sigma_s(1)} \end{bmatrix}$$

$$= \frac{\tilde{x}^2(o|H_1)}{\sigma_s^2(o)} + \frac{\tilde{x}^2(1|H_1)}{\sigma_s^2(1)-\sigma_s^2(o)}$$

Again, this calculation may be repeated for other values of N-1 and it is concluded that

$$|\underline{R}_\rho(N-1)| = \prod_{k=1}^{N-1} \frac{(\sigma_s^2(k)-\sigma_s^2(k-1))}{\sigma_s^2(k)}$$

and $(\underline{y}-\underline{\xi})\underline{R}_\rho^{-1}(N-1)(\underline{y}-\underline{\xi})^T = \sum_{k=o}^{N-1} \frac{\tilde{x}^2(k|H_1)}{\sigma_s^2(k)-\sigma_s^2(k-1)}$

where $\sigma_s^2(-1) \triangleq 0$

By defining $\sigma^2(k|H_1) \triangleq \sigma_s^2(k)-\sigma_s^2(k-1)$

$$|\underline{R}_\rho(N-1)| = \prod_{k=o}^{N-1} \frac{\sigma^2(k|H_1)}{\sigma_s^2(k)}$$

and $(\underline{y}-\underline{\xi})\underline{R}_\rho^{-1}(N-1)(\underline{y}-\underline{\xi})^T = \sum_{k=o}^{N-1} \frac{\tilde{x}^2(k|H_1)}{\sigma^2(k|H_1)}$

APPENDIX 2

THE ADAPTIVE LEAST SQUARES LATTICE PREDICTION ERROR FILTER

The following lattice algorithm is a generalisation of the lattice structure as derived by Schichor [58] to include both complex data and exponential weighting of the data. This exponential weighting allows the past data to be forgotten, thereby enabling the tracking of time variations in the data's statistical properties.

The following symbols are defined where (Figure 4.2):

$P_{e,M}(n)$ is the M-th order forward prediction error power at time n

$P_{r,M}(n)$ is the M-th order backward prediction error power at time n

$K_M(n)$ is the M-th order unnormalised reflection coefficient at time n

$e_M(n)$ is the M-th order forward prediction error at time n

$r_M(n)$ is the M-th order backward prediction error at time n

$\varepsilon_M(n)$ is the M-th order forward pseudo prediction error at time n

$\rho_M(n)$ is the M-th order backward pseudo prediction error at time n

μ is the damping cosntant

Then, with the following initial parameters:

$$\varepsilon_o(n) = \rho_o(n) = e_o(n) = r_o(n) = x(n)$$

$$K_o(1) = 0$$

$$P_{r,o}(0) = P_{e,o}(1) = |x(o)|^2$$

$$0 \leq \mu \leq 1$$

the ordering of the recursive equations for updating the lattice parameters when a new data sample x(n) is input to the lattice is as follows:

$$\varepsilon_M(n) = \varepsilon_{M-1}(n) - \frac{K_{M-1}(n-1)}{P_{r,M-1}(n-2)} \rho_{M-1}(n-1)$$

$$\rho_M(n) = \rho_{M-1}(n-1) - \frac{K^*_{M-1}(n-1)}{P_{e,M-1}(n-1)} \varepsilon_{M-1}(n)$$

$$K_{M-1}(n) = \mu K_{M-1}(n-1) + \varepsilon_{M-1}(n) r^*_{M-1}(n-1)$$

$$P_{e,M-1}(n) = \mu P_{e,M-1}(n-1) + e_{M-1}(n) \varepsilon^*_{M-1}(n)$$

$$P_{r,M-1}(n-1) = \mu P_{r,M-1}(n-2) + r_{M-1}(n-1)\rho^{*}_{M-1}(n-1)$$

$$e_M(n) = e_{M-1}(n) - \frac{K_{M-1}(n)}{P_{r,M-1}(n)} r_{M-1}(n)$$

$$r_M(n) = r_{M-1}(n-1) - \frac{K^{*}_{M-1}(n)}{P_{e,M-1}(n)} e_{M-1}(n)$$

$$P_{e,M}(n) = P_{e,M-1}(n) - \frac{\left|K_{M-1}(n)\right|^2}{P_{r,M-1}(n-1)}$$

$$P_{r,M}(n-1) = P_{r,M-1}(n-2) - \frac{\left|K_{M-1}(n-1)\right|^2}{P_{e,M-1}(n-1)}$$

APPENDIX 3

THE KALMAN PREDICTION ERROR TAPPED DELAY-TIME FILTER

The following adaptive linear prediction error algorithm results from the use of Kalman prediction theory where the assumption is made that the optimum tap-weight vector is randomly varying about a mean value (Zhang and Haykin [69]).

