

ADAPTIVE DETECTION FOR RADAR

by

Qi-Tu Zhang

A thesis
submitted to the School of Graduate Studies
in partial fulfilment of the requirements
for the degree
Doctor of Philosophy

McMaster University



November 1985

ADAPTIVE DETECTION FOR RADAR

DOCTOR OF PHILOSOPHY (1985)
(Electrical Engineering)

MCMASTER UNIVERSITY
Hamilton, Ontario

TITLE: ADAPTIVE DETECTION FOR RADAR

AUTHOR: Qi-Tu Zhang, B.Sc. (Tsinghua University, Peking, China)
M.Eng. (South China Institute of Technology,
Canton, China)

SUPERVISOR: Dr. S. Haykin

NUMBER OF PAGES: xiv, 189

ABSTRACT

A new adaptive procedure for the detection of a random signal in a background of colored interference plus white noise, all having unknown statistics, is derived. The procedure is based on the development of linear models for the two hypotheses H_0 and H_1 . Under hypothesis H_0 a wanted signal is absent and under H_1 the wanted signal is present. After parameterizing the input signal under the two hypotheses, the log likelihood ratio is easily established in terms of the innovations processes. The test statistic thus obtained has an intuitive interpretation, which conversely provides a basis for the choice of model candidates for the signal detection problem. Different modelling approaches generally result in different receiver performance. Therefore, choosing appropriate type of model for the two hypotheses is a key step in order to enhance detection. A computationally efficient technique has been developed for this purpose.

This adaptive detection method is applied to construct a radar detection scheme for detecting a moving target in a surveillance radar environment. The detector

is composed of a pair of adaptive whitening filters followed by a log likelihood ratio calculator. The new scheme is tested by using actual radar data, including weather and ground clutter. Comparison is made with the previously described scheme known as innovations-based detection algorithm (IBDA). The results show that the performance of the new scheme is much better than the IBDA.

In a related study, the spatial correlation between radar clutter from adjacent rings is exploited to construct two new adaptive clutter suppression schemes. The schemes are tested by real radar data. Experimental results show that these schemes are around 5 dB and 10 dB better than the adaptive lattice prediction-error filter in terms of improvement factor, respectively.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Dr. Simon Haykin for his supervision, help and encouragement during this research. I am also grateful to Drs. P. Yip, K.M. Wong and D.P. Taylor for many useful discussions and suggestions for improving this thesis. I would also like to thank: Dr. Peter Metford for his helpful assistance; the Ministry of Colleges and Universities of the Ontario Government for their financial support, and Mrs. Aloma Munian for the expert typing of this thesis.

Finally, to my wife and son, Huisi and Howard, I wish to express my sincere appreciation of their support and understanding during my study for the Ph.D. degree.

CONTENTS

	<u>Page</u>
LIST OF PRINCIPAL SYMBOLS	viii
ABBREVIATIONS	xi
LIST OF FIGURES	xii
CHAPTER 1. INTRODUCTION	1
1.1 Purpose of the Research	1
1.2 Traditional and Adaptive Signal Detection	2
1.3 Organization of the Thesis	11
CHAPTER 2. ADAPTIVE WHITENING FILTER FOR NONSTATIONARY PROCESSES	13
2.1 LLR Test and Whitening Filter	13
2.2 Determination of the Filter Parameters	18
2.3 Explicit Solution for the Tap-Weight Vector	22
2.4 Analysis of the Algorithm	25
Appendix 2A: The Proof of a Lemma	29
CHAPTER 3. LOG-LIKELIHOOD RATIO RECEIVERS BASED ON LINEAR MODELLING	30
3.1 Introduction	30
3.2 LLR Test	35
3.2.1 Maximum Likelihood Function Under H_1	36
3.2.2 Maximum Likelihood Function Under H_0	39
3.2.3 Test Statistic	40
3.3 Test Statistic Based on ARAR Modelling	41
3.4 Adaptive LLR Detection Algorithm	42
3.5 Properties of LLR Test Based on REAR Modelling	58
3.6 Examples	67

CHAPTER 4	LINEAR MODEL BUILDING FOR THE ADAPTIVE DETECTION PROBLEM	78
4.1	Geometric Interpretation of Signal Detection	79
4.2	Model for the Input Signal Under H_1	83
4.3	Model for the Input Signal Under H_0	89
4.4	Summarizing Remarks	100
CHAPTER 5	APPLICATION TO RADAR DETECTION	104
5.1	Information Involved in Radar Clutter	105
5.2	Summary of the Adaptive Detection Algorithm	114
5.3	Determination of the Matrices Q^0 and Q^1	117
5.4	Computational Complexity	120
5.5	Estimation of the CFAR Threshold	121
5.6	Statistical Evaluation Using Actual Radar Data	123
5.6.1	Clutter Data	123
5.6.2	Target Signal	127
5.6.3	System Performance Evaluation	131
5.7	Test With Simulated Data	132
CHAPTER 6	RADAR CLUTTER SUPPRESSION SCHEMES	139
6.1	Formulation and Normal Equation	143
6.2	Random Signal Passes Through the Clutter Suppression System I	146
6.3	Analysis of System Improvement Factor	149
6.4	Implementation of Clutter Suppression Scheme I	152
6.5	System Output in the Frequency Domain	153
6.6	System Output in the Time Domain	167
6.7	Improvement Factor	170
6.8	Further Improvement -- Suppression Scheme II	176
CHAPTER 7	CONCLUSION	180
REFERENCES		183

LIST OF PRINCIPAL SYMBOLS

$\underline{a}_p$ = $(a_1, a_2, \dots, a_p)^T$, the model coefficient vector.

$\underline{\hat{a}}_p(\underline{y}|H_1)$ An estimate of $\underline{a}_p$ given observation data $\underline{y}$ under hypothesis H_1 .

$\underline{\hat{a}}_p(\underline{y}|H_0)$ An estimate of $\underline{a}_p$ given observation data $\underline{y}$ under hypothesis H_0 .

$\underline{b}_q$ = $(b_1, b_2, \dots, b_q)^T$, the model coefficient vector.

$\underline{\hat{b}}_q(\underline{y}|H_1)$ An estimate of $\underline{b}_q$ given observations $\underline{y}$ under hypothesis H_1 .

$\underline{\hat{b}}_q(\underline{y}|H_0)$ An estimate of $\underline{b}_q$ given observations $\underline{y}$ under hypothesis H_0 .

$\underline{c}(k)$ measurement matrix at time k .

$\underline{d}_{yx}$ = $E[\underline{y}(n)\underline{x}(n)^*]$

$\underline{d}_{cx}$ = $E[\underline{c}(n)\underline{x}(n)^*]$

$E[\cdot]$ Expectation operator.

$e(k)$ The error process of a linear model.

$\hat{e}(k|k-1)$ One-step forward prediction-error given observation up to time $k-1$.

$\hat{e}_0(k|k-1)$ Prediction-error based on H_0 -model.

$\hat{e}_1(k|k-1)$ Prediction-error based on H_1 -model.

$\underline{g}(k)$ Kalman gain vector at time k .

$\underline{I}$ Identity matrix.

k	Discrete time variable.
$l(k)$	Test statistic.
m	Discrete time variable.
$N(0, \sigma^2)$	Gaussian distribution with zero-mean and variance σ^2 .
$P(k)$	Predicted weight-vector covariance matrix.
$p(\cdot)$	Probability density function.
$R(k)$	Correlation function of a stationary process with time lag k .
$R(l, m)$	Correlation function.
$S(k)$	Filtered weight-vector covariance matrix.
$s(k)$	Discrete-time signal process.
$s(t)$	Continuous-time signal process.
subscript p	Model order = p .
subscript q	Model order = q .
$\sup \{ \cdot \}$	Supremum operator.
superscript H	Hermitian transposition.
superscript t	Transposition.
superscript $*$	Complex conjugate.
t	Continuous-time variable.
$w(k)$	System state vector in Chs. 2, 6; noise vector in Ch. 3.
$\underline{x}_q(k)$	$= (x(k), x(k-1), \dots, x(k-q+1))^t$, reference data vector of order q .
$x_r(k)$	Radar data at time k taken on range ring r .
$y(k)$	Observed data being processed.

$\underline{y}_p(k)$	$= (y(k), y(k-1), \dots, y(k-p+1))^t$, observed data vector.
$\underline{y}_N$	Short form for data vector $\underline{y}_N(N)$, i.e.,
$\underline{y}$	$\underline{y}_N \underline{y} = (y(N), y(N-1), \dots, y(1))^t$.
μ_w	Memory factor of an adaptive algorithm.
λ_i	The i^{th} eigenvalue of a square matrix.
η	Detection loss function.
$\phi(k k-1)$	Transition matrix of a system state.
$\underline{\theta}$	Unknown parameter vector of a linear model.
σ_e^2	The variance of a stationary error process $\{e(k)\}$.
$\sigma_e^2(k k-1)$	The variance of prediction-error process $\{e(k k-1)\}$.
$ \cdot $	Absolute value operator.
$\ \cdot\ $	Norm operator.

ABBREVIATIONS

AIC	Akaike information criterion.
AR model	Autoregressive model.
ARMA model	Autoregressive moving-average model.
CAT	Autoregressive transfer function criterion.
CFAR	Constant false alarm rate.
IBDA	Innovations-based detection algorithm.
H_0 -Model	The model used to describe the input signal under hypothesis H_0 .
H_1 -Model	The model used to describe the input signal under hypothesis H_1 .
H_0 -Space	The linear subspace associated with the H_0 -model.
H_1 -Space	The linear subspace associated with the H_1 -model.
LPF	Linear prediction filter.
MDL	Minimum description length criterion.
MTD	Moving target detector.
MTI	Moving target indication.
P_d	Probability of detection.
pdf	Probability density function.
PEF	Prediction-error filter.
P_{fa}	Probability of false alarm.
PRF	Pulse repetition frequency.
SNR	Signal-to-noise ratio.
SCR	Signal-to-clutter ratio.
UMP Test	Uniformly most powerful test.

LIST OF FIGURES.

<u>Figure</u>		<u>Page</u>
3.1	A-Block Diagram of the Adaptive LLR Receiver	57
3.2	Baseband Radar Data Taken from Adjacent Range Rings	68
3.3	Curves of Normalized LLR v.s. Target Doppler Frequency for Data Set I	71
3.4	Curves of Normalized LLR v.s. Target Doppler Frequency for Data Set II	72
3.5	The Effect of Relative Position of Target Peak and Clutter Peak on Normalized LLR: Moderate Separation	74
3.6	The Effect of Relative Position of Target Peak and Clutter Peak on Normalized LLR: The Target is Completely Obscured by Clutter Peak	75
3.7	The Effect of Relative Position of Target Peak and Clutter Peak on Normalized LLR: Large Separation	76
4.1	A Geometric Interpretation of Signal Detection	84
4.2	The Spectra Used for Whitening-Test of H_1 -Model Candidates	87
4.3	The MEM Spectra Used for Whitening-Test of Error Processes in Different Models	92
4.4	The MEM Spectra Used for Whitening-Test	93
4.5	n-Loss Test for H_0 -Model Candidates	95
4.6	Comparison of LLR vs. Target Doppler Curves for Different Schemes	98, 99

<u>Figure</u>		<u>Page</u>
5.1	The Correlation Coefficient (Amplitude) Between the Ground Clutter Samples From Adjacent Range Rings as a Function of The Time Lag for Data Set I	109
5.2	The Correlation Coefficient (Amplitude) Between the Ground Clutter Samples From Adjacent Range Rings as a Function of The Time Lag for Data Set II	110
5.3	The Correlation Coefficient (Amplitude) Between the Weather Clutter Samples From Adjacent Range Rings as a Function of Time Lag for Data Set I	111
5.4	The Correlation Coefficient (Amplitude) Between the Weather Clutter Samples From Adjacent Range Rings as a Function of The Time Lag for Data Set II	112
5.5	Block Diagram of the New Detection Scheme Implementation	118
5.6	A Flowchart of the Data Acquisition and Storage System	125
5.7	Photographs of the TRACS PPI Display	128
5.8	The Amplitude Envelope of the First Scan of the Weather Clutter Dominated Radar Data	129
5.9	The Amplitude Envelope of the First Scan of the Ground Clutter Dominated Radar Data	130
5.10	Comparison of New Detection Scheme with the IBDA Using Actual Radar Data $P_{fa}=10^{-5}$	133
5.11	Testing the New Detection Scheme Using Simulated Data	137
6.1	New Radar Clutter Rejection Schemes	144

<u>Figure</u>		<u>Page</u>
6.2	Power Spectra of the Input and Output Signal of the Clutter Suppression Scheme-I Weather Clutter 'WTRNWI' Range Ring = 25 Azimuth = 2520-2540	155-160
6.3	Power Spectra of the Input and Output Signal of the Clutter Suppression Scheme I Ground Clutter 'WTRNGI' Range Ring = 28 Azimuth = 1500-1520	162-163
6.4	Filter Outputs in Frequency in Response to Ground Clutter	165
6.5	Filter Outputs in Frequency in Response to Weather Clutter	166
6.6	Filter Outputs in Time Domain in Response to Ground Clutter	168
6.7	Filter Outputs in Time Domain in Response to Weather Clutter	171
6.8	Comparison of the Performance of Different Radar Clutter Filters. Results Shown are for Weather Clutter	173
6.9	Comparison of the Performance of Different Radar Clutter Filters. Results Shown are for Ground Clutter	174

CHAPTER 1

INTRODUCTION

1.1 PURPOSE OF THE RESEARCH

This dissertation develops a general procedure for detecting a discrete time-random signal in a background of coloured interference plus white Gaussian noise. In particular, an adaptive radar detection scheme is proposed.

In theory, the problem of detecting a random signal in coloured and/or white Gaussian noise is straightforward if all the statistics of the stochastic processes of interest are completely known. This, however, is not the case for most engineering problems, where the environmental statistics are either unknown or time-varying. Multipath channels in communications and clutter process in surveillance radar systems are two typical examples of such an environment. A useful method that can be employed to solve the detection problem in environment of this kind should have the following desirable characteristics:

- (1) The method has a capability of learning

the environmental statistic adaptively.

(2) It is computationally efficient.

It is the purpose of this study to provide adaptive detection procedures. The emphasis is placed not only on theoretical analysis but on their applications. The procedures are illustrated by designing an adaptive radar detection scheme.

In a related part of the study, we describe two new adaptive radar clutter-suppression schemes. The approach taken here is to suppress the clutter (of unknown statistics), thereby reducing the problem (in effect) to that of detecting a signal in additive white noise.

1.2 TRADITIONAL AND ADAPTIVE SIGNAL DETECTION

Detecting a random signal, not necessarily Gaussian, in additive white Gaussian noise (AWGN) has been studied extensively in the literature; for example, see Kailath [33-34] and Root [64]. The detection problem is described as follows: given the set of observations $\{y(t), 0 \leq t \leq T\}$, determine which of the two following hypotheses is true:

$$(1.1) \quad H_1: y(t) = s(t) + n(t)$$

$$H_0: y(t) = n(t)$$

where $\{n(t)\}$ is white Gaussian noise with zero mean, and $s(t)$ is a random process that is independent of $n(t)$.

According to the innovation theorem [33], under very weak conditions, hypotheses testing (1.1) can be transformed without loss of information into the problem of detecting a conditionally known signal in a background of innovation process $v(t)$ which has exactly the same statistics as the original noise $n(t)$, namely

$$(1.2) \quad H_1: y(t) = \hat{s}(t) + v(t)$$

$$H_0: y(t) = v(t)$$

where $\hat{s}(t)$ is the predicted estimate of signal $s(t)$ given past observations, i.e.,

$$(1.3) \quad \hat{s}(t) = E[s(t) | y(\tau)] \quad \tau \leq t, H_1$$

Hypothesis testing (1.2) resembles that of detecting a known signal in AWGN process. The solution to this problem may be interpreted as an "estimator-correlator". The test statistic is calculated by first estimating the signal process and then using this estimate (as if it were

*According to linear prediction theory, a random sample $y(n)$ can be decomposed into two components. One component is estimated from the past observed data samples, while the other one is unpredictable. Thus, given the past samples, the unpredictable component $a(n)$ contains all the new information in $y(n)$. We refer to $\{a(n)\}$ as the innovation process.

perfect) to compute its inner product with the received signal.

In [64] it is shown that the estimator-correlator receiver can be extended to discrete-time signal detection problem: given observations $(y(1), y(2), \dots, y(N))$, determine which one of the following hypotheses is true:

$$(1.4) \quad H_1: y(k) = s(k) + n(k);$$

$$H_0: y(k) = n(k);$$

where $s(k)$ is a random signal with zero mean and covariance matrix, R_s , and $n(k)$ is zero mean Gaussian noise with covariance matrix, R_n . As before, the solution to this problem has the general inner product structure:

$$(1.5) \quad \lambda(\underline{y}) = \underline{\hat{s}}^t R_n^{-1} \underline{y}$$

where $\underline{\hat{s}} = (\hat{s}(1), \hat{s}(2), \dots, \hat{s}(N))^t$ is no longer a predicted estimate of signal process but redefined according to combined minimum mean-square error criterion [64]. The test statistic (1.5) remains to be interpreted as an estimator-correlator. The estimate of the signal is not causal, being a function of future as well as past observation data; hence, it cannot be evaluated recursively. This makes it unsuitable for engineering problems.

Since structure (1.5) is not easy to implement, we turn back to a predicted estimate again. Consider a general discrete-time signal detection problem:

$$(1.6) \quad H_1: y(k) = s_1(k) + n(k)$$

$$H_0: y(k) = s_0(k) + n(k)$$

where $s_0(k)$ and $s_1(k)$ are random processes with finite power, $n(k)$ is a zero-mean white Gaussian noise. Using a one-step forward prediction estimate and the pertinent innovation process, detection problem (1.6) can be rewritten as follows:

$$(1.7) \quad H_1: y(k) = \hat{s}_1(k) + v_1(k)$$

$$H_0: y(k) = \hat{s}_0(k) + v_0(k) \quad k = 1, 2, \dots, N$$

where $\hat{s}_1(k)$ and $\hat{s}_0(k)$ are the predicted estimates of the signals given the observations $y(1), y(2), \dots, y(k-1)$ under hypotheses H_1 and H_0 , respectively. The innovation processes $\{v_1(k)\}$ and $\{v_0(k)\}$ are defined by

$$(1.8) \quad v_1(k) = s_1(k) - \hat{s}_1(k) + n(k)$$

$$v_0(k) = s_0(k) - \hat{s}_0(k) + n(k)$$

It is shown in [33] that, unlike the case of continuous-time signal, $v_1(k)$ and $v_0(k)$ are no longer Gaussian due to lack of continuity (except, of course, when $s_1(k)$ and $s_0(k)$ are Gaussian). In the following, we will assume that $s_1(k)$ and $s_0(k)$ are Gaussian distributed. This is not a serious limitation since it is shown in [48] that as the sample size tends to infinity, the innovation processes $v_1(k)$ and $v_0(k)$ approach a Gaussian distribution rapidly. Thus, a log likelihood ratio (LLR) receiver based on the Gaussian signal assumption provides a good basis for solving most practical problems. In [33], it is also verified that the variance of $v_1(k)$ and $v_0(k)$ are not equal to that of $n(k)$ and, in general, they are not equal to each other. As a result, the test statistic obtained from LLR does not have a correlator structure.

A convenient way for solving the prediction problem is to parameterize the received signal process under hypotheses H_0 and H_1 in terms of mathematical models. The benefit from doing so is that well developed adaptive algorithms can be employed to recursively and efficiently estimate the unknown model parameters. The likelihood ratio test based on two exactly known autoregressive (AR) models is considered by Dickinson [12]. But his result is not suited to most practical cases where the model coeffi-

7
clients are unknown or time-varying. Porat and Friedlander [60] consider the application of AR and autoregressive moving average (ARMA) modelling to the problem of detecting a Gaussian signal in white Gaussian noise, both with unknown statistics. The basic idea of this second method is that, when H_1 is true, the LLR normalized by the sample size asymptotically approaches the signal-to-noise ratio (SNR), and the SNR is very easy to express in terms of the model parameters. Thus, the SNR is chosen to be the test statistic and the required unknown model parameters of signal and noise component are estimated by using a parametric spectral estimation technique [6, 15]. The philosophy of this method is elegant, but it suffers from two limitations:

- (1) It is used for solving a random signal-in-white Gaussian noise problem.
- (2) The spectral matching technique used for estimating the unknown parameters is computationally complicated.

To overcome these drawbacks, Metford and Haykin [49] describe an adaptive detection algorithm known as the innovations-based detection algorithm (IBDA). IBDA assumes the use of AR model for the signals under both hypotheses H_1 and H_0 , and it is observed [27, 49] that

the model for the received signal under H_1 , in general, has a longer model order than that under H_0 . The model orders are empirically determined and the unknown model parameters are recursively estimated using a Kalman filter or recursive least-squares algorithm. The LLR test is established in terms of the innovation processes resulting from the adaptive filters. This method has been successfully applied to solve the radar detection problem using real data.

In IBDA, the models used for fitting the signals under the two hypotheses are chosen from the AR family. This implies that the characteristic used for discriminating two random processes is the difference in the model order. This form of discrimination, however, limits the detection performance of IBDA.

Furthermore, in IBDA it is not realized that the LLR used is a generalized one and thus, like other model parameters, the unknown variance of observation model error can be estimated from input data. Rather, great effort is made to establish a prediction-error-power map (called clutter map in the radar detection context) for the hypotheses H_0 and H_1 ; consequently, a great deal of memory is required. In a nonstationary environment, the prediction-error-power map for hypothesis H_0 does help a lot for improving the receiver performance; but under

H_1 , it varies with SNR and should therefore be estimated from the input sequence.

Generalized LLR based on linear models does not lead to a uniformly most powerful test (UMP) [38, 44].

Consequently, the use of different modelling techniques gives rise to a different parameter estimator and different receiver performance.

Modelling the received signal for a detection problem is quite different from and more difficult than the time series-model-fitting problem. For the latter, the data are given and an appropriate type of model is assumed to be known; the only problem left is to estimate the model parameters. For the detection problem, however, only one set of input data, under hypothesis H_0 or H_1 , is available and it must be used to estimate two sets of model parameters. When the received signal under H_1 is used, the adaptive algorithm will give correct estimates for H_1 -model parameters since (for this time) the model is matched to the input signal; but the estimates for H_0 -model parameters may be wrong since the model and input sequence are not matched. Accordingly, an ideal H_0 -model should be designed such that the estimates of the model parameters are dependent only on the noise component and independent of the signal component. When hypothesis H_0 is true, the H_0 -model is matched to the input signal and the estimates of H_0 -model parameters

are correct, but the estimates for H_1 -model parameters are probably incorrect. A linear model may also be viewed as a linear filter. In addition to the requirements discussed above, the filters corresponding to models H_0 and H_1 should be able to whiten the received sequence under hypotheses H_0 and H_1 , respectively. Thus, an ideal signal model performs the following functions:

(1) It correctly estimates the model parameters, irrespective of whether the input corresponds to hypothesis H_0 or H_1 .

(2) It whitens the input signal under the pertinent hypothesis.

For practical problems, it is difficult to find such ideal models; however, we should bear in mind that the major aim of mathematical analysis is to gain insight into the approach to the solution. Such insight can then be used to intelligently modify and adapt the mathematical solution to a particular physical situation. Given a practical detection environment, we should make full use of the information, a priori and a posteriori, as much as possible, to design a pair of good models so that a satisfactory detection performance is achieved.

After determining the models for the two hypotheses H_0 and H_1 , two adaptive filters may be used to

recursively estimate the unknown parameters and produce the corresponding error processes. These outputs are then applied to calculate the LLR.

The contribution of this thesis is the development of a general technique, based on linear modelling for detecting a random signal in coloured plus white Gaussian noise, all having unknown statistics. The procedure consists of the following steps:

- a) Choosing model candidates.
- b) Testing these candidates to identify the kind of models that are worthy of further investigation.
- c) Estimating all the parameters of the tentative models.
- d) Constructing an LLR receiver based on these models.

An adaptive algorithm is developed to recursively and efficiently calculate the LLR. To illustrate the use of this technique, the procedures are used to construct an adaptive radar detection scheme for detecting a moving target in a surveillance radar environment.

1.3 ORGANIZATION OF THE THESIS

The thesis is organized as follows: In chapter 2, we discuss the use of a Kalman filter, assuming random

walk model as an adaptive whitening filter, and consider its application to a nonstationary signal process. The maximum-likelihood principle is employed to specify some unknown parameters of the filter. In chapter 3, we discuss the generalized LLR based on linear modelling and its major properties. Geometrical interpretation of the LLR and an efficient algorithm are presented also. Chapter 4 provides some methods for linear model building for solving the detection problem. We discuss the application of this technique to radar detection in chapter 5. In chapter 6, two new adaptive radar clutter suppression schemes are presented. Conclusions and some suggestions for further research are given in chapter 7.

CHAPTER 2

ADAPTIVE WHITENING FILTER FOR NONSTATIONARY PROCESSES

2.1 LLR TEST AND WHITENING FILTER

The calculation of LLR for a binary detection problem requires knowledge of the likelihood functions under the two hypotheses H_0 and H_1 . The likelihood function is essentially the probability density function (pdf) of the observations, with the parameter vector treated as the unknown. Thus, to determine the LLR, it is necessary to know the pdfs of the observed data. Now consider the following problem: Given observations $y(1), y(2), \dots, y(N)$ generated by a zero-mean Gaussian process, find their joint pdf. This problem becomes trivial if the covariance matrix is known a priori. When the covariance matrix is not available, we can first estimate it and then use it to form the pdf. This brute force evaluation misses the point and is inefficient in computation. A more effective method is to proceed as follows: Let $\underline{y}_N$ denote the vector of the observations available at time N and let $\underline{\theta}$ be the vector of unknown parameters of the random

process of interest. The multiplication rule for conditional probability gives

$$(2.1) \quad p(\underline{y}_N | \underline{\theta}) = p[y(N) | \underline{y}_{N-1}, \underline{\theta}] p(\underline{y}_{N-1} | \underline{\theta})$$

Repeated use of the multiplication rule yields the likelihood function $L(\underline{\theta}; \underline{y}_N)$.

$$(2.2) \quad L(\underline{\theta}; \underline{y}_N) = p[y(N) | \underline{y}_{N-1}] \cdots p[y(2) | y(1)] p[y(1)]$$

where $\underline{\theta}$ has been suppressed to simplify the notation.

Equation (2.2) indicates that for sequential observations the likelihood function can be written as a product of conditional pdfs. Assume that the conditional densities are Gaussian, as shown by

$$(2.3) \quad p[y(k) | \underline{y}_{k-1}] = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left[-\frac{1}{2\sigma_k^2} (y(k) - \hat{y}(k))^2\right]$$

where the predicted estimate

$$\hat{y}(k) = E[y(k) | \underline{y}_{k-1}]$$

and the variance

$$\sigma_k^2 = \text{cov}[y(k) | \underline{y}_{k-1}]$$

The logarithm of the likelihood function is written as

$$(2.4) \quad -\log L(\underline{\theta}; Y_N) = -\frac{1}{2} \sum_{k=1}^N \left[\frac{|e(k)|^2}{\sigma_k^2} + \log \sigma_k^2 \right] + \frac{N}{2} \log 2\pi$$

where the prediction error equals

$$e(k) = \begin{cases} y(1), & k = 1 \\ y(k) - \hat{y}(k), & k = 2, 3, \dots, N \end{cases}$$

From previous discussion, the LLR can be easily set up by decorrelating the input sequence through the use of a whitening filter. By using this filter, the original signal process is transformed, without loss of information, into an innovation process. The whitening filter performs a function analogous to the the Gram-Schmidt orthogonalization procedure in mathematics. The predicted estimate is causal since it depends only on past observations, and hence can readily be implemented by a recursive algorithm.

In order to analyze and design a whitening filter, a key step is to represent the input signal in terms of some kind of prediction models. There are many possible model structures appropriate for this purpose; among them, the AR model family and "state-space"

approach are two most popular methods in engineering applications.

An AR model for fitting a given time series is obtained by first identifying an appropriate subclass of models for the general AR model family and then determining the model coefficients employing the maximum-likelihood principle [Box and Jenkins, 9]. The model coefficients define a finite impulse response (FIR) filter, and this filter can then be used to orthogonalize the input sequence. This method is suitable for a stationary signal process. To extend its application to nonstationary environments, a memory factor (in the form of decaying exponential) is introduced into the definition of the index of performance to be minimized. In doing so, data in the distant past tend to be ignored, and the filter tap-weights will be updated mainly depending on the newly received data. This method makes use of the concept of local stationarity, i.e., for a short period of time, a nonstationary process can be treated as a stationary one [68].

The first successful approach to systematically solve the prediction and whitening problem of nonstationary processes was made in 1960 by Kalman [36]. According to Kalman, a nonstationary signal process can be expressed in terms of the "state-space" approach. The state is described by the process equation:

$$(2.5) \quad \underline{w}(k+1) = \underline{\phi}(k+1, k) \underline{w}(k) + \Delta \underline{w}(k)$$

where $\underline{w}(k)$ is an M -by-1 state vector at time k , $\underline{\phi}(k+1, k)$ the M -by- M state transition matrix, and $\Delta \underline{w}(k)$ is M -by-1 process noise vector with zero-mean values and covariance matrix $\underline{Q}$. The observed output of the system is related to state variable by

$$(2.6) \quad y(k) = \underline{c}(k) \underline{w}(k) + e(k)$$

where $\underline{c}(k)$ is 1-by- M measurement vector, and the measurement error process $\{e(k)\}$ is assumed to be an independent and Gaussian process with zero-mean and covariance σ_0^2 that is independent of $\{\Delta \underline{w}(k)\}$.

For many practical situations, the measurement matrix can be determined through some kind of physical and/or geometrical interpretation. The choice of $\underline{c}(k)$ is particularly important to a detection problem and therefore, will be discussed in length in chapters 4 and 5. The transition matrix $\underline{\phi}(k+1, k)$ is generally unknown and must be estimated from the observations. It is conceptually simple but its formulation is usually quite complicated. Accordingly, it is customary to assume that it equals the identity matrix. To account for nonstationarity,

we may assume that the state vector executes a random walk from one iteration to the next. In particular, the system is modelled as

$$(2.7a) \quad \underline{w}(k+1) = \underline{w}(k) + \Delta \underline{w}(k)$$

$$(2.7b) \quad y(k) = \underline{c}(k) \underline{w}(k) + e(k)$$

This model is well suited for slowly varying random signals and has been successfully used in communications [30, 53], radar detection [49, 78, 80], system tracking and other areas [45, 52, 79]: Apparently, signal model (2.7) is more general than an autoregression model and includes an AR model as a special case.

2.2 DETERMINATION OF THE FILTER PARAMETERS

Let

$$(2.8a) \quad E|e(k)|^2 = r$$

$$(2.8b) \quad E[\Delta \underline{w}(k) \Delta \underline{w}^H(k)] = \underline{Q}$$

where H denotes Hermitian transposition. The tap-weight vector can be recursively calculated by applying the Kalman filtering algorithm [29, 32, 74] to the linear

model of (2.7), yielding:

$$(2.9a) \quad \underline{g}(k) = \underline{P}(k) \underline{c}^H(k) (r + \underline{c}(k) \underline{P}(k) \underline{c}^H(k))^{-1}$$

$$(2.9b) \quad \underline{S}(k) = \underline{P}(k) - \underline{g}(k) \underline{c}(k) \underline{P}(k)$$

$$(2.9c) \quad \underline{P}(k+1) = \underline{S}(k) + \underline{Q}$$

$$(2.9d) \quad \underline{e}(k|k-1) = \hat{y}(k) - \underline{c}(k) \underline{w}(k-1)$$

$$(2.9e) \quad \underline{w}(k) = \underline{w}(k-1) + \underline{g}(k) \underline{e}(k|k-1)$$

where $\underline{g}(k)$ = Kalman gain at k^{th} step.

$\underline{P}(k)$ = predicted weight-error covariance matrix.

$\underline{S}(k)$ = filtered weight-error covariance matrix.

If we normalize the covariance matrices by dividing them by r , the variance of the measurement error, then equations (2.9a) to (2.9c) may be rewritten as follows:

$$(2.10a) \quad \underline{g}(k) = \left(\frac{1}{r} \underline{P}(k) \underline{c}^H(k) \left(1 + \underline{c}(k) \left(\frac{1}{r} \underline{P}(k) \right) \underline{c}^H(k) \right)^{-1} \right)$$

$$(2.10b) \quad \underline{S}(k)/r = \underline{P}(k)/r - \underline{g}(k) \underline{c}(k) \underline{P}(k)/r$$

$$(2.10c) \quad \underline{P}(k+1)/r = \underline{S}(k)/r + \underline{Q}/r$$

We note from these equations that the utilization of the normalized covariance matrices does not change the Kalman gain and therefore the recursive

estimate of $\underline{w}(k)$. For convenience of notation, we continue to denote these covariance matrices by $\underline{P}(k)$, $\underline{S}(k)$, $\underline{Q}$, respectively; thus, the adaptation algorithm (2.9) becomes:

$$(2.11a) \quad \underline{v}(k) = \underline{P}(k) \underline{c}^H(k)$$

$$(2.11b) \quad \underline{g}(k) = \underline{v}(k) / (1 + \underline{c}(k) \underline{v}(k))$$

$$(2.11c) \quad \underline{S}(k) = \underline{P}(k) - \underline{g}(k) \underline{v}^H(k)$$

$$(2.11d) \quad \underline{P}(k+1) = \underline{S}(k) + \underline{Q}$$

$$(2.11e) \quad e(k|k-1) = y(k) - \underline{c}(k) \underline{w}(k-1)$$

$$(2.11f) \quad \underline{w}(k) = \underline{w}(k-1) + \underline{g}(k) e(k|k-1)$$

Although the Kalman algorithm provides the solution to the tap-weight vector, it cannot be used to solve for the covariance matrix $\underline{Q}$. The Kalman filter transforms the original sequence $\{y(k)\}$ to an innovation process $\{e(k|k-1)\}$. Under the Gaussian assumption of $\{y(k)\}$, $\{e(k|k-1)\}$ is also Gaussian with zero-mean. It should be noted, however, that variance of the innovation process is no longer equal to that of $\{e(k)\}$. According to Kalman theory, we have [43, 29]

$$(2.12) \quad E|e(k|k-1)|^2 \triangleq \sigma_{k,k-1}^2 = \sigma_e^2 (1 + \underline{c}(k) \underline{P}(k) \underline{c}^H(k))$$

where

$$(2.13) \quad \sigma_e^2 = E\{|e(k)|^2\}$$

Using the same argument as in Section 2.1, the log likelihood function becomes

$$(2.14) \quad \log L(\underline{w}, \underline{Q}, \sigma_e^2; \underline{y}_N) = -\frac{1}{2} \sum_{k=1}^N \left[\frac{|e(k|k-1)|^2}{\sigma_{k,k-1}^2} + \log \sigma_{k,k-1}^2 \right] - \frac{n}{2} \log 2\pi$$

Clearly, maximum-likelihood estimates of $\underline{Q}$ and σ_e^2 are those values that maximize the log function given in (2.14).

For many practical environments, covariance $\underline{Q}$ changes slowly with time. Thus, we may use the following strategy. Estimate $\underline{Q}$ during a training period, and then use it as if it were perfect to calculate the LLR during the operating mode.