The following symbols are defined where, if given an M-th order tapped delay-line prediction error filter then:

$x(n)$ is the received complex data sample at time n

$\underline{x}(n-1)$ is the M length vector of the past M data samples

$\underline{K}(n)$ is the M length Kalman gain vector

$\hat{h}(n|n-1)$ is the optimal (MMSE) M length vector of the tap weights of time n given the past data

$\underline{P}(n)$ is the MxM predicted prediction error covariance matrix

$\underline{\Gamma}(n)$ is the MxM filtered prediction error covariance matrix

$q\underline{I}$ is the MxM covariance matrix of the white noise tap weight process r is the variance of the prediction error process.

Given $\hat{h}(o)$, $\underline{P}(o)$, r; q these quantities are recursively updated as follows:

$$\underline{K}(n) = \underline{P}(n)\underline{x}^H(n-1)[\underline{x}(n-1)\underline{P}(n)\underline{x}^H(n-1)+r]^{-1}$$

$$\hat{\underline{h}}(n+1|n) = \hat{\underline{h}}(n|n-1)+\underline{K}(n)[x(n)-\underline{x}(n-1)\hat{\underline{h}}(n|n-1)]$$

$$\underline{\Gamma}(n) = \underline{P}(n)-\underline{K}(n)\underline{x}(n-1)\underline{P}(n)$$

$$\underline{P}(n+1) = \underline{\Gamma}(n)+q\underline{I}$$

REFERENCES

[1] H. Aasnaes and T. Kailath, "An Innovations Approach to Least Squares Estimation, Part 7, Some Applications of Vector Autoregressive Moving Average Models", IEEE Trans. on Automatic Control, Vol. AC-18, pp. 601-607, (1983).

[2] J. Abbot and J. Blum, "On a Theorem of Renyi Concerning Mixing Sequences of Sets", Annals Math. Stat., Vol. 32, pp. 757-760, (1961).

[3] E. Ataman, V. K. Aatre and K. M. Wong, "Some Statistical Properties of Median Filters", IEEE Trans. on ASSP, Vol. ASSP-29, pp. 1073-1083, (1981).

[4] D. Barton, "Radar Clutter", Radars, Vol. 5, Artech House, Inc., Dedham, Massachusetts, (1975).

[5] H. Bode and C. Shannon, "A Simplified Derivation of Linear Least Square Smoothing and Prediction Theory", Proc. IRE, Vol. 38, pp. 417-425, (1950).

[6] G. Box and G. Jenkins, "Time Series Analysis, Forecasting and Control", Holden Day, San Francisco, (1976).

[7] L. Brooks and I. Reed, "Equivalence of the Likelihood Ratio Processor, the Maximum Signal-to-Noise Ratio Filter and the Wiever Filter", IEEE Trans. on Aerospace and Electronic Systems, pp. 690-692, (1972).

[8] D. deLong and E. Hofstetter, "On the Design of Optimum Radar Waveforms for Clutter Rejection", IEEE Trans. on Information Theory, Vol. IT-13, pp. 454-463, (1967).

[9] J. Doob, "Stochastic Processes", John Wiley and Sons, New York, (1953).

[10] W. Drury, "Improved MTI Radar Signal Processor", MIT Lincoln Laboratory Report ATC-39, Report #FAA-RD-74-185, (available through NTIS).

[11] W. Feller, "An Introduction to Probability Theory and its Applications", Vol. 2, John Wiley and Sons, New York, (1971).

[12] A. Fogel and S. C. Schwartz, "Limiting Forms of the Discrete-Time Estimator Correlator Detector", IEEE Trans. on Information Theory, Vol. IT-30, pp. 296-299, (1983).

[13] B. Friedlander, "Lattice Filters for Adaptive Processing", Proc. IEEE, Vol. 70, No. 8, pp. 829-867, (1982).

[14] P. Frost and T. Kailath, "An Innovations Approach to Least Squares Estimation, Part 3, Nonlinear Estimation in White Gaussian Noise", IEEE Trans. on Automatic Control, Vol. AC-16, pp. 217-226, (1971).

[15] M. Gevers and T. Kailath, "An Innovations Approach to Least Squares Estimation, Part 4, Discrete-Time Innovations Representation and Recursive Estimation", IEEE Trans. on Automatic Control, Vol. AC-18, pp. 588-600, (1973).

[16] C. Gibson, "Adaptive Lattice Filtering for Radar Applications", Ph.D. Thesis, McMaster University, (1982).