2.3

EXPLICIT SOLUTION FOR THE TAP-WEIGHT VECTOR

Now we seek an explicit expression for the tap-weight vector $\underline{w}(k)$. Substituting equation (2.11a) into (2.11b) to eliminate $\underline{v}(k)$ and rearranging it, we have

$$(2.15) \quad \{ [\underline{I} - \underline{q}(k) \underline{c}(k)] \underline{p}(k) \} \underline{c}(k)^H = \underline{q}(k)$$

But from equation (2.11c)

$$(2.16) \quad [\underline{I} - \underline{q}(k) \underline{c}(k)] \underline{p}(k) = \underline{s}(k)$$

Thus, we obtain

$$(2.17) \quad \underline{q}(k) = \underline{s}(k) \underline{c}(k)^H$$

Substituting $\underline{q}(k)$ from equation (2.17) and $\underline{v}(k)$ from (2.11a) into (2.11c), we obtain

$$(2.18) \quad \underline{s}^{-1}(k) = \underline{p}^{-1}(k) + \underline{c}(k)^H \underline{c}(k)$$

Eliminating $\underline{p}(k)$ by using equation (2.11d)

$$(2.19) \quad \underline{s}^{-1}(k) = \underline{p}(k-1) \underline{s}^{-1}(k-1) + \underline{c}(k)^H \underline{c}(k)$$

where

$$(2.20) \quad \underline{D}(k-1) = (\underline{I} + \underline{S}^{-1}(k-1)\underline{Q})^{-1}$$

Equation (2.19) is a first-order difference equation in $\underline{S}^{-1}(k)$ and can be solved by using an iterative approach, as shown by

$$(2.21) \quad \underline{S}^{-1}(k) = \underline{D}(k-1) \dots \underline{D}(0) \underline{S}^{-1}(0) + \underline{D}(k-1) \dots \underline{D}(1) \underline{c}(1)^H \underline{c}(1) \\ + \dots + \underline{D}(k-1) \underline{c}(k-1)^H \underline{c}(k-1) + \underline{c}(k)^H \underline{c}(k)$$

Generally, we have no a priori information about the optimum tap-weight vector and $\underline{S}(0)$ is assumed to be a diagonal matrix with large positive elements. As such, the first term on the right side of (2.21) may be ignored from the summation and so we have

$$(2.22) \quad \underline{S}(k) = \left(\sum_{\ell=1}^k \underline{F}(\ell) \underline{c}(\ell)^H \underline{c}(\ell) \right)^{-1}$$

where $\underline{F}(\ell)$ is a forgetting factor and defined as

$$(2.23) \quad \underline{F}(\ell) = \underline{D}(k-1) \underline{D}(k-2) \dots \underline{D}(\ell) \quad 1 \leq \ell \leq k-1 \\ = \underline{I} \quad \ell = k$$

Now we try to find a recursive equation for $\underline{w}(k)$. Substituting $\underline{q}(k)$ from equation (2.17), and $\underline{e}(k|k-1)$ from

(2.11e) into equation (2.11f), we have

$$(2.24) \quad \underline{\hat{w}}(k) = \underline{\hat{w}}(k-1) + \underline{S}(k) \underline{c}(k)^H \underline{y}(k) - \underline{S}(k) \underline{c}(k)^H \underline{c}(k) \underline{\hat{w}}(k-1)$$

Then eliminating $\underline{c}(k)^H \underline{c}(k)$ in the third term on the right side of equation (2.24), by using equation (19), we get

$$(2.25) \quad \underline{S}^{-1}(k) \underline{\hat{w}}(k) = \underline{D}(k-1) \underline{S}^{-1}(k-1) \underline{\hat{w}}(k-1) + \underline{c}(k)^H \underline{y}(k)$$

Using iterative method again to solve for $\underline{S}^{-1}(k) \underline{\hat{w}}(k)$, we obtain

$$(2.26) \quad \underline{S}^{-1}(k) \underline{\hat{w}}(k) = \underline{F}(0) \underline{S}^{-1}(0) \underline{\hat{w}}(0) + \underline{F}(1) \underline{c}(1)^H \underline{y}(1) + \dots + \underline{F}(k-1) \underline{c}(k-1)^H \underline{y}(k-1) + \underline{c}(k)^H \underline{y}(k)$$

It is customary to assume that $\underline{\hat{w}}(0)$ equals the null vector. Thus equation (2.26) reduces to

$$(2.27) \quad \underline{\hat{w}}(k) = \underline{S}^{-1}(k) \sum_{\ell=1}^k \underline{F}(\ell) \underline{c}(\ell)^H \underline{y}(\ell)$$

Combining equations (2.22) and (2.27), we get

$$(2.28) \quad \underline{\hat{w}}(k) = \left(\sum_{\ell=1}^k \underline{F}(\ell) \underline{c}(\ell)^H \underline{c}(\ell) \right)^{-1} \left(\sum_{\ell=1}^k \underline{F}(\ell) \underline{c}(\ell)^H \underline{y}(\ell) \right)$$

It is obvious from equation (2.28) that the Kalman algorithm assuming random walk model leads to a Wiener solution with operators as weighting factors.

$\underline{c}(k)$ is the input vector and $y(k)$ is the desired response. The definition of $\underline{c}(k)$ depends on the purpose of the filter and the environment in which the algorithm operates. For instance, when used as a prediction-error filter, $\underline{c}(k)$ is a 1-by-M vector defined by

$$\underline{c}(k) = (y(k-M), \dots, y(k-M-1), \dots, y(k-1)).$$

When used as a noise canceller, however, $\underline{c}(k)$ is a 1-by-M vector with elements from reference input $\{x(k)\}$:

$$\underline{c}(k) = (x(k-M-1), \dots, x(k)).$$

2.4 ANALYSIS OF THE ALGORITHM

For convenience of analysis, we rewrite the Wiener solution as

$$(2.29) \quad \underline{w}(k) = \left(\sum_{\ell=1}^k \underline{F}(\ell) \underline{c}^H(\ell) \underline{c}(\ell) \right)^{-1} \left(\sum_{\ell=1}^k \underline{F}(\ell) \underline{c}^H(\ell) y(\ell) \right)$$

where

$$(2.30) \quad \underline{F}(\ell) = \underline{D}(k-1) \underline{D}(k-2) \dots \underline{D}(\ell) \quad 1 \leq \ell \leq k-1 \\ = \underline{I} \quad \ell = k$$

$$(2.31) \quad \underline{D}(k-1) = (\underline{I} + \underline{S}^{-1}(k-1)\underline{Q}(k-1))^{-1}$$

This solution has the following characteristics:

- (1) When $\underline{Q}(k) = \underline{0}$ for all k , it reduces to the Wiener solution for a stationary environment.

Proof. At this time, $\underline{F}(\ell) = \underline{I}$ $1 \leq \ell \leq k$

Thus,

$$(2.32) \quad \underline{w}(k) = -\left(\sum_{\ell=1}^k \underline{c}^H(\ell)\underline{c}(\ell)\right)^{-1} \left(\sum_{\ell=1}^k \underline{c}^H(\ell)y(\ell)\right)$$

- (2) The weighted recursive least-squared (RLS) solution is a special case of equation (2.29) when $\underline{Q}(\ell) \equiv \mu \underline{S}(\ell)$ $\mu > 0$

Proof. Since $\underline{Q}(\ell) = \mu \underline{S}(\ell)$

we have $\underline{D}(\ell) = (\underline{I} + \underline{Q}(\ell)\underline{S}^{-1}(\ell))^{-1} = \left(\frac{1}{1+\mu}\right)\underline{I}$

and $\underline{F}(\ell) = \left(\frac{1}{1+\mu}\right)^{k-\ell}\underline{I}$

Thus

$$(2.33) \quad \underline{w}(k) = \left(\sum_{\ell=1}^k \left(\frac{1}{1+\mu}\right)^{k-\ell} \underline{c}^H(\ell)\underline{c}(\ell)\right)^{-1} \left(\sum_{\ell=1}^k \left(\frac{1}{1+\mu}\right)^{k-\ell} \underline{c}^H(\ell)y(\ell)\right)$$

(3) The algorithm given in equation (2.11)

will result in an operator-weighted

Wiener solution if the matrix $\underline{Q}(k)$ satisfies the following condition:

$$\underline{Q}(k) = \text{diag}(q, q, \dots, q)$$

where $q \geq 0$. This implies that in some

situations the algorithm has the capability

of forgetting data in the distant past.

Proof. In Appendix 2A it is shown that under the constraint $\underline{Q}(k) = \text{diag}(q, \dots, q)$, $q \geq 0$, the norm of the operator $\underline{D}(k)$ satisfies the following condition:

$$(2.34) \quad \|\underline{D}(k)\| = \frac{1}{1 + q/\lambda_{k, \max}} \leq 1 \text{ for all } k$$

where $\lambda_{k, \max} > 0$ is the largest eigenvalue of the positive definite matrix $\underline{S}(k)$. Accordingly, we have

$$(2.35) \quad \begin{aligned} \|\underline{F}(k)\| &= \|\underline{D}(k-1)\underline{D}(k-2)\dots\underline{D}(k)\| \\ &\leq \|\underline{D}(k-1)\| \dots \|\underline{D}(k)\| \\ &\leq 1 \end{aligned}$$

$$\begin{aligned} \|\underline{F}(k)\| &= \|\underline{D}(k-1)\dots\underline{D}(k+1)\underline{D}(k)\| \\ &\leq \|\underline{F}(k-1)\| \|\underline{D}(k)\| \end{aligned}$$

That is

$$(2.36) \quad \|F(k)\| \leq \|F(k-1)\|$$

We conclude from equations (2.35) and (2.36) that: When $q \neq 0$, all the "weights" for the past data are less than one. Furthermore, the more distant the past data, the smaller the weight. Let

$$\omega = \sup_k [\|D(k)\|]$$

Then, if $q \neq 0$, we find that $\omega < 1$; thus

$$(2.37) \quad \|F(k) C^H(k) C(k)\| \leq \omega^{k-l} \|C^H(l) C(l)\|$$

$$(2.38) \quad \|F(k) C^H(k) Y(k)\| \leq \omega^{k-l} \|C^H(l) Y(l)\|$$

These norms approach zero as k tends to ∞ .

APPENDIX 2A

Lemma. Let $\underline{S}$ be a Hermitian positive-definite matrix with $\lambda_{\max}$ as its largest eigenvalue, and q be an arbitrary positive number; then,

$$(A-1) \quad \|(\underline{I} + q\underline{S}^{-1})^{-1}\| = (1 + \frac{q}{\lambda_{\max}})^{-1} < 1$$

where the norm of the matrix $\underline{S}$ is defined [54] as

$$\|\underline{S}\| = \text{maximum eigenvalue of } (\underline{S}^H \underline{S})^{1/2}$$

Proof. Since $\underline{S}$ is a positive-definite Hermitian matrix, there must be a unitary matrix $\underline{A}$ such that

$$(A-2) \quad \underline{A}^H \underline{S} \underline{A} = \underline{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

where $\lambda_i > 0$ are eigenvalues of $\underline{S}$. From (A-2), we have

$$\underline{S}^{-1} = \underline{A} \underline{\Lambda}^{-1} \underline{A}^H$$

Accordingly,

$$\begin{aligned} (\underline{I} + q\underline{S}^{-1})^{-1} &= (\underline{I} + q\underline{A} \underline{\Lambda}^{-1} \underline{A}^H)^{-1} \\ &= \underline{A} (\underline{I} + q\underline{\Lambda}^{-1})^{-1} \underline{A}^H \\ &= \underline{A} \text{diag}(\frac{1}{1+q/\lambda_1}, \dots, \frac{1}{1+q/\lambda_n}) \underline{A}^H \end{aligned}$$

Hence

$$\|(\underline{I} + q\underline{S}^{-1})^{-1}\| = \frac{1}{1 + q/\lambda_{\max}} < 1$$

CHAPTER 3

LOG LIKELIHOOD RATIO (LLR)

RECEIVERS BASED ON LINEAR MODELLING

3.1 INTRODUCTION

Detecting a random signal in the presence of coloured noise with unknown statistics is widely encountered in communications, radar and other disciplines. Traditionally, there are two ways to tackle this problem. The first one is to estimate the correlation function of the coloured noise and the waveform of the signal, and then construct a correlator-receiver. This type of receiver has been well developed for detecting a random signal in white Gaussian noise [33]. However, this approach is of limited utility when the background consists of correlated noise with unknown statistics.

In the second method, the unwanted noise is suppressed by using a noise canceller before the received signal enters an envelope detector.

A new method is developed in this thesis. It first parameterizes the received signal under hypotheses H_0 and H_1 by using two different models. Then, an adaptive

algorithm is developed to perform the following tasks:

1) Estimation of the unknown statistics of the noise and the signal, or equivalently, the model parameters and the variances of the corresponding error processes.

2) Transformation of the correlated received signal into innovation process.

3) Recursive calculation of the log likelihood ratio (LLR). These features make this method highly suitable for real-time signal processing.

In this chapter we develop an adaptive detection algorithm; and the modelling problem is considered in chapter 4.

In a certain class of detection problems, reference inputs are available to the designer. For radar detection in weather or ground clutter dominated environment, radar data from adjacent range rings are correlated with the clutter component of the data under test and this can be used as reference inputs [78].

We denote the reference inputs by $x(1), x(2), \dots, x(n)$ and the received data under test by $y(1), y(2), \dots, y(n)$. $y(k)$ is composed of noise $n(k)$, when H_0 is true and composed of noise $n(k)$ plus signal $s(k)$ when H_1 is true; that is,

$$\begin{aligned} (3.1) \quad H_0: y(k) &= n(k) \\ H_1: y(k) &= s(k) + n(k) \end{aligned}$$

It is assumed that signal $s(k)$ is uncorrelated with $n(k)$ and reference $x(k)$, but $x(k)$ is correlated with $n(k)$. Using these assumptions and regressive and auto-

regressive model theory [14, 67, 9], it is

possible to describe the H_0 and H_1 hypotheses by two pairs of models:

1) The first pair is

$$(3.2) \quad H_0: y(k) = \underline{b}_q^t \underline{x}(k) + e_0(k)$$

$$H_1: y(k) = \underline{a}_p^t \underline{y}_p(k-1) + e_1(k)$$

where $\underline{a}_p = (a_1, a_2, \dots, a_p)^t$

$$\underline{b}_q = (b_1, b_2, \dots, b_q)^t$$

$$\underline{y}_p(k-1) = (y(k-1), y(k-2), \dots, y(k-p))^t$$

$$\underline{x}(k) = (x(k), x(k-1), \dots, x(k-q+1))^t$$

The error processes $\{e_0(k)\}$ and $\{e_1(k)\}$ are Gaussian distributed with zero-mean and variances

$$E(|e_0(k)|^2) = \sigma_0^2$$

$$E(|e_1(k)|^2) = \sigma_1^2$$

Furthermore, $\{e_0(k)\}$ and $\{e_1(k)\}$ are assumed to be mutually independent, i.e.,

$$E|e_0(k)e_1^*(k)| = E|e_1(k)e_0^*(k)| = 0.$$

Equation (3.2) models the received signal under H_0 by a regressive (RE) model and the received signal under H_1 by an autoregressive (AR) model. We thus refer to this method as REAR modelling.

2). The received signal under hypotheses H_0 and H_1 is modelled by two AR processes of different orders. For brevity, we refer to this second approach as ARAR modelling. ARAR modelling was used by Metford and Haykin [49]. In this case we have:

$$H_0: y(k) = \underline{b}_q^t \underline{y}_q(k-1) + e_0(k) \quad (3.3)$$

$$H_1: y(k) = \underline{a}_p^t \underline{y}_p(k-1) + e_1(k)$$

where $\{e_0(k)\}$ and $\{e_1(k)\}$ are zero-mean Gaussian processes with

$$E(|e_0(k)|^2) = \sigma_0^2$$

$$E(|e_1(k)|^2) = \sigma_1^2$$

$\{e_0(k)\}$ and $\{e_1(k)\}$ are mutually independent. $\underline{a}_p$ and $\underline{b}_q$ are model coefficient vectors defined by

$$\underline{a}_p = (a_1, a_2, \dots, a_p)^t$$

$$\underline{b}_q = (b_1, b_2, \dots, b_q)^t.$$

The data to be treated are assumed to be complex valued. In many situations, radar detection for instance,

the environment is nonstationary so that $\underline{b}_q$ and σ_0^2 will change spatially and temporarily. $\underline{a}_p$ and σ_1^2 are also varying not only due to nonstationarity of the background but also due to changes in signal-to-noise ratio and the signal waveform. However, in some cases, radar detection for example, σ_0^2 for each range-azimuth cell changes slowly from scan to scan. In order to enhance detection, we may store σ_0^2 for each cell and update it whenever hypothesis H_0 is made. In the following discussion, we assume that σ_0^2 is a known parameter. Thus, we may rewrite REAR- and ARAR-modelling as follows:

REAR Modelling

$$(3.4) \quad H_0: \quad y(k) = \underline{b}_q^t x(k) + e_0(k)$$

$$e_0(k) \sim N(0, E[|e_0(k)|^2])$$

$$E[|e_0(k)|^2] = \sigma_0^2$$

$$H_1: \quad y(k) = \underline{a}_p^t y_p(k-1) + e_1(k)$$

$$e_1(k) \sim N(0, E[|e_1(k)|^2])$$

ARAR-Modelling

$$(3.5) \quad H_0: y(k) = \underline{b}_q^t y(k-1) + e_0(k)$$

$$e_0(k) \sim N(0, E[|e_0(k)|^2])$$

$$E[|e_0(k)|^2] = \sigma_0^2$$

$$H_1: y(k) = \underline{a}_p^t y_p(k-1) + e_1(k)$$

$$e_1(k) \sim N(0, E[|e_1(k)|^2])$$

$\underline{a}_p$, $\underline{b}_q$ and σ_1^2 are unknown parameters and have to be estimated before the LLR can be computed.

3.2 LLR TEST

We first develop LLR test for the REAR modelling approach. The generalized LLR is defined as

$$(3.6) \quad \Lambda(\underline{y}) = \frac{\max_{\underline{a}_p, \sigma_1^2} \{p(\underline{y}|H_1)\}}{\max_{\underline{b}_q} \{p(\underline{y}|H_0)\}} = \frac{p(\underline{y}|\hat{\underline{a}}_p, \hat{\sigma}_1^2)}{p(\underline{y}|\hat{\underline{b}}_q, \hat{\sigma}_0^2)}$$

where $\underline{y}$ denotes the vector composed of the given samples $y(1), y(2), \dots, y(n)$, and $\hat{\underline{a}}_p, \hat{\sigma}_1^2$ are ML estimates assuming H_1 is true, while $\hat{\underline{b}}_q$ is the ML estimate, assuming that H_0 is true.

We will calculate the numerator and denominator of eq. (3.6), respectively.