[17] L. Griffiths, "A Continuously Adaptive Filter Implemented as a Lattice Structure", Proc. 1977 IEEE Int. Conf. on ASSP, pp. 683-686, (1977).

[18] P. Halmos, "Measure Theory", Springer-Verlag, New York, (1974).

[19] S. Haykin, "Non-linear Methods of Spectral Analysis", Springer-Verlag, New York, (1979).

[20] S. Haykin, "Communication Systems", John Wiley and Sons, New York, (1978).

[21] S. Haykin, B. W. Currie and S. P. Kesler, "Maximum-Entropy Spectral Analysis of Radar Clutter", Proc. IEEE, Vol. 70, No. 9, pp. 953-962, (1982).

[22] C. Helstrom, "Statistical Theory of Signal Detection", 2-nd edition, Pergammon Press, New York, (1968).

[23] T. Kadota and L. Shepp, "Conditions for Absolute Continuity Between a Certain Pair of Probability Measures", Z. Wahr, Verw. Geb. Vol. 16, pp. 250-260, (1970).

[24] T. Kailath, "A View of Three Decades of Linear Filtering Theory", IEEE Trans. on Information Theory, Vol. IT-20, pp. 146-181, (1974).

[25] T. Kailath, "The Innovations Approach to Detection Theory and Estimation Theory", Proc. IEEE, Vol. 58, pp. 680-695, (1970).

[26] T. Kailath, "A Note on Least Squares Estimation by the Innovations Method", Siam J. Control, Vol. 10, pp. 477-486, (1972).

[27] T. Kailath, "An Innovations Approach to Least Squares Estimation, Part 1: Linear Filtering in Additive White Noise", IEEE Trans. on Automatic Control, Vol. AC-13, pp. 646-654, (1968).

[28] T. Kailath and P. Frost, "An Innovations Approach to Least Squares Estimation, Part 2: Linear Smoothing in Additive White Noise", IEEE Trans. on Automatic Control, Vol. AC-13, pp. 655-660, (1968).

[29] T. Kailath and R. Geesey, "An Innovations Approach to Least Squares Estimation, Part 4: Recursive Estimation Given Lumped Covariance Functions", IEEE Trans. on Automatic Control, Vol. AC-16, pp. 720-727, (1971).

[30] T. Kailath and R. Geesey, "An Innovations Approach to Least Squares Estimation, Part 5: Innovations Representations and Recursive Estimation in Coloured Noise", IEEE Trans. on Automatic Control, Vol. AC-18, pp. 435-453, (1973).

[31] T. Kailath, "Some Extensions of the Innovations Theorem", The Bell Systems Technical Journal, Vol. 50, pp. 1487-1494, (1971).

[32] T. Kailath, "A General Likelihood Ratio Formula for Random Signals in Gaussian Noise", IEEE Trans. on Information Theory, Vol. IT-15, pp. 350-361, (1969).

[33] T. Kailath, "A Further Note on a General Likelihood Formula for Random Signals in Gaussian Noise", IEEE Trans. on Information Theory, Vol. IT-16, pp. 393-396, (1970).

[34] T. Kailath, "Likelihood Ratios for Gaussian Processes", IEEE Trans. on Information Theory, Vol. IT-16, pp. 276-288, (1970).

[35] R. Kalman, "A New Approach to Linear Filtering and Prediction Problems", Trans. ASME, J. Basic Engineering, Vol. 82, pp. 34-45, (1960).

[36] R. Kalman and R. Bucy, "New Results in Linear Filtering and Prediction Theory", Trans. ASME, J. Basic Engineering, Vol. 83, pp. 95-107, (1961).

[37] D. Karp and J. Anderson, "Moving Target Detector (Mod II) Summary Report", MIT Lincoln Laboratory Project Report ATC-95, (available NTIS).

[38] S. Kay, "Asymptotically Optimal Detection in Unknown Coloured Noise via Autoregressive Modelling", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-31, No. 4, pp. 927-940, (1983).

[39] E. Kelly, J. Reed and W. Root, "The Detection of Radar Echos in Noise, Part 1", J. Soc. Indust. Appl. Math., Vol. 8, pp. 309-338, (1960).

[40] E. Kelly, J. Reed and W. Root, "The Detection of Radar Echose in Noise, Part 2", J. Soc. Indust. Appl. Math., Vol. 8, pp. 481-507, (1960).

[41] S. Kesler, "Nonlinear Spectral Analysis of Radar Clutter", Ph.D. Thesis, McMaster University, (1978).

[42] E. Lehmann, "Testing Statistical Hypotheses", John Wiley and Sons, New York, (1959).

[43] M. Loeve, "Probability Theory", D. Van Nostrand, Princeton, (1960).

[44] P. A. S. Metford, S. Haykin and D. P. Taylor, "An Innovations Approach to Discrete-Time Detection Theory", IEEE Trans. on Information Theory, Vol. IT-28, pp. 376-380, (1982).

[45] P.A.S. Metford and S. Haykin, "Experimental Analysis of an Innovations Based Detection Algorithm for Surveillance Radar", Proc. IEE (London) accepted for publication.

[46] MIT Radiation Laboratory Series, Vol. 1, "Radar System Engineering", McGraw-Hill, New York, (1947).

[47] C. Muehe, "Moving Target Detection, An Improved Signal Processor", AGARD Conf. Proceeding, #197, pp. 14, 1-14, 10, (1976).

[48] C. Muehe et al., "New Techniques Applied to Air Traffic Control Radars", Proc. IEEE, Vol. 62, pp. 716-723, (1974).

[49] R. O'Donnell and C. Muehe, "Automated Tracking for Aircraft Surveillance Radar Systems", IEEE Trans. on Aerospace and Electronic Systems, Vol. AES-15, pp. 508-517, (1979).

[50] A. Papoulis, "Probability, Random Variables and Stochastic Processes", McGraw-Hill, New York, (1955).

[51] B. Porat and B. Friedlander, "Parametric Techniques in Adaptive Signal Detection", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-31, No. 3, pp. (1983).

[52] L. R. Rabiner and B. Gold, "Theory and Application of Digital Signal Processing", Prentice Hall, New Jersey, (1975).

[53] R. A. Scholtz, J. J. Kappl and N. E. Nahi, "The Detection of Moderately Fluctuating Rayleigh Targets", IEEE Trans. on AES, Vol. AES-12, pp. 117-125, (1976).

[54] S. C. Schwartz, "Conditional Mean Estimates and Bayesian Hypothesis Testing", IEEE Trans. on Information Theory, Vol. IT-21, pp. 663-665, (1975).

[55] S. C. Schwartz, "The Estimator Correlator for Discrete-Time Problems", IEEE Trans. on Information Theory, Vol. IT-23, pp. 93-100, (1977).

[56] M. Sekine, S. Ohtani and T. Musha, "Weibull-Distributed Ground Clutter, IEEE Trans., in Aerospace and Electronic Systems, Vol. AES-17, No. 4, pp. 596-599, (1981).

[57] S. Sherman, "Non-Mean-Square Error Criterion", IRE Trans. on Information Theory, pp. 125-125, (1958).

[58] F. Shichor, "Fast Recursive Estimation Using the Lattice Structure", The Bell System Technical Journal, Vol. 61, pp. 97-114, (1982).

[59] M. Skolnik, "Introduction to Radar Systems", McGraw Hill, New York, (1962).

[60] V. Statulevicius, "Limiting Theorems and Their More Specific Forms for Additive Stochastic Functions and the Sum of the Weakly Dependent Quantities", (in Russian), Trans. 3-rd Prague Conf. in Information Theory, Statistical Decision Functions and Random Functions, pp. 683-687.

[61] T. Taylor and G. Brunins, "Long Range Surveillance Radars for Automatic Control Systems", IEEE International Radar Conf., pp. 312-317, (1975).

[62] J. W. Taylor, "Sacrifices on Radar Clutter Suppression due to Compromises in Implementation of Digital Doppler Filters", IEE INt. Radar Conf., (1982).

[63] C. W. Therrien, "A Sequential Approach to Target Discrimination", IEEE Trans. on AES, Vol. AES-14, pp. 433-440, (1978).

[64] H. Van Trees, "Detection, Estimation and Modulation Theory, Part 1, Detection, Estimation and Linear Modulation Theory", John Wiley and Sons, New York, (1968).

[65] H. Van Trees, "Optimum Signal Design and Processing for Reverberation-Limited Environments", IEEE Trans. on Military Electronics, Vol. MIL-9, pp. 212-229, (1965).

[66] A. Whalen, "Detection of Signals in Noise", Academic Press, New York, (1980).

[67] N. Wiener, The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications", John Wiley and Sons, New York, (1949).

[68] P. Woodward, "Probability and Information Theory with Applications to Radar", Pergamon Press, Oxford, (1964).

[69] Qi-tu Zhang and S. Haykin, "Tracking Characteristics of the Kalman Filter in a Nonstationary Environment for Adaptive Filter Applications", Proc. 1983 IEEE Int. Conf. on ASSP, Vol. CH1841-6, pp. 671-674, (1983).