3.2.1 Maximum Likelihood Function Under H_1

By using the linear model of (3.4b), the conditional probability density function under H_1 can be expressed as

$$(3.7) \quad p(y(1), \dots, y(n) | \underline{a}_p, \sigma_1^2) = p(y(1), y(2), \dots, y(p) | \underline{a}_p, \sigma_1^2) \cdot p(y(p+1), \dots, y(n) | \underline{y}_p, \underline{a}_p, \sigma_1^2).$$

$$= (2\pi\sigma_1^2)^{-n/2} (\det \underline{M}_p)^{1/2} \exp \left[-\frac{1}{2\sigma_1^2} \left(\sum_{k=p+1}^n |e(k)|^2 + \underline{y}_p^H \underline{M}_p^{-1} \underline{y}_p \right) \right]$$

$$= (2\pi\sigma_1^2)^{-n/2} (\det \underline{M}_p)^{1/2} \exp \left(-\frac{1}{2\sigma_1^2} \underline{a}_{p+1}^H \underline{R}_{p+1} \underline{a}_{p+1} \right)$$

where $\sigma_1^2 \underline{M}_p^{-1}$ is the autocovariance matrix of vector,

$\underline{y}_p = (y(p), \dots, y(1))^t$ and the augmented matrices

$\underline{a}_{p+1}$ and $\underline{R}_{p+1}$ are defined by

$$(3.8a) \quad \underline{a}_{p+1} = \underline{a} (1, a_1, \dots, a_p)^t.$$

$$(3.8b) \quad \underline{R}_{p+1} \triangleq \begin{bmatrix} R(0,0) & -R(0,1) & -R(0,2) & \dots & -R(0,p) \\ -R(1,0) & R(1,1) & R(1,2) & \dots & R(1,p) \\ -R(p,0) & R(p,1) & R(p,2) & \dots & R(p,p) \end{bmatrix}$$

where

$$(3.8c) \quad R(l,m) = \sum_{t=1}^{n-l-m} y(t+l)y^*(t+m)$$

Theoretically, it is straightforward to obtain $\hat{a}_p$ and $\hat{\sigma}_1^2$ by differentiating eq. (3.7) with respect to the unknown parameters, setting the results equal to zero and solving for the ML estimates. In practice, however, this approach is complicated and we have to resort to the use of approximations. Three approaches can be used [9]:

(1) Least Squares Estimate

$$(3.9) \quad \hat{a}_p = \begin{bmatrix} R(1,1) & R(1,2) & \dots & R(1,p) \\ R(2,1) & R(2,2) & \dots & R(2,p) \\ \vdots & \vdots & \ddots & \vdots \\ R(p,1) & R(p,2) & \dots & R(p,p) \end{bmatrix}^{-1} \begin{bmatrix} R(0,1) \\ R(0,2) \\ \vdots \\ R(0,p) \end{bmatrix}$$

where

$$(3.10) \quad R(l,m) = \sum_{t=1}^{n-l-m} y(t+l)y^*(t+m)$$

(2) The estimate in (3.9) becomes an approximate ML one if $R(l, m)$ in (3.10) is replaced by

$$(3.11) \quad R(l, m) = \frac{n}{n-l-m} \sum_{t=1}^{n-l-m} y(t+l) y^*(t+m)$$

(3) Yule-Walker Estimate

$$(3.12) \quad \hat{a}_p = \begin{pmatrix} R(0) & R(-1) & \dots & R(-p+1) \\ R(1) & R(0) & \dots & R(-p+1) \\ \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & \dots & R(0) \end{pmatrix}^{-1} \begin{pmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{pmatrix}$$

where

$$(3.13) \quad R(k) = \sum_{t=1}^{n-k} y(t+k) y^*(t)$$

Usually, for moderate and large sample sizes, the differences between the estimates given by these approximations will be small. The use of least-squares estimate is attractive since it is easy to implement by using an adaptive algorithm.

Differentiating equation (3.7) with respect to σ_1^2 and equating the result to zero we obtain

$$(3.14) \quad \hat{\sigma}_1^2 = \frac{1}{n} \hat{a}_{p+1}^H R_{p+1}^{-1} \hat{a}_{p+1}$$

Using any of the above estimates for $\hat{a}_p$ and $\hat{\sigma}_1^2$ in equation (3.14), the likelihood function in (3.7) simplifies as

$$(3.15) \quad p(\underline{y} | \underline{\hat{a}}_p, \sigma_1^2) = (2\pi\sigma_1^2)^{-n/2} (\det \underline{M}_p)^{1/2} \exp\left(-\frac{n}{2}\right)$$

3.2.2 Maximum Likelihood Function Under H_0

Based on the linear model of (3.4a), the conditional probability density function under hypothesis H_0 is given by

$$(3.16) \quad p(\underline{y} | H_0) = p[y(1), y(2), \dots, y(n) | \underline{b}_q, \sigma_0^2] \\ = (2\pi\sigma_0^2)^{-n/2} \exp\left[-\frac{1}{2\sigma_0^2} \sum_{k=1}^n |y(k) - \underline{b}_q^T x(k)|^2\right]$$

Differentiating $p(\underline{y} | H_0)$ with respect to $\underline{b}_q$'s and equating to zero, the ML estimate of $\underline{b}_q$ equals

$$(3.17) \quad \underline{b}_q = \begin{bmatrix} R_{xx}(1,1) & R_{xx}(1,2) & \dots & R_{xx}(1,q) \\ R_{xx}(2,1) & R_{xx}(2,2) & \dots & R_{xx}(2,q) \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}(q,1) & R_{xx}(q,2) & \dots & R_{xx}(q,q) \end{bmatrix}^{-1} \begin{bmatrix} R_{yx}(1) \\ R_{yx}(2) \\ \vdots \\ R_{yx}(q) \end{bmatrix}$$

where

$$(3.18) \quad R_{xx}(l,m) = \sum_{t=1}^n x(t-l+1)x^*(t-m+1)$$

$$R_{yx}(k) = \sum_{t=1}^n y(t)x^*(t-k+1)$$

Combining equations (3.16) and (3.17), it follows that

$$(3.19) \quad p(\underline{y} | \underline{b}_q, \sigma_0^2) = (2\pi\sigma_0^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{k=1}^n |y(k) - \underline{b}_q^t \underline{x}(k)|^2\right\}$$

3.2.3 Test Statistic

Substituting equations (3.15), (3.19) into (3.6) for the generalized LLR, then taking logarithm and ignoring the terms independent of the input samples, we get the test statistic

$$(3.20) \quad \tau = -n \log \hat{\sigma}_1^2 + \frac{1}{2\hat{\sigma}_1^2} \sum_{k=1}^n |y(k) - \underline{b}_q^t \underline{x}(k)|^2$$

From equations (3.3) and (3.14)

$$(3.21) \quad \hat{\sigma}_1^2 = \frac{1}{n} \underline{a}_{p+1}^H \underline{R}_{p+1} \underline{a}_{p+1} = \frac{1}{n} \left[\sum_{k=p+1}^n |e_1(k)|^2 + \underline{v}_{p-p}^H \underline{M}_{p-p} \underline{v}_{p-p} \right]$$

where $\underline{v}_p$ is a short form of $\underline{v}_p(p)$, that is

$$\underline{v}_p = (y(p), y(p-1), \dots, y(1))^t$$

In equation (3.21), the expected value of $\sum_{k=p+1}^n |e_1(k)|^2$ is proportional to $(n-p)$, while the second term $\underline{v}_{p-p}^H \underline{M}_{p-p} \underline{v}_{p-p}$ is independent of n . Thus, for moderate and large sample size, (3.21) is dominated by the term $\sum_{k=p+1}^n |e_1(k)|^2$, and

so we may use the approximation:

$$(3.22) \quad \hat{\sigma}_1^2 = \frac{1}{n} \sum_{k=1}^n |e_1(k)|^2 = \frac{1}{n} \sum_{k=1}^n |y(k) - y_p^t(k-1) \hat{a}_p|^2$$

where

$$y_p^t(k-1) = (y(k-1), y(k-2), \dots, y(k-p))^T$$

In summary, the test statistic for the REAR modelling approach equals

$$(3.23a) \quad \ell = -n \log \hat{\sigma}_1^2 + \frac{1}{\sigma_0^2} \sum_{k=1}^n |y(k) - \hat{b}_q^t x(k)|^2$$

$$(3.23b) \quad \hat{\sigma}_1^2 = \frac{1}{n} \sum_{k=1}^n |y(k) - y_p^t(k-1) \hat{a}_p|^2$$

where $\hat{a}_p$ and $\hat{b}_p$ are given by (3.9) and (3.17), respectively.

3.3

TEST STATISTIC BASED ON ARAR MODELLING

Using a similar method, it is straightforward to obtain the test statistic for the ARAR modelling approach of equation (3.5) as:

$$(3.24a) \quad \ell = -n \log \hat{\sigma}_1^2 + \frac{1}{\sigma_0^2} \sum_{k=1}^n |y(k) - \hat{b}_q^t y(k)|^2$$

where

$$(3.24b) \quad \hat{\sigma}_1^2 = \frac{1}{n} \sum_{k=1}^n |y(k) - \hat{a}_p^t y_p^t(k-1)|^2$$

$\hat{a}_p$ and $\hat{b}_q$ are least squares estimates for a_p and b_q under hypotheses H_1 and H_0 , respectively. They are given by

$$(3.25) \quad \hat{a}_p = \begin{bmatrix} R(1,1) & R(1,2) & \dots & R(1,p) \\ R(2,1) & R(2,2) & \dots & R(2,p) \\ \vdots & \vdots & \ddots & \vdots \\ R(p,1) & R(p,2) & \dots & R(p,p) \end{bmatrix}^{-1} \begin{bmatrix} R(0,1) \\ R(0,2) \\ \vdots \\ R(0,p) \end{bmatrix}$$

and

$$(3.26) \quad \hat{b}_q = \begin{bmatrix} R(1,1) & R(1,2) & \dots & R(1,q) \\ \vdots & \vdots & \ddots & \vdots \\ R(q,1) & R(q,2) & \dots & R(q,q) \end{bmatrix}^{-1} \begin{bmatrix} R(0,1) \\ \vdots \\ R(0,q) \end{bmatrix}$$

where

$$R(l,m) = \sum_{t=1}^{n-l-m} y(t+l)y^*(t+m)$$

3.4

ADAPTIVE LLR DETECTION ALGORITHM

In section 3.3, the procedures for generalized LLR test were derived. The results given there involve the use of inverse matrices. However, for most applications, the recommended practice is to avoid the computation of the inverse of matrices. Moreover, it is desirable to compute up-to-date estimates for the unknown parameters and test statistic from previous estimates and the new received data samples, and so only the previous estimate needs to be stored. In addition to eliminating the need for storing the entire past observation data sequence, this recursive

algorithm is computationally more efficient than computing the estimates directly from the entire past observed data sequence at each step.

For ease of analysis, the derivation of the adaptive detection algorithm will be done for REAR modelling; nevertheless, the method is general in principle and the algorithm can be used with other linear models, requiring only minor modifications.

We reformulate the linear model for hypothesis H_0 , equation (3.4), as follows: Given observed data samples $y(k-n+1), y(k-n+2), \dots, y(k)$, we may write

$$(3.27a) \quad y(k) = \underline{x}^T(k) \underline{b}_q(k) + e_0(k)$$

$$(3.27b) \quad \underline{b}_q(k+1) = \underline{b}_q(k) + \Delta \underline{b}(k)$$

where $\underline{b}_q(k)$ is state column vector of order q at time k ,

$\{e_0(k)\}$ is a sequence of independent Gaussian variable with zero mean and variance σ_0^2 .

$\{\Delta \underline{b}(k)\}$ is a multiple Gaussian process $N(0, \sigma_0^2 \underline{Q}^0)$,

where $\underline{Q}^0$ is a positive definite diagonal matrix;

$e_0(k)$ is assumed to be independent of $\Delta \underline{b}(k)$.

A random walk model is used to describe the process so that we can track the slow changes in the statistics of the nonstationary environment.

The state $\underline{b}_q(k)$ is recursively computed by Kalman filter 1:

$$(3.28a) \quad \underline{v}^0(k) = \underline{P}^0(k) \underline{x}^*(k)$$

$$(3.28b) \quad f^0(k) = 1 + \underline{x}^t(k) \underline{v}^0(k)$$

$$(3.28c) \quad \underline{q}^0(k) = \underline{v}^0(k) / f^0(k)$$

$$(3.28d) \quad \underline{S}^0(k) = \underline{P}^0(k) - \underline{q}^0(k) \underline{v}^0(k)^H$$

$$(3.28e) \quad \underline{P}^0(k+1) = \underline{S}^0(k) + \underline{Q}^0$$

$$(3.28f) \quad e_0(k|k-1) = y(k) - \underline{x}^t(k) \underline{b}_q(k|k-1)$$

$$(3.28g) \quad \underline{b}_q(k+1|k) = \underline{b}_q(k|k-1) + \underline{q}^0(k) e_0(k|k-1)$$

The terms with superscript 0 are used to represent matrices and vectors in the Kalman Algorithm for estimating model parameters under hypothesis H_0 . Later on, superscript 1 is used for hypothesis H_1 . The term $\underline{P}^0(k)$ is the error covariance matrix for prediction estimate of $\underline{b}_q(k)$ from past data up to time $k-1$. Similarly, $\underline{S}^0(k)$ is the error covariance matrix for filtered estimate of $\underline{b}_q(k)$ from data up to time k . The term $\underline{q}^0(k)$ is the gain vector.

The Kalman filter transforms the correlated input data $\{y(k)\}$ into an innovation process $\{e_0(k|k-1)\}$, whose probability density function is $N(0, \sigma_0^2(k))$. It must be noted that the variance of $e_0(k|k-1)$ is not the same as that of $e_0(k)$ in that [26, 29, 43, 74].

$$\begin{aligned}
 E[|e_0(k)|^2] &= \sigma_0^2 \\
 (3.29) \quad E[|e_0(k|k-1)|^2] &= \sigma_0^2 [1 + \underline{x}^t(k) \underline{P}^0(k) \underline{x}(k)^H] = \sigma_0^2 f^0(k)
 \end{aligned}$$

In some applications, for example, moving target detection in airport surveillance radar system, to increase the detection performance, the prediction power of the clutter under hypothesis H_0 can be measured in advance and stored in the library. This clutter map, in the case of ground clutter-dominated situation [49], is updated for each scan by some simple algorithm whenever decision H_0 is made. We may thus assume that $\sigma_0^2(k)$ is known.

The coefficient $f^0(k)$ is a parameter in the algorithm that is determined by the nonstationarity of the input process. When $\{y(k)\}$ is stationary $\Delta b(k)$ is assigned to be zero. $\underline{P}^0(k)$ tends to zero after a moderate number of iterations [19]. This implies that

$$(3.40) \quad E[|e_0(k|k-1)|^2] \approx \sigma_0^2$$

since $f^0(k) = 1 + \underline{x}^t(k) \underline{P}^0(k) \underline{x}^*(k)$. For a nonstationary process, however, $\underline{P}^0(k)$ does not tend to zero; thus $f^0(k)$ is always larger than one. This implies that a nonstationary process is more difficult to be predicted than a stationary one.

When H_0 holds, the probability density function of $y(k-n+1), \dots, y(k)$, conditional on initial values of the Kalman

filter, can be expressed in terms of the innovation process,

$\{e_0(i|i-1)\}$ as

$$(3.41) \quad p[y(k-n+1), \dots, y(k) | \sigma_0^2(k|k-1), \underline{Q}^0, \underline{b}_q(0|-1), \underline{P}^0(0)] \\ = \frac{1}{\prod_{i=k-n+1}^k (2\pi\sigma_0^2(i|i-1))^{1/2}} \exp \left[-\frac{1}{2} \sum_{i=k-n+1}^k \frac{|e_0(i|i-1)|^2}{\sigma_0^2(i|i-1)} \right]$$

The set $\{\sigma_0^2(i|i-1)\}$ can be taken from the library, and

$\{e_0(i|i-1)\}$ from the output of the Kalman filter.

Similarly, the linear model for the received signal under H_1 can be rewritten as

$$(3.42a) \quad y(k) = \underline{y}_p^t(k-1) \underline{a}_p(k) + e_1(k)$$

$$(3.42b) \quad \underline{a}_p(k+1) = \underline{a}_p(k) + \Delta \underline{a}(k)$$

where $\underline{a}_p(k)$ is the state column vector of order p at time k .

$\{e_1(k)\}$ is a sequence of independent Gaussian random variables $N(0, \sigma_1^2)$.

$\{\Delta \underline{a}(k)\}$ is a multiple random Gaussian process.

$N(\underline{0}, \sigma_1^2 \underline{Q}^1)$ where $\underline{Q}^1$ is a positive definite diagonal matrix.

$\{e_1(k)\}$ is assumed independent of $\{\Delta \underline{a}(k)\}$.

$$\underline{y}_p^t(k-1) = (y(k-1), \dots, y(k-p))^t$$

The input sequence, $y(k-n+1), \dots, y(k)$, under hypothesis H_1 , can be transformed into the innovation process $\{e_1(k|k-1)\}$ by using Kalman filter IF that is described by

$$(3.42a) \quad \underline{v}^1(k) = \underline{p}^1(k) \underline{y}_p^*(k-1)$$

$$(3.42b) \quad f^1(k) = 1 + \underline{y}_p^t(k-1) \underline{v}^1(k)$$

$$(3.42c) \quad \underline{q}^1(k) = \underline{v}^1(k) / f^1(k)$$

$$(3.42d) \quad \underline{s}^1(k) = \underline{p}^1(k) - \underline{q}^1(k) \underline{v}^1(k)^H$$

$$(3.42e) \quad \underline{p}^1(k+1) = \underline{s}^1(k) + \underline{Q}^1$$

$$(3.42f) \quad e_1(k|k-1) = y(k) - \underline{y}_p^t(k-1) \underline{a}_p(k|k-1)$$

$$(3.42g) \quad \underline{a}_p(k+1|k) = \underline{a}_p(k|k-1) + \underline{q}^1(k) e_1(k|k-1)$$

Since $\{e_1(i|i-1)\}$ is Gaussian distributed with zero mean and

$$(3.43) \quad E[|e_1(i|i-1)|^2] = \sigma_1^2 f^1(i)$$

thus, the probability function of $y(k-n+1), \dots, y(k)$ under H_1 is given by

$$(3.44) \quad p[y(k-n+1), y(k-n+2), \dots, y(k) | \sigma_1^2, \underline{a}_p(0|-1), \underline{p}^1(0)] \\ = \frac{1}{\pi^k [2\pi\sigma_1^2 f^1(i)]^{1/2}} \exp\left[-\frac{1}{2} \sum_{i=k-n+1}^k \frac{|e_1(i|i-1)|^2}{\sigma_1^2 f^1(i)}\right]$$

From equation (3.44), the ML estimate of σ_1^2 is given by

$$(3.45) \quad \hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=k-n+1}^k \frac{|e_1(i|i-1)|^2}{f^1(i)}$$

n is a preassigned length of data block based on which decision is made. Obviously, the estimate $\hat{\sigma}_1^2$ is function of time index k . To emphasize this we will rewrite it as $\hat{\sigma}_{1,k}^2$.

Combining equations (3.44) and (3.45), we obtain

$$(3.46) \quad p[y(k-n+1), \dots, y(k) | \hat{\sigma}_1^2, \underline{a}_p(0|-1), \underline{p}^1(0)] \\ = \frac{1}{\pi^k [2\pi\hat{\sigma}_{1,k}^2 f^1(i)]^{1/2}} \exp\left(-\frac{n}{2}\right)$$

The generalized LLR is obtained by dividing equation (3.46) by (3.41). Taking the natural logarithm of this ratio and ignoring the terms that are independent of input data, we get the test statistic:

$$(3.47) \quad \ell(k) = \alpha(k) + \beta(k) - \theta(k) - n \log_e \gamma(k)$$

where

$$(3.48) \quad \alpha(k) = \sum_{i=k-n+1}^k |e_0(i|i-1)|^2 / \sigma_0^2(i|i-1)$$

$$(3.49) \quad \beta(k) = \sum_{i=k-n+1}^k \log \sigma_0^2(i|i-1)$$

$$(3.50) \quad \theta(k) = \sum_{i=k-n+1}^k \log f^1(i)$$

$$(3.51) \quad \gamma(k) = \frac{1}{n} \sum_{i=k-n+1}^k |e_1(i|i-1)|^2 / f^1(i)$$

Now we derive recursions for computing the test statistic $\ell(k)$. Consider $\alpha(k)$ in equation (3.48) first. We write

$$(3.52) \quad \alpha(k) = \sum_{i=k-n+1}^k |e_0(i|i-1)|^2 / \sigma_0^2(i|i-1)$$

$$= \sum_{i=k-n}^{k-1} [|e_0(i|i-1)|^2 / \sigma_0^2(i|i-1)]$$

$$+ |e_0(k|k-1)|^2 / \sigma_0^2(k|k-1)$$

$$- |e_0(k-n|k-n-1)|^2 / \sigma_0^2(k-n|k-n-1)$$

$$= \alpha(k-1) + \frac{|e_0(k|k-1)|^2}{\sigma_0^2(k|k-1)} - \frac{|e_0(k-n|k-n-1)|^2}{\sigma_0^2(k-n|k-n-1)}$$

Similarly, $\beta(k)$, $\theta(k)$ and $\gamma(k)$ in equations (3.49-3.51) can be written in the following recursive versions:

$$(3.53) \quad \beta(k) = \beta(k-1) + \log \frac{\sigma_0^2(k|k-1)}{\sigma_0^2(k-n|k-n-1)}$$

$$(3.54) \quad \theta(k) = \theta(k-1) + \log \frac{f^1(k)}{f^1(k-n)}$$

$$(3.55) \quad \gamma(k) = \gamma(k-1) + \frac{1}{n} \frac{|e_1(k|k-1)|^2}{f^1(k)} - \frac{1}{n} \frac{|e_1(k-n|k-n-1)|^2}{f^1(k-n)}$$

Equations (3.47) and (3.52-3.55) constitute the recursive formulae for updating the test statistic $\lambda(k)$. Clearly, these equations require shift registers of length n to store $\{\sigma_0^2(i|i-1)\}$, $\{e_0(i|i-1)\}$, $\{f^1(i)\}$ back to time $i = k-n$. From equations (3.48-3.51), we see that the summations may be viewed as sums with rectangular window of length n . It is possible to reduce the memory requirement by replacing the rectangular window with an exponential one. Modifying the summations in (3.48-3.51) in this way, we may write

$$(3.56) \quad \alpha(k) = \sum_{i=1}^k \mu^{k-i} \frac{|e_0(i|i-1)|^2}{\sigma_0^2(i|i-1)}$$

$$(3.57) \quad \beta(k) = \sum_{i=1}^k \mu^{k-i} \log \sigma_0^2(i|i-1)$$

$$(3.58) \quad \theta(k) = \sum_{i=1}^k \mu^{k-i} \log f^1(i)$$

$$(3.59) \quad \gamma(k) = (1-\mu) \sum_{i=1}^k \mu^{k-i} \frac{|e_1(i|i-1)|^2}{f^1(i)}$$

where μ is a constant used to control the memory length of past data. In particular, $(1-\mu)^{-1}$ is roughly a measure of the memory of the filter. The multiplying factor $(1-\mu)$ is introduced in equation (3.59) so that the expected value of $\gamma(k)$ equals the variance $\sigma_1^2(k|k-1)$. The updating formulae resulting from the use of equations (3.56-3.59) are as follows:

$$(3.60) \quad \alpha(k) = \mu \alpha(k-1) + \frac{|e_0(k|k-1)|^2}{\sigma_0^2(k|k-1)}$$

$$(3.61) \quad \beta(k) = \mu \beta(k-1) + \log \sigma_0^2(k|k-1)$$

$$(3.62) \quad \theta(k) = \mu \theta(k-1) + \log f^1(k)$$

$$(3.63) \quad \gamma(k) = \mu \gamma(k-1) + (1-\mu) \frac{|e_1(k|k-1)|^2}{f^1(k)}$$

Comparing the algorithm for updating test statistic given in equations (3.52-3.55) with that in equations (3.60-3.63), we may make the following observations:

(1) In the context of radar detection where a target echo generally lasts for a short period of time (say 15-20 sample points), the use of a rectangular window is beneficial for making full use of the target information from the received data at the expense of more memory.

(2) The use of exponential decaying window helps reduce the memory requirement. Its drawback is that it has a tail longer than the duration of the target echo. Therefore, it is possible to weaken the effect of the target echo and enhance the clutter, which is undesirable. Accordingly, we should use a value for the coefficient μ such that the effective memory length approximately equals the duration of the target echo. For instance, for a target duration of 20 data samples, the effective memory length is

$$\frac{1}{1-\mu} = 20,$$

which gives $\mu = 0.95$.

We now have all the formulae we need to recursively compute the test statistic. In summary, the procedure for adaptive detection is as follows:

- (1) Start the detection algorithm with the following initial values.

- a) for Kalman filters

$$\underline{p}^0(0) = \text{diag}(d_0, \dots, d_0)$$

$$\underline{p}^1(0) = \text{diag}(d_1, \dots, d_1)$$

$$\underline{b}_q(0|-1) = \underline{0}$$

$$\underline{a}_p(0|-1) = \underline{0}$$

$$\underline{y}_p(0) = \underline{0}$$

$$\underline{x}(0) = \underline{0}$$

$$\underline{Q}^0 = \text{diag}(q_0, q_0, \dots, q_0)$$

$$\underline{Q}^1 = \text{diag}(q_1, q_1, \dots, q_1)$$

Without a priori knowledge, we assign large positive values to d_0 and d_1 , and small positive values to q_0 and q_1 , depending on the nonstationarity of the input process. Normally, we choose

$$0 \leq q_0, q_1 \leq 0.2$$

Larger values are suited for highly nonstationary environments. In theory, q_0 and q_1 can be considered as unknown parameters and thereby estimated together with others by using the maximum-likelihood approach. However, this will lead to a complicated nonlinear optimization problem that increases the computational burden substantially. Experience shows that the test statistic $\lambda(k)$ is not very sensitive to small changes in q_0 and q_1 . Thus, we may choose appropriate values for them during the training period and then fix them in the process of signal processing. q_1 is normally larger than q_0 since the received signal under H_1 , composed of noise plus signal, involves a sudden change in statistics.

$$b) \quad \alpha(0) = 0$$

$$\beta(0) = 0$$

$$\delta(0) = 0$$

$$\gamma(0) = 0$$

- (2) Compute the innovation under H_0 , namely, $e_0(k|k-1)$, using equation (3.28).

- (3) Compute the innovation under H_1 , namely, $e_1(k|k-1)$, and parameter $f^1(k)$ by using equation (3.42).

- (4) Update the test statistic $\ell(k)$ by using equations (3.52-3.55), or equations (3.60-3.63).

- (5) Make a decision whether the signal of interest is present or not, based on the rule

$$\ell(k) \underset{H_0}{\overset{H_1}{\geq}} TH$$

where TH is a prescribed threshold. The value assigned to TH depends on the false alarm rate required.

- (6) Increment k by 1, go back to step (1) and repeat the procedure.

Several criteria may be used to determine a threshold for the signal detector. Among them, the Neyman-Pearson criterion is the most appropriate one for radar detection. According to this criterion, the threshold is determined by fixing the probability of making an error in favour of hypothesis H_1 when H_0 is true. In the radar context, this probability is known as the false alarm rate.

A block diagram of the adaptive LLR detection receiver based on the algorithm described is shown in Figure 3.1. Kalman filter 1 is composed of equation (3.28), and Kalman filter 2 is composed of equation (3.42). The output from the first Kalman filter equals the innovation $e_0(k|k-1)$, and that from the second Kalman filter equals $e_1(k|k-1)$ and $f^1(k)$. $e_0(k|k-1)$, $e_1(k|k-1)$ and $f^1(k)$ together with $\sigma_0^2(k)$ from the memory are applied to the recursive computer to find $\ell(k)$. Finally, $\ell(k)$ is compared with threshold TH in the decision device.

Although the adaptive detection algorithm described above is developed for use with the REAR modelling approach it is applicable to other situations (e.g., the ARAR modelling approach) by making simple modifications:

- (1) Modify the corresponding matrices in Kalman filters according to the linear models to be used.
- (2) Establish the test statistic $\ell(k)$ in accordance with the constraint conditions given in the hypotheses and the estimated parameters from the Kalman filters.

3.5 PROPERTIES OF LLR TEST BASED ON REAR MODELLING

The REAR modelling approach plays an important role in the LLR detection scheme design. It has some desirable properties that make the receiver based on this approach better than others in performance.

It is therefore important that we understand its various properties and their implications. In particular, using linear model (3.4) and the ML estimates of the model parameters given in equations (3.17)-(3.18), we find that the main properties of the receiver are as follows:

Property 1. The ML estimate of model parameter b_q , given the input $y(1), y(2), \dots, y(n)$ under H_0 and denoted by $\hat{b}_q(y|H_0)$, equals that based on $y(1), y(2), \dots, y(n)$ under H_1 and denoted by $\hat{b}_q(y|H_1)$. That is,

$$(3.67) \quad \hat{b}_q(y|H_1) = \hat{b}_q(y|H_0)$$

Proof. We rewrite the solution for linear model (3.4a) as follows (see equations (3.17)-(3.18)):

$$(3.17) \quad \hat{b}_q = \left(\sum_{i=1}^n \underline{x}(i) \underline{x}(i)^H \right)^{-1} \left(\sum_{i=1}^n y(i) \underline{x}(i)^* \right).$$

where $\underline{x}(i) = (x(i-q+1), x(i-q+2), \dots, x(i))^T$

When H_0 is true, the received signal is only composed of noise $w(i)$, that is,

$$y(i) = w(i).$$

Accordingly,

$$(3.68) \quad \underline{b}_q(y|H_0) = \left(\sum_{i=1}^n \underline{x}(i) \underline{x}(i)^H \right)^{-1} \sum_{i=1}^n w(i) \underline{x}(i)^*$$

When H_1 is true, $y(i)$ consists of signal plus noise, that is,

$$y(i) = w(i) + s(i).$$

$$(3.69) \quad \underline{b}_q(y|H_1) = \left(\sum_{i=1}^n \underline{x}(i) \underline{x}(i)^H \right)^{-1} \left[\sum_{i=1}^n w(i) \underline{x}(i)^* + \sum_{i=1}^n s(i) \underline{x}(i)^* \right]$$

Now, signal $s(i)$ is assumed to be independent of reference

$\underline{x}(i)$; this implies that

$$\sum_{i=1}^n s(i) \underline{x}(i)^* \approx 0$$

Therefore,

$$(3.70) \quad \underline{b}_q(y|H_1) = \left(\sum_{i=1}^n \underline{x}(i) \underline{x}(i)^H \right)^{-1} \sum_{i=1}^n w(i) \underline{x}(i)^*$$

Comparison of (3.68) with (3.70) completes the proof.

It is noted that when H_0 is true, model (3.4a) is a correct model for the input sequence; as a result, $\underline{b}_q(\underline{y}|H_0)$ is a correct estimate of $\underline{b}_q$. Thus, property 1 implies that linear model (3.4a) always leads to correct estimates for model parameters whether the signal component is absent or present.

This property is not shared by other models for H_0 -hypothesis. Take ARAR modelling for example; the least squares estimate for model (3.5a) is given by (3.9)-(3.10), and approximated by

$$(3.71) \quad \underline{b}_q = \left(\sum_{k=1}^n \underline{y}_q(k-1) \underline{y}_q(k-1)^H \right)^{-1} \left(\sum_{k=1}^n y(k) \underline{y}_q(k-1)^* \right)$$

When H_0 holds, $y(k) = w(k)$; thus $\underline{b}_q$ is given by

$$(3.72) \quad \underline{b}_q(\underline{y}|H_0) = \left[\sum_{k=1}^n \underline{w}_q(k-1) \underline{w}_q(k-1)^H \right]^{-1} \left[\sum_{k=1}^n w(k) \underline{w}_q(k-1)^* \right]$$

where $\underline{w}_q(k-1) = (w(k-q), w(k-q+1), \dots, w(k-1))^T$

When H_1 holds, $y(k) = w(k) + s(k)$, and $\underline{b}_q$ equals

$$(3.73) \quad \underline{b}_q(\underline{y}|H_1) = \left[\sum_{k=1}^n (\underline{w}_q(k-1) + \underline{s}_q(k-1)) (\underline{w}_q(k-1) \right.$$

(continued)

$$\begin{aligned}
& + \underline{s}_q(k-1) \underline{s}_q(k-1)^H \Big]^{-1} \left[\sum_{k=1}^n (w(k) + s(k)) (\underline{w}_q(k+1) + \underline{s}_q(k-1))^* \right] \\
& - \sum_{k=1}^n \underline{w}_q(k-1) \underline{w}_q(k-1)^H \\
& + \sum_{k=1}^n \underline{s}_q(k-1) \underline{s}_q(k-1)^H \Big]^{-1} \left(\sum_{k=1}^n w(k) \underline{w}_q(k-1)^* \right. \\
& \left. + \sum_{k=1}^n s(k) \underline{s}_q(k-1)^* \right)
\end{aligned}$$

where $\underline{s}_q(k-1) = (s(k-q), s(k-q+1), \dots, s(k-1))^T$ is the signal component vector. The last equality holds due to independence of $s(k)$ from $w(k)$.

From equations (3.72) and (3.73), we conclude for ABAR modelling that:

$$1) \quad \underline{b}_q(\underline{y}|H_1) \neq \underline{b}_q(\underline{y}|H_0)$$

2) When SNR increases, equation (3.73) shows that $\underline{b}_q(\underline{y}|H_1)$ is dominated by the signal component.

Therefore, the larger SNR, the less is $\underline{b}_q(\underline{y}|H_1)$ related to $\underline{b}_q(\underline{y}|H_0)$.

We rewrite LR (3.6) as follows

$$(3.6) \quad \Lambda(\underline{y}) = \frac{p(\underline{y}|\underline{a}_q, \sigma_1^2)}{p(\underline{y}|\underline{b}_q, \sigma_0^2)}$$

We note that $\hat{a}_p$, $\hat{b}_q$, $\hat{\sigma}_1^2$ are estimated from the input data $\{y(k)\}$; the input under H_0 and H_1 generally give different estimates. When H_0 is true, LR is given by:

$$(3.74) \quad \Lambda(y|H_0) = \frac{p(y|\hat{a}_p(y|H_0), \hat{\sigma}_1^2(y|H_0))}{p(y|\hat{b}_q(y|H_0), \hat{\sigma}_0^2)}$$

This will be employed to determine the constant false alarm rate (CFAR) threshold. When H_1 holds, the LR is given by

$$(3.75) \quad \Lambda(y|H_1) = \frac{p(y|\hat{a}_p(y|H_1), \hat{\sigma}_1^2(y|H_1))}{p(y|\hat{b}_q(y|H_1), \hat{\sigma}_0^2)}$$

In this equation, all the unknown parameters, for both H_0 -model and H_1 -model, are estimated from the observed samples under hypothesis H_1 , since only this set of data is available. For H_1 -model parameters, the signal model is matched to the received signal, and therefore, the estimates $\hat{a}_p(y|H_1)$ and $\hat{\sigma}_1^2(y|H_1)$ would be good approximates of their true values if the sample size is large. For the H_0 -model parameter, however, the signal model is mismatched to the received signal and thus, the estimate $\hat{b}_q(y|H_1)$ may have a serious deviation from its true value. The degree of such a deviation depends on what kind of H_0 -model we use. To analyze the effect of the deviation of this kind

on the LR, we define

$$(3.76) \quad \Lambda_0(y|H_1) = \frac{R[\hat{y}|\hat{a}_p(y|H_1), \sigma_1^2(y|H_1)]}{p[y|\hat{b}_q(y|H_0), \sigma_0^2]}$$

$\Lambda_0(y|H_1)$ represents the LR in which $\hat{a}_p$ and σ_1^2 are evaluated from $y(1), \dots, y(n)$ under hypothesis H_1 , while $\hat{b}_q$ is estimated from $y(1), \dots, y(n)$ under H_0 . In order to describe the difference between $\Lambda(y|H_1)$ and $\Lambda_0(y|H_1)$, we define a parameter n as

$$(3.77) \quad n = \ln \left[\frac{\Lambda(y|H_1)}{\Lambda_0(y|H_1)} \right]$$

Then using equations (3.75) and (3.76), we have

$$(3.78) \quad n = \ln \left[\frac{p(y|\hat{b}_q(y|H_0), \sigma_0^2)}{p(y|\hat{b}_q(y|H_1), \sigma_0^2)} \right]$$

When H_1 is true, $\hat{b}_q(y|H_1)$ is an ML estimate of parameter b_q , but $\hat{b}_q(y|H_0)$ is not. Thus,

$$(3.79) \quad p(y|\hat{b}_q(y|H_1), \sigma_0^2) \geq p(y|\hat{b}_q(y|H_0), \sigma_0^2)$$

and

$$(3.80) \quad n \leq 0$$

Apparently, $|n|$ represents the loss in LLR due to erroneous estimate of $\hat{b}_q$ when H_1 holds. Consequently, we may refer to $|n|$ as the loss parameter, and so state

Property 2. The loss parameter $|n|$ is zero for the LLR test based on REAR modelling.

This result is a direct consequence of Property 1.

For most linear models, however, $\hat{b}_q(y|H_1) \neq \hat{b}_q(y|H_0)$; therefore, the loss parameter is non-zero. When using ARAR modelling, the larger the SNR, the less is $\hat{b}_q(y|H_1)$ related to $\hat{b}_q(y|H_0)$, and the more serious the loss n .

Property 3. The test statistic based on REAR modelling increases approximately in proportion to $E|s(k)|^2/\sigma_0^2$, where $E|s(k)|^2$ is signal power and $\sigma_0^2 = E|e_0(k)|^2$, the power of measurement noise $e_0(k)$, when H_0 holds.

This is a highly desirable property of the REAR modelling approach. To verify it, we rewrite the test statistic (3.23) as follows:

$$(3.81) \quad \lambda(n) = -n \log \frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2} + \frac{1}{\hat{\sigma}_0^2} \sum_{k=1}^n |y(k) - \hat{b}_q^t x(k)|^2$$

$$\hat{\sigma}_1^2 = \frac{1}{n} \sum_{k=1}^n |y(k) - \hat{a}_{p-1}^t \hat{z}_{p-1}(k-1)|^2$$

We analyze first the second term in the summation of the first line. When H_1 is true, we have

$$\begin{aligned} & \frac{1}{\sigma_0^2} \sum_{k=1}^n |y(k) - \hat{b}_q^t x(k)|^2 \\ &= \frac{1}{\sigma_0^2} \sum_{k=1}^n |w(k) + s(k) - \hat{b}_q(y|H_1) x(k)|^2 \\ &= \frac{1}{\sigma_0^2} \sum_{k=1}^n |(w(k) - \hat{b}_q(y|H_0) x(k)) + s(k)|^2 \end{aligned}$$

In the above equality, we have used Property 1,

i. e., $\hat{b}_q(y|H_1) = \hat{b}_q(y|H_0)$. Noting that $s(k)$ is independent of $w(k)$, $x(k)$ and $\hat{b}_q(y|H_0)$. We therefore have

$$\begin{aligned} (3.82) \quad & \frac{1}{\sigma_0^2} \sum_{k=1}^n |y(k) - \hat{b}_q(y|H_1) x(k)|^2 \\ &= \frac{1}{\sigma_0^2} \sum_{k=1}^n |w(k) - \hat{b}_q(y|H_1) x(k)|^2 + \frac{1}{\sigma_0^2} \sum_{k=1}^n |s(k)|^2 \\ &+ \frac{1}{\sigma_0^2} \sum_{k=1}^n (w(k) - \hat{b}_q(y|H_1) x(k))^* s(k) \\ &+ \frac{1}{\sigma_0^2} \sum_{k=1}^n (w(k) - \hat{b}_q(y|H_1) x(k)) s^*(k) \end{aligned}$$

.....(continued)

$$= \frac{1}{2} \sum_{k=1}^n |w(k)|^2 + \frac{1}{2} \sum_{k=1}^n |s(k)|^2$$

$$= \frac{1}{2} \cdot n \sigma_0^2 + \frac{1}{2} n E |s(k)|^2$$

$$= n [1 + E |s(k)|^2 / \sigma_0^2]$$

In general, prediction error power σ_0^2 is very small compared to the average power of the original sequence $\{w(k)\}$. Thus, when signal power is moderate, ratio $E |s(k)|^2 / \sigma_0^2$ is much greater than 1.

Now, consider the first term in the summation in equations (3.81); σ_1^2 is an approximate of the linear prediction error power, $E |e_1(k)|^2$, under hypothesis H_1 . According to [13, 47], the prediction error power of a stochastic process can be expressed in term of its spectrum as:

$$(3.83) \quad E |e_1(k)|^2 = \exp \left\{ \int_{-0.5}^{0.5} \log [S(f) + W(f)] df \right\}$$

where $S(f)$ is the power spectrum of signal component and $W(f)$ is that of the noise component of $\{y(k)\}$. The spectrum of $y(k)$ is the sum of the spectra of the signal and noise

component since $s(k)$ is independent of $w(k)$.

Combining equations (3.81) and (3.83), we write the first term as

$$(3.84) \quad -n \log \hat{\sigma}_1^2 = -n \int_{-0.5}^{0.5} \log[S(f) + W(f)] df$$

Because of the logarithm operation, this term decreases very slowly with an increase in signal power. The test statistic $\lambda(n)$ is therefore dominated by the second term for moderate to large signal component. Hence, it is proportional to $E|s(k)|^2/\sigma_0^2$. It is interesting to note that $E|s(k)|^2/\sigma_0^2$ equals the signal-to-noise prediction-error power ratio, rather than the conventional signal-to-noise power ratio.

3.6 EXAMPLES

The purpose of this section is to illustrate the usage of the adaptive algorithm and experimentally test the properties described above by applying the algorithm to real baseband radar data.

We are given a scan of radar clutter data, referred to as 'WTRNW1'. It has 32 range rings, each of which contains 3296 samples, and partly shown in Figure 3.2.

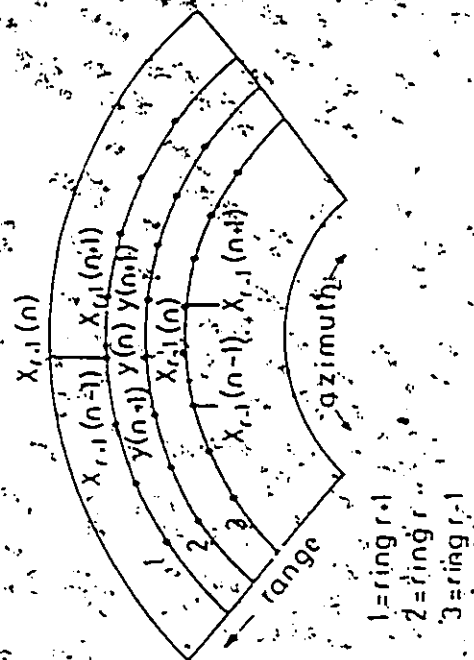


Fig.3.2 Baseband radar data taken from

adjacent range ring; $y(n)$ is received

signal to be processed.

The experiments were done under the following conditions:

- (1) The test statistic $\lambda(k)$ was calculated for 20 samples, i.e., $n = 20$, by using the algorithm developed in Section 3.4. $\lambda(k)$ was observed at the end of each data block.

- (2) The REAR model used is defined by

$$H_0: y(k) = (x_{r-1}(k), x_{r+1}(k)) \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + e_0(k)$$

$$H_1: y(k) = \sum_{i=1}^5 a_i y(k-i) + e_1(k)$$

The ARAR model used is defined by

$$H_0: \hat{y}(k) = b_1 y(k-1) + e_0(k)$$

$$H_1: y(k) = \sum_{i=1}^5 a_i y(k-i) + e_1(k)$$

- (3) Data sets used for the experiments were as follows:

Data Set I: Weather clutter dominated data is composed of 20 samples, located between azimuth 3100-3119

at range ring 25. The power spectrum of this data set is shown in Figure 3.3a.

Data Set II: Weather clutter is mixed with ground clutter, with 20 samples located between azimuth 1400-1419 at ring 15. The power spectrum of this second data set is given in Figure 3.4a.

- (4) An artificial target model was used to generate the target echo. It was an analytical signal with envelope complex Gaussian-shaped and central frequency adjustable. For more detail, see Section 5.6. The target spectrum is given in Figure 3.5a, 3.6a and 3.7a, its position depending on the Doppler shift.
- (5) The normalized LLR is defined as the value of test statistic $\lambda(k)$ for given SCR minus that for SCR = $-\infty$ dB, namely, absence of target.

We first study the test statistic as a function of target Doppler frequency. Figure 3.3 shows the results of normalized LLR vs. target Doppler for data Set I and Figure 3.4 for data Set II.

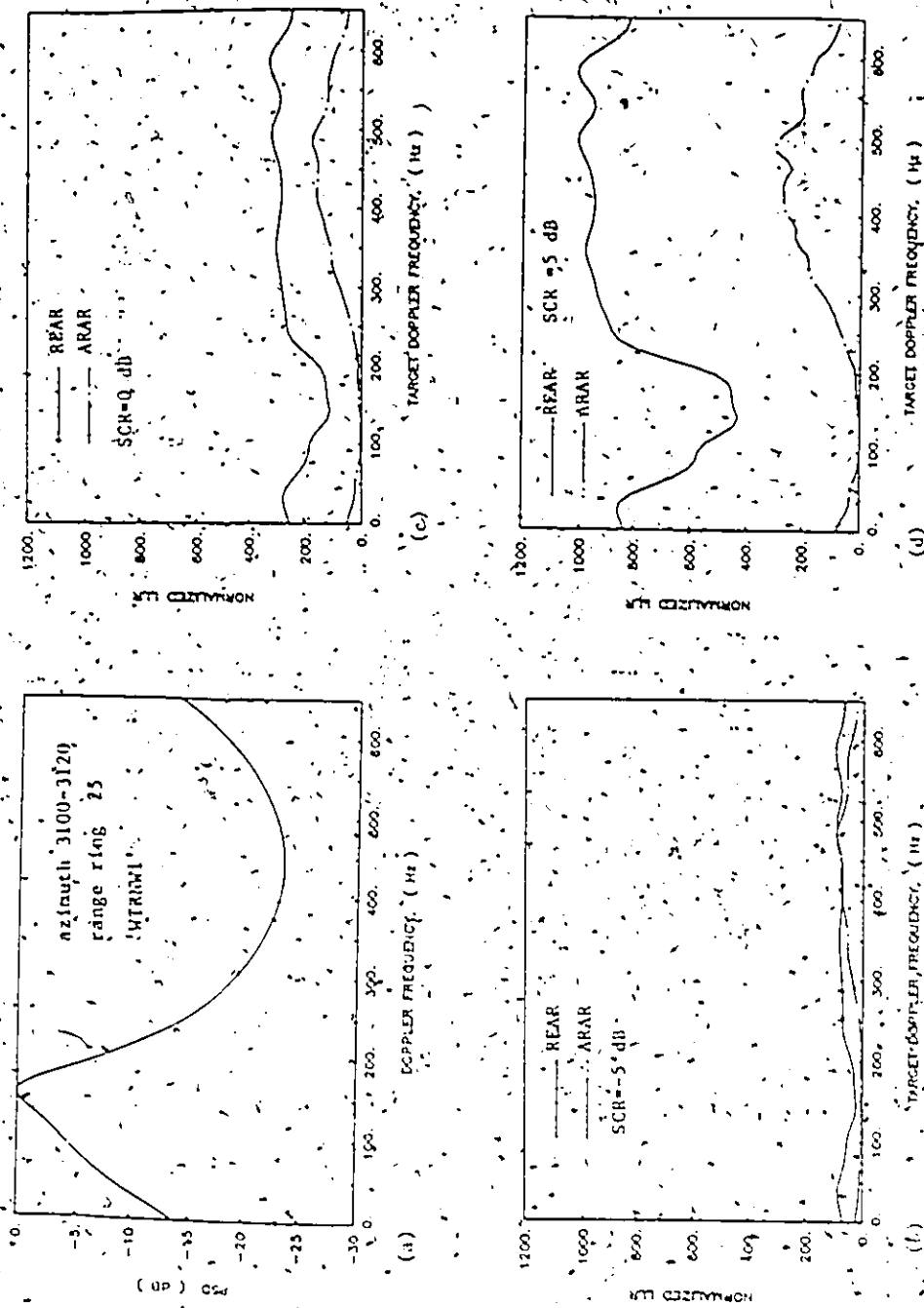
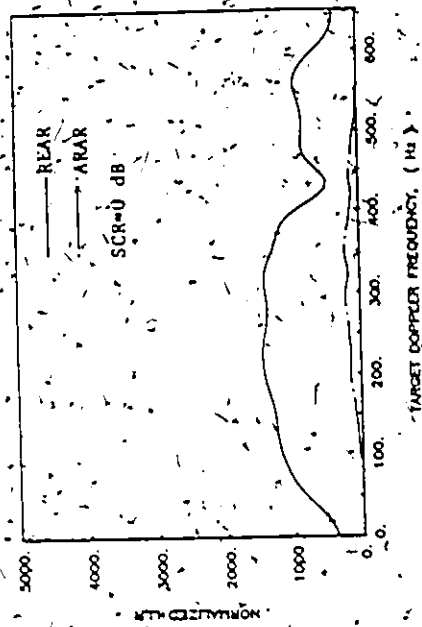
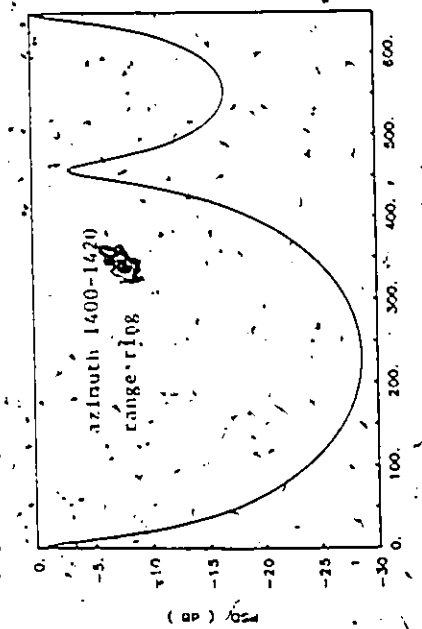


FIG. 3.3 Curves of normalized LLR v.s. target doppler frequency.
 (a) Power spectrum of weather clutter taken from azimuth 3100-3120 on ring 25
 (b) LLR v.s. target doppler curve for SCR=5 dB
 (c) LLR v.s. target doppler curve for SCR=0 dB
 (d) LLR v.s. target doppler curve for SCR=5 dB
 (e) LLR v.s. target doppler curve for SCR=5 dB
 (f) LLR v.s. target doppler curve for SCR=5 dB



(a)



(b)



(c)

(d)

Fig. 3. Curves of normalized LL v.s. target doppler frequency.

(a) Power spectrum of weather clutter taken from azimuth 1400-1420 on ring 15.

(b) Normalized LL v.s. target doppler for SCR=-5 dB

(c) SCR=0 dB (d) SCR=5 dB

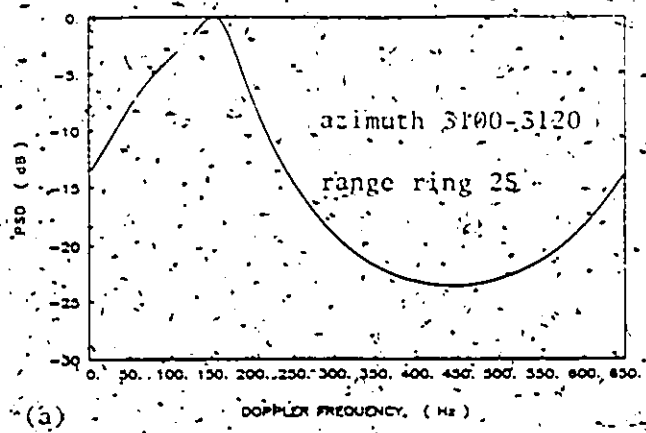
From Figures 3.3 and 3.4, we may make the following observations:

- (1) The LLR increases with SCR.
- (2) The test statistic based on REAR modelling is nearly always better than that based on ARAR model.
- (3) A drop in LLR occurs when the target peak is masked by a clutter peak.

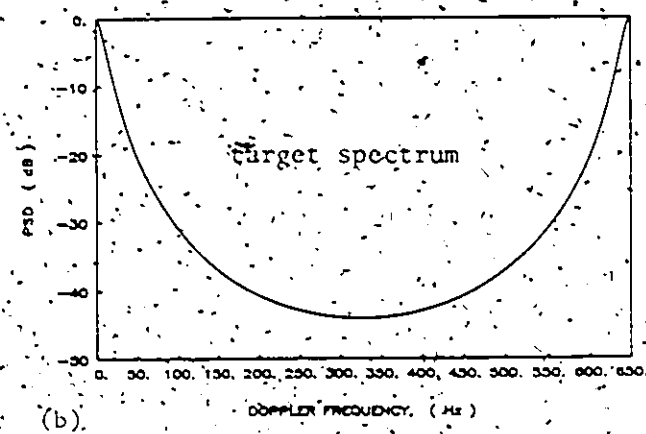
Next, we investigate the effect of the separation of target from the clutter peak (in the frequency domain) on the value of normalized LLR.

Experimental results using data Set I are drawn in Figure 3.5. For purposes of comparison, spectra of clutter and target are included in Figure 3.5a. The target doppler equals zero HZ while clutter peak is around 175 HZ. It represents a moderate separation case. We see that for REAR modelling LLR increases nearly linearly with SCR in dB; for ARAR modelling, however, it goes slightly up with SCR.

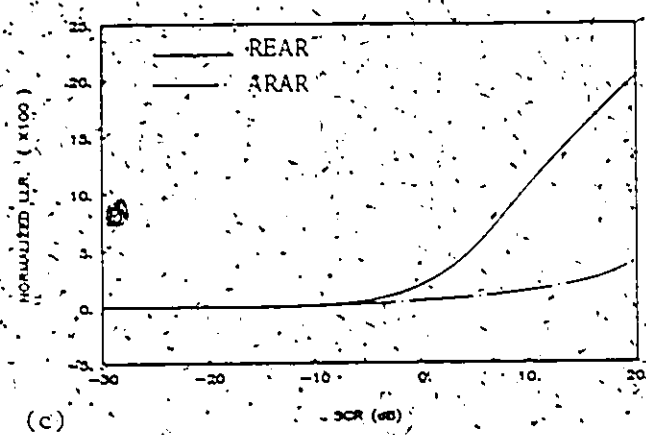
Figure 3.6 shows the case in which the target, centered at 150 HZ, is completely obscured by the clutter peak. It is seen that the LLR based on ARAR modelling does not change essentially as SCR increases; for the REAR modelling the LLR increases, although not so quickly, with SCR. Finally, the case where the target is completely



(a)



(b)



(c)

Fig. 3.5 The effect of relative position of target peak and clutter peak on normalized LLR.

Moderate separation: target peak 0 Hz, clutter peak 170Hz.

(a) Clutter spectrum (b) target spectrum

(c) LLR v.s: SCR

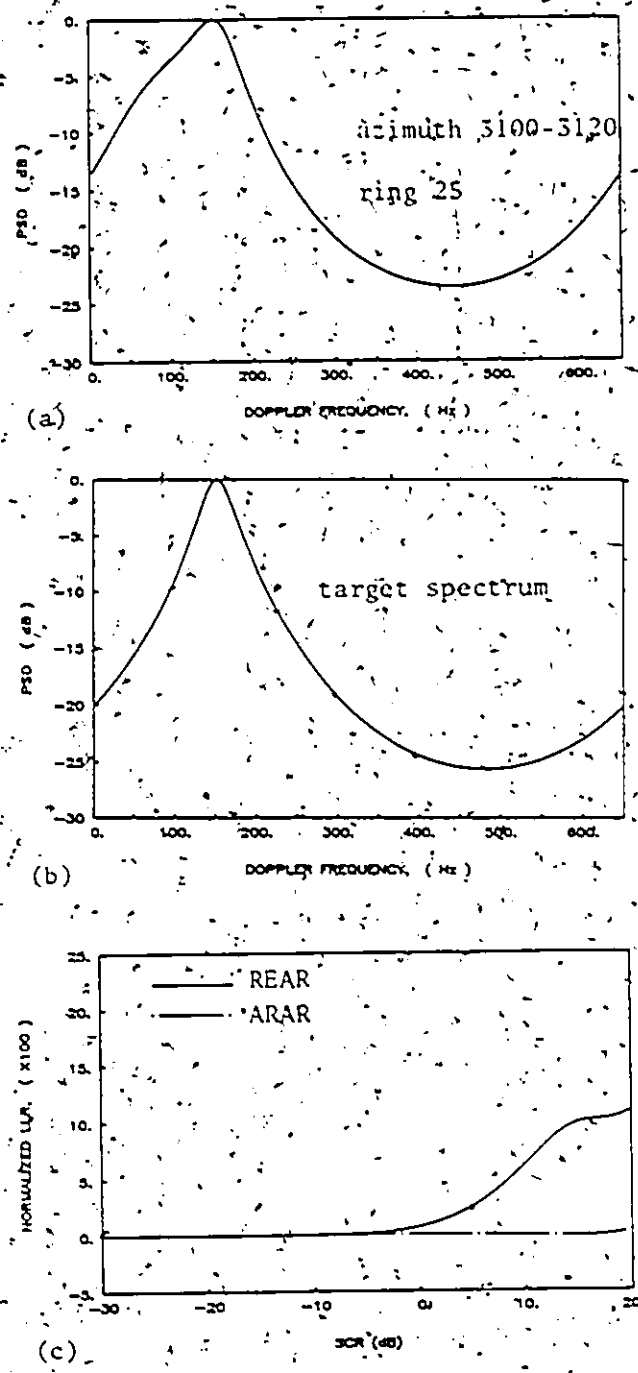
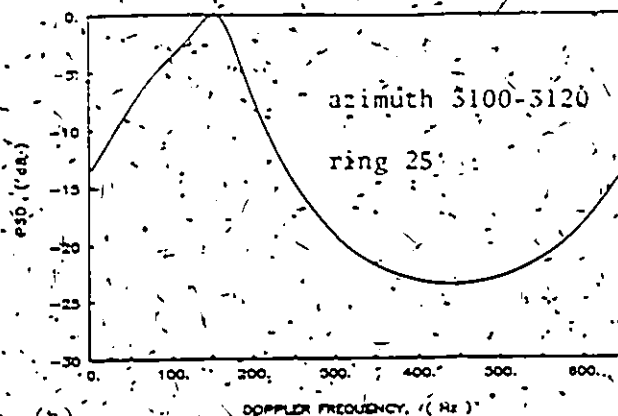


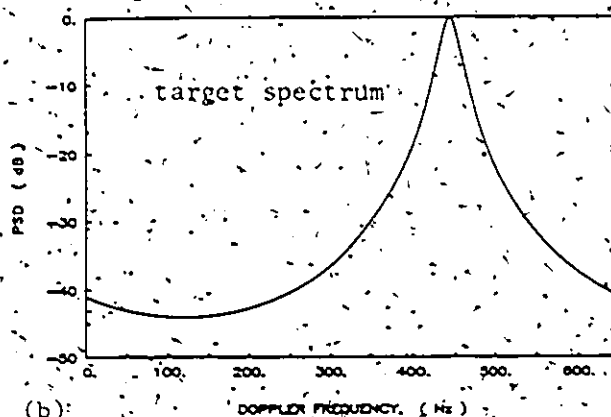
Fig. 3.6 The effect of relative position of target peak and clutter peak on normalized LLR.

Shown here is the case of the target completely obscured by clutter peak.

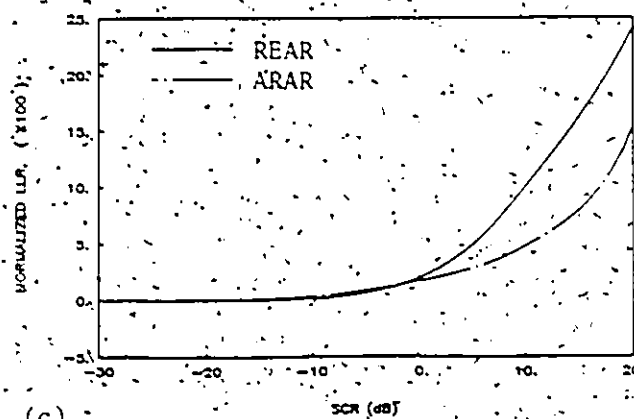
- (a) Clutter spectrum
- (b) target spectrum
- (c) LLR v.s. SCR



(a)



(b)



(c)

Fig. 3.7 The effect of relative position of target peak and clutter peak on normalized LLR.

Large separation: target peak 450 Hz,
clutter peak 170 Hz.

(a) clutter spectrum (b) target spectrum

(c) LLR v.s. SCR.

separated from the clutter peak is given in Figure 3.7.

As expected, the LLR goes up very fast with SCR for both REAR and ARAR modelling approaches. But the REAR approach is found to be somewhat better than the ARAR method.

Summarizing the above experiments, we conclude that:

- (1) The value of test statistic based on REAR modelling increases with SCR much faster than that in the ARAR modelling approach.
- (2) For a clutter dominated environment, a receiver based on the REAR modelling approach operates consistently better than that based on the ARAR modelling approach.

Point (2) requires some explanation. We recall the key idea of ARAR method is that the signal under H_1 should be fitted with an AR model of longer order than that under H_0 . When the target is masked by a clutter peak, the received signal under H_1 has the same number of peaks as that under H_0 . Therefore, if model order is suited for the return under H_0 , it is also well suited for H_1 . In other words, the difference between the prediction-error from AR(p) and AR(q) model is small; as a result, the LLR does not increase with SCR.

CHAPTER 4-

LINEAR MODEL BUILDING FOR THE ADAPTIVE DETECTION PROBLEM

Thus far, we have described the method to establish the test statistic for solving the signal detection problem and the relevant recursive algorithm for implementing it. For convenience of description, this method was illustrated in chapter 3 by using two pairs of special linear models, namely, ARAR and REAR modelling. This, however, does not imply either that the choice of the models for solving detection problem has to be confined to two special modelling techniques, or that a REAR model can always be found for any situation.

Apparently, the most important step in the design of model-based receivers is to determine a pair of linear models for the received signal under hypotheses H_0 and H_1 by exploiting the information available in a given environment as much as possible so as to achieve good receiver performance. A straightforward method is to make a guess at a pair of models for hypotheses H_0 and H_1 , evaluate the detection performance by using the adaptive detection algorithm developed in chapter 3, and repeat the computation

until satisfactory models are obtained. This method by trial and error is extremely time-consuming, since statistical evaluation of a system is computationally complicated. We will therefore not pursue this approach further.

It is the objective of the present chapter to provide procedures for linear model building. First of all, a method is suggested to construct linear model candidates. Secondly, rather than evaluating these models one by one, in terms of probabilistic performance, we provide some techniques for initially testing them. Based on the results of the initial test, the poor models are given up and only one or two pairs of models are reserved for final statistical evaluation. The initial test is quite simple and therefore the above method is computationally efficient. The procedures will be illustrated by using the moving target detection problem encountered in an airport surveillance radar system.

4.1 GEOMETRICAL INTERPRETATION OF SIGNAL DETECTION

We will develop a geometric interpretation of the LLR receiver described previously so that we may gain insight into the way the receiver works. This understanding will offer us a basis for the search for good models.

Recall the test statistic (3.24) which is based on REAR modelling, reproduced here for convenience:

$$(4.1a) \quad \hat{\ell}(k) = \frac{n}{\hat{\sigma}_0^2} \hat{\sigma}_0^2 - n \log \hat{\sigma}_1^2$$

$$(4.1b) \quad \hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=k-n+1}^k |y(i) - \hat{a}_p^t y_p(i-1)|^2$$

$$(4.1c) \quad \hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=k-n+1}^k |y(i) - \hat{b}_q^t x(k)|^2$$

If the sample size n is moderate or large, we have approximately

$$(4.2a) \quad \hat{\sigma}_1^2 = E |y(k) - \hat{a}_p^t y_p(k-1)|^2 = E |e_1(k)|^2$$

$$(4.2b) \quad \hat{\sigma}_0^2 = E |y(k) - \hat{b}_q^t x(k)|^2 = E |e_0(k)|^2$$

Let ϕ_1 denote the space spanned by past random samples $y(k-1), y(k-2), \dots, y(k-p)$, and ϕ_0 denote the space spanned by reference samples $x(k), x(k-1), \dots, x(k-q+1)$. Then using the geometric projection explanation of a linear estimation [Doob, 13], [Papoulis, 55], [Anderson and Moore, 4], [Kailath, 35], $\hat{a}_p^t y_p(k-1)$ is the projection of $y(k)$ onto the space ϕ_1 and $E |e_1(k)|^2$ is the squared distance of $y(k)$ to space ϕ_1 . Similarly, $E |e_0(k)|^2$

represents the squared distance measure of $y(k)$ to space ϕ_0 . We note that $\lambda(k)$ in equation (4.1a) is the difference between these two terms, with each of them being a monotonic function of the distance measure. Obviously, for the model-based LLR receiver, the decision is made by comparing the distance of the observation $y(k)$ from space ϕ_0 with that from space ϕ_1 .

It is desirable that when hypothesis H_1^* is true, the test statistic $\lambda(k)$ increases rapidly with signal-to-noise ratio (SNR). This can be done if $E|e_0(k)|^2$ increases rapidly while $E|e_1(k)|^2$ remains basically unchanged with SNR. For ease of description, we introduce two definitions:

Definition 1.

The H_0 -characteristic projection space is defined as a space onto which the noise component of the received signal can project but the signal component does not.

Definition 2.

The H_1 -characteristic projection space is defined as a space onto which the signal plus noise can project.

We call these spaces the H_0 -space and H_1 -space, respectively, for brevity. From the above discussion, the essence of linear model design is to find characteristic projection spaces for the received signal under hypotheses H_0 and H_1 . In this context, it is worthwhile to compare different types of detection problems.

When detecting which one of two known deterministic signal is received in the presence of additive white noise, the deterministic signals are represented by two predefined points in Euclidean space, and a decision is made by comparing the distances of the random point, representing the received noise signal, to these two signal points.

The detection of an unknown random signal in coloured noise with unknown statistics, on the other hand, is much more complicated. The signal under hypotheses H_1 and H_0 cannot be simply represented by two fixed points; rather, it is represented by an unknown vector in the corresponding characteristic projection space. This vector is in effect the projection of the received signal onto the pertinent characteristic space. The distance from the received signal to its projection vector is exactly the same as that from the signal to the space. In this sense, we may also say the signal under hypotheses H_0 and H_1 is represented by its related characteristic spaces. The decision is made by comparing the distances of received

signal to these two spaces; this is illustrated in Figure 4.1. When using an adaptive detection algorithm, the projection and error vector are simultaneously calculated and recursively updated.

4.2 MODEL FOR THE INPUT SIGNAL UNDER HYPOTHESIS H_1

The design of linear model for input data under hypothesis H_1 is comparatively simple. Box and Jenkins [9] show that most practical time series are well fitted by ARMA model with low order. According to the Wold decomposition theorem [13, 77, 59], this ARMA model can itself be approximated by an AR model of appropriate order. Expressing it in mathematical terms, we have

$$(4.3) \quad y(k) = \sum_{i=1}^p a_i y(k-i) + e_1(k)$$

Order p can be determined by using a number of techniques, say, AIC [1, 2, 3, 40, 59], CAT [56-59] and MDL [61, 62, 63, 67]. For radar data under hypothesis H_1 , a typical value for the order p is 5 [49]. The projection space associated with model (4.3) is that spanned by past data samples $y(k-1), \dots, y(k-p)$. However, the AR family is not a unique choice for describing hypothesis H_1 . Using the

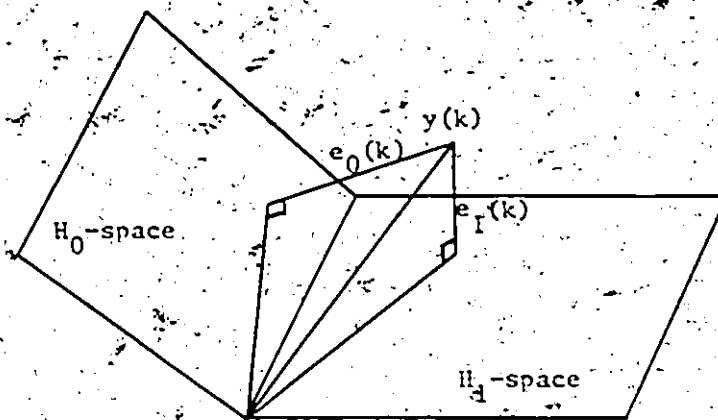


Fig. 4.1 A geometric interpretation of signal detection.

concept of characteristic projection space, we can construct other possible models. The H_1 -space may constitute not only the past observations, but also some reference data. Reconsider the radar data record "WTRNW". One possible model of this kind is described by

$$(4.4) \quad y(k) = [y(k-1), y(k-2), y(k-3), x_{r+1}(k), x_{r-1}(k)] \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} + e_1(k)$$

where $\{y(k)\}$ denotes the target plus clutter data sampled at range ring r , and $\{x_{r-1}(k)\}$ and $\{x_{r+1}(k)\}$ denote the reference samples taken from adjacent rings $r-1$ and $r+1$, respectively. (See Figure 3.2.) The projection space associated with model (4.4) is composed of two subspaces (not necessary to be mutually orthogonal). One subspace is spanned by the past data $y(k-1)$, $y(k-2)$, $y(k-3)$, while the other is spanned by the reference samples $x_{r+1}(k)$ and $x_{r-1}(k)$. In a similar way, we may add or take out some elements in the row vector in the right side of equation (4.4), thereby changing the order of the model. This kind of models is known as the mixed model.

When hypothesis H_1 is true, the received signal

is composed of clutter returns plus target echoes. Since the target signal has no correlation with the reference data, the adaptive algorithm will adjust the model parameters such that the target component project only onto the subspace spanned by the past data, while the clutter component project mainly onto the other one. Consequently, the algorithm has enough degrees of freedom to optimize the model coefficients according to two signal components separately, resulting in a model error process with small variance.

A linear model may be related to a filter with $\{y(k)\}$ as its input and error process as its output. From the discussion presented in chapter 3, we recall that the filter should be able to whiten the input sequence $\{y(k)\}$. Thus, for a model candidate, we should examine whether it performs such an operation. We refer to this procedure as the whitening test. An effective method for performing such a test is to draw the power spectrum of the error process by using the maximum-entropy method (NEM) [25].

Figure 4.2 shows some typical results. The received signal consists of actual radar clutter plus an artificially generated target echo with complex Gaussian envelope. Clutter data used for test are taken from different locations. The signal-to-clutter ratio also takes on different values. These conditions have been

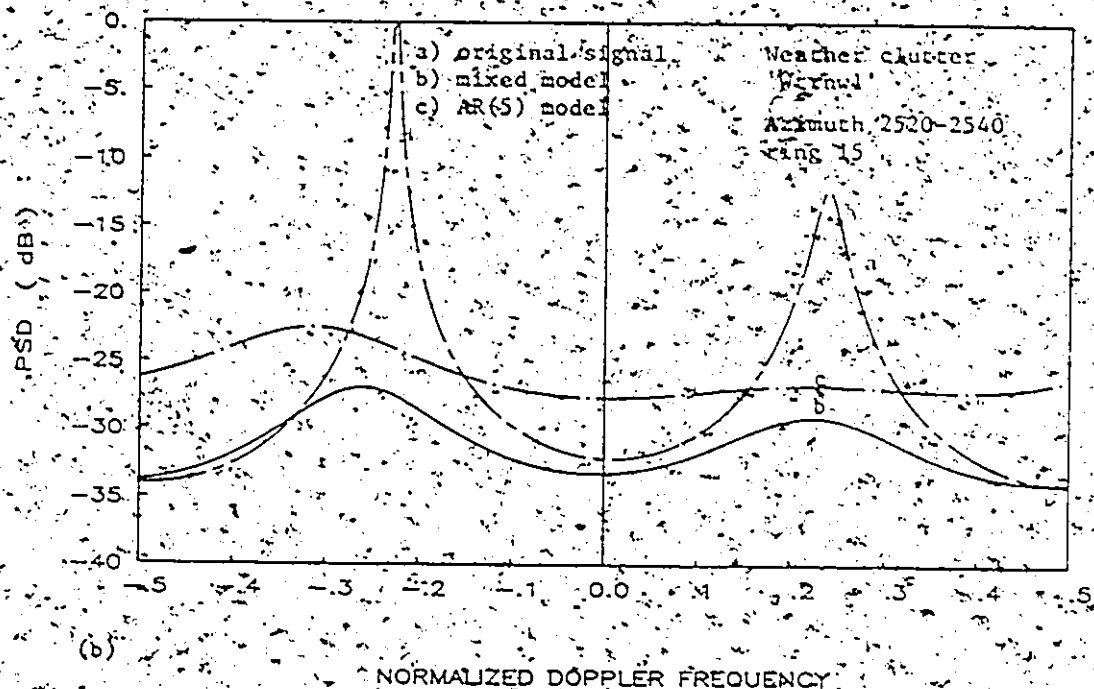
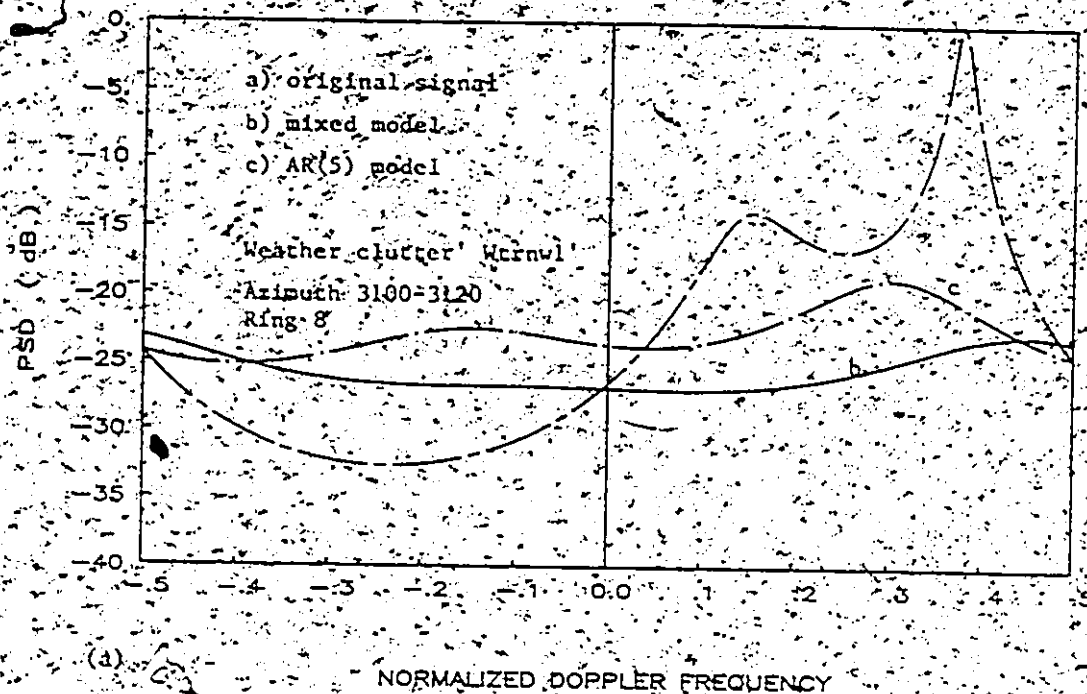


Fig. 4.2: The spectra used for whitening-test of N_1 model candidates.

(a) target Doppler = 250 Hz, SCR = -2 dB.

(b) target Doppler = 150 Hz, SCR = 0 dB.

labelled in the figures. In Figure 4.2a, the target peak is close to the clutter peak in frequency. Figure 4.2b shows the case where target peak is well separated from the clutter peak. We see from Figure 4.2b that the filters, represented by the AR(5) and mixed models, can well decorrelate the input sequence under H_1 hypothesis. The mixed model is comparatively better since it has a flatter spectrum and smaller error power as shown in Figure 4.2a, or similar shape of spectrum but smaller error variance as given in Figure 4.2b. This can be interpreted as follows: In the mixed model, the pertinent space is composed of two subspaces. The first subspace is spanned by the past observation data and the second one, spanned by the samples from the reference sequence. The adaptive algorithm will adjust the model coefficient a_1, a_2, a_3 by projecting the signal component onto the first subspace and adjust coefficient a_4, a_5 by projecting the clutter component mainly onto the second subspace. The adjustment of parameters a_1, a_2, a_3 and adjustment of a_4, a_5 have little effect on each other, and so two sets of parameters can achieve their own optimum values. This makes the error power relatively smaller. It is important to realize that these two adjustments are not completely independent, since the two subspaces are not orthogonal.

For the AR model, however, the model coefficients are affected and adjusted by both components at the same time. The minimum error power cannot be obtained unless a large model order is used.

4.3. MODEL FOR INPUT SIGNAL UNDER H_0

The procedure for building the H_0 -model is much more complicated than that for the H_1 -model, in that we have to proceed as follows:

- (1) Choose model candidates.
- (2) Apply Whitening Test to the error process.
- (3) Compute the η -loss in LLR due to erroneous estimation of H_0 -model parameters when H_1 is true.

We will illustrate the procedures by considering the record of radar data 'WTRNW1' again. We make use of the correlation of received signal under H_0 hypothesis with those from adjacent range rings. We may project $y(k)$, the presently received sample, either onto the space spanned by the data on one side of reference rings, or onto the space spanned by reference data from both side rings.

A few possible candidates for H_0 -model are given below.

$$(1) \quad y(k) = [b_1, b_2] \begin{bmatrix} x_{r+1}(k) \\ x_{r+1}(k-1) \end{bmatrix} + e_0(k)$$

$$(2) \quad y(k) = [b_1, b_2] \begin{bmatrix} x_{r-1}(k) \\ x_{r-1}(k-1) \end{bmatrix} + e_0(k)$$

$$(3) \quad y(k) = [b_1, b_2] \begin{bmatrix} x_{r+1}(k) \\ x_{r-1}(k) \end{bmatrix} + e_0(k)$$

$$(4) \quad y(k) = [b_1, b_2, b_3, b_4] \begin{bmatrix} x_{r+1}(k) \\ x_{r+1}(k-1) \\ x_{r-1}(k) \\ x_{r-1}(k-1) \end{bmatrix} + e_0(k)$$

$$(5) \quad y(k) = [b_1, b_2, b_3, b_4] \begin{bmatrix} x_{r+1}(k) \\ x_{r+2}(k) \\ x_{r-1}(k) \\ x_{r-2}(k) \end{bmatrix} + e_0(k)$$

Also, we may project $y(k)$ onto the space spanned by past data in a given range ring, resulting in an AR model. These models are included for the purpose of comparison, although they obviously do not satisfy the

conditions for H_0 characteristic projection space. As examples, we may thus consider the following:

$$(6) \quad y(k) = b_1 y(k-1) + e_0(k)$$

$$(7) \quad \hat{y}(k) = [b_1, b_2] \begin{bmatrix} y(k-1) \\ y(k-2) \end{bmatrix} + e_0(k)$$

The degree of independence of the error process can be intuitively shown by its MEM spectrum. Fig. 4.3a is for the data located between azimuth 3100 to 3120 on range ring 22 while Figure 4.3b for data located between azimuth 2520 to 2540 on range ring 15.

Figure 4.3 shows the spectra for the error process resulting from model (2-7); model (1) is similar to (2) and ignored from plotting. In Figure 4.3 we show the spectra for models (2) and (6), for the purpose of comparison. In both figures, the spectrum of the original signal is included.

Figure 4.3 shows that models (1), (2), obtained by projecting the received signal $y(k)$ onto the space spanned by the data sample from only one adjacent ring, are very poor in their whitening capability; we may therefore discard them.

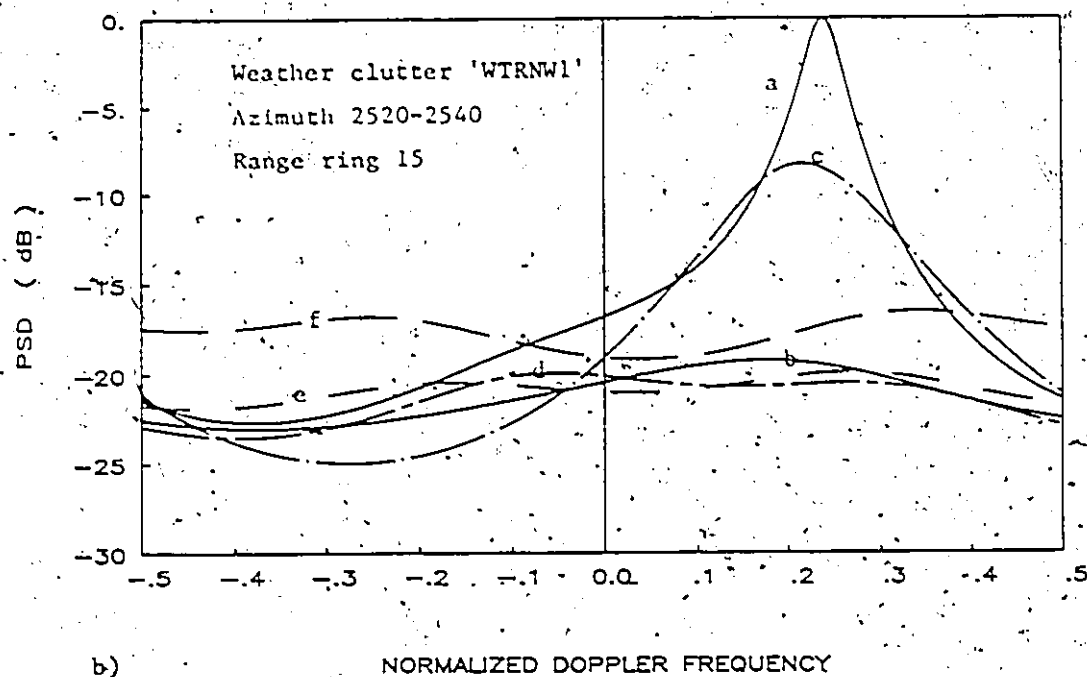
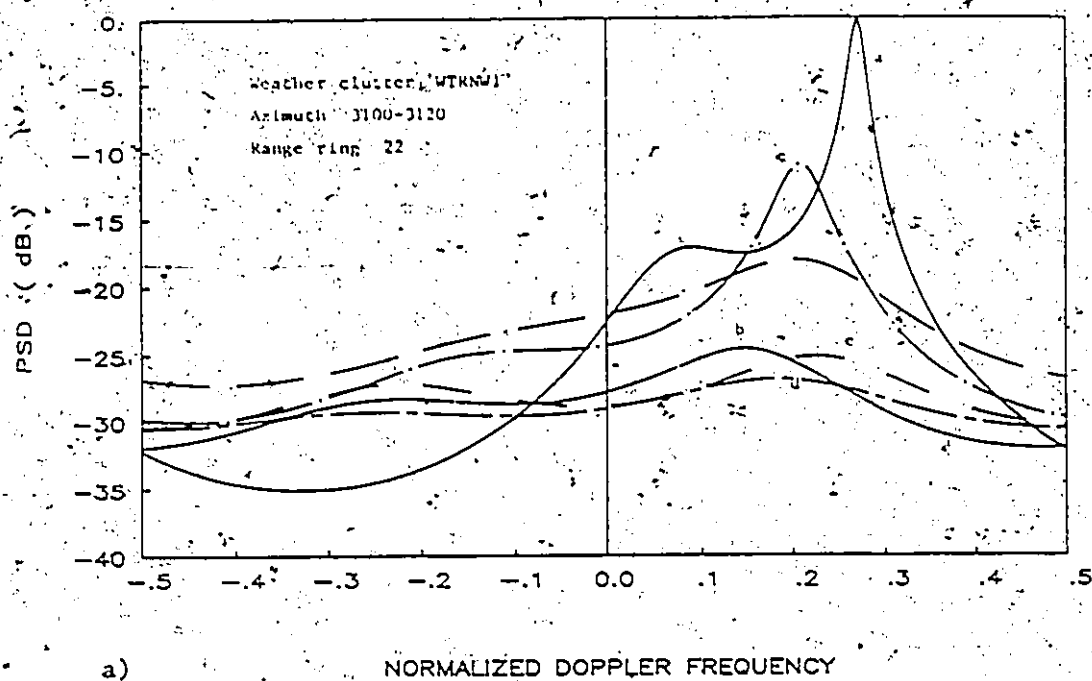


Fig. 4.3. The MEM spectra used for whitening-test of error processes in different models.

Curve a: for original signal

Curve b: for model (3)

Curve c: for model (2)

Curve d: for model (4)

Curve e: for model (5)

Curve f: for model (6)

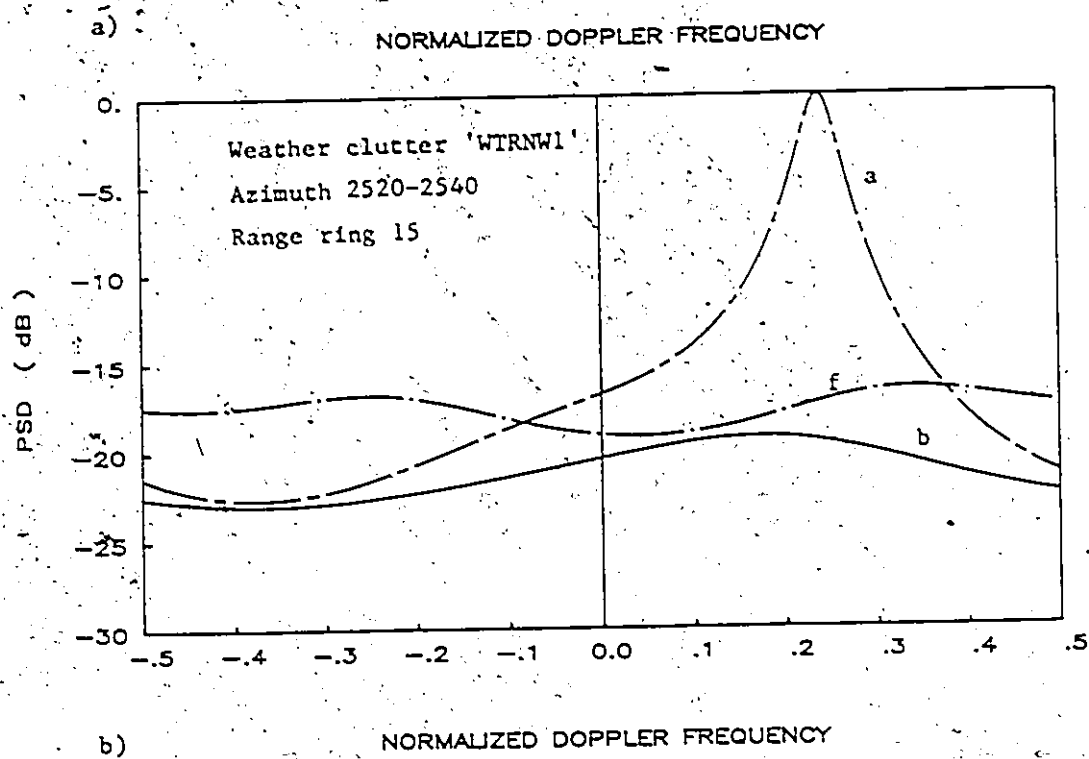
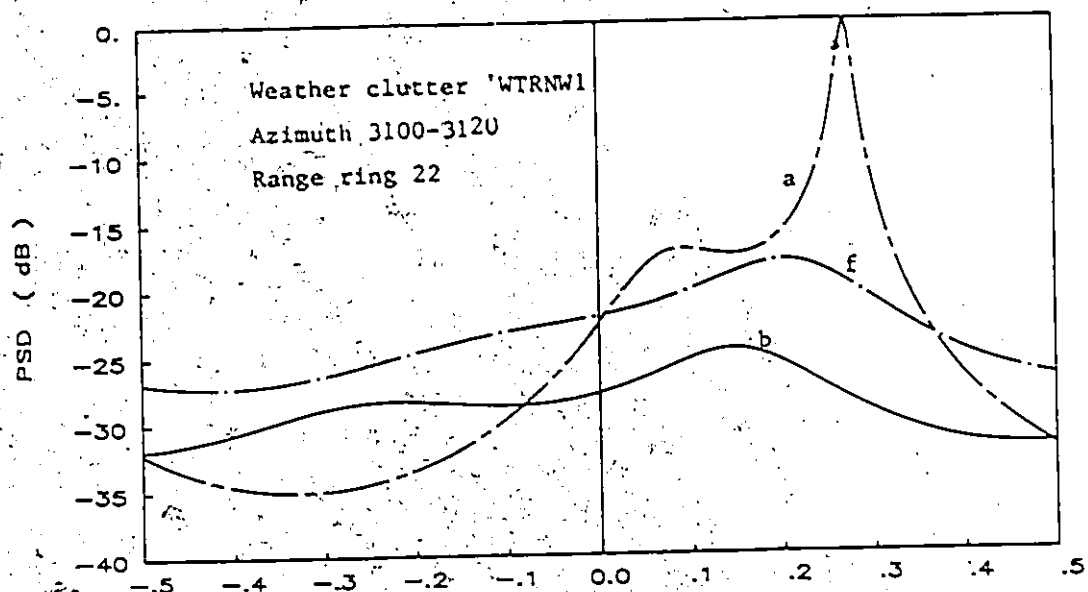


Fig. 4.4 The MEM spectra used for whitening test

Curve a: for original signal

Curve b: for model (3)

Curve f: for model (6)

Model (6), i.e., autoregressive model of order one, is rather better than models (1), (2), but inferior to models (3), (4), (5). Accordingly, models (3), (4), (5) and (6) will be retained for the next step, namely the η -loss test. For clarity, the results of whitening test for models (3) and (6) are redrawn in Figure 4.4.

The η -loss test is conducted in accordance with formula (3.78), reproduced here for convenience:

$$\eta = \ln \left[\frac{p(\underline{y} | \underline{b}_q(\underline{y} | H_0), \sigma_0^2)}{p(\underline{y} | \underline{b}_q(\underline{y} | H_1), \sigma_0^2)} \right]$$

An artificial target signal is used during the test.

Coefficient $\underline{b}_q(\underline{y} | H_0)$ is found and stored when H_0 is true, and then $\underline{b}_q(\underline{y} | H_1)$ is estimated when the target signal is added to the clutter data. The η -test is performed for the radar data located at different parts of the record.

Figure 4.5 shows typical results. It is seen that the losses in LLR for the LLR test based on models (3), (4), (5), coincide with each other and are practically independent of SCR.

For the autoregressive models (6), (7), namely, AR(1) and AR(2), on the other hand, the η -loss increases rapidly with SCR. This is in good agreement with our previous theoretical analysis given in Section 3.5. We also note that when order q of the AR model increases, so does the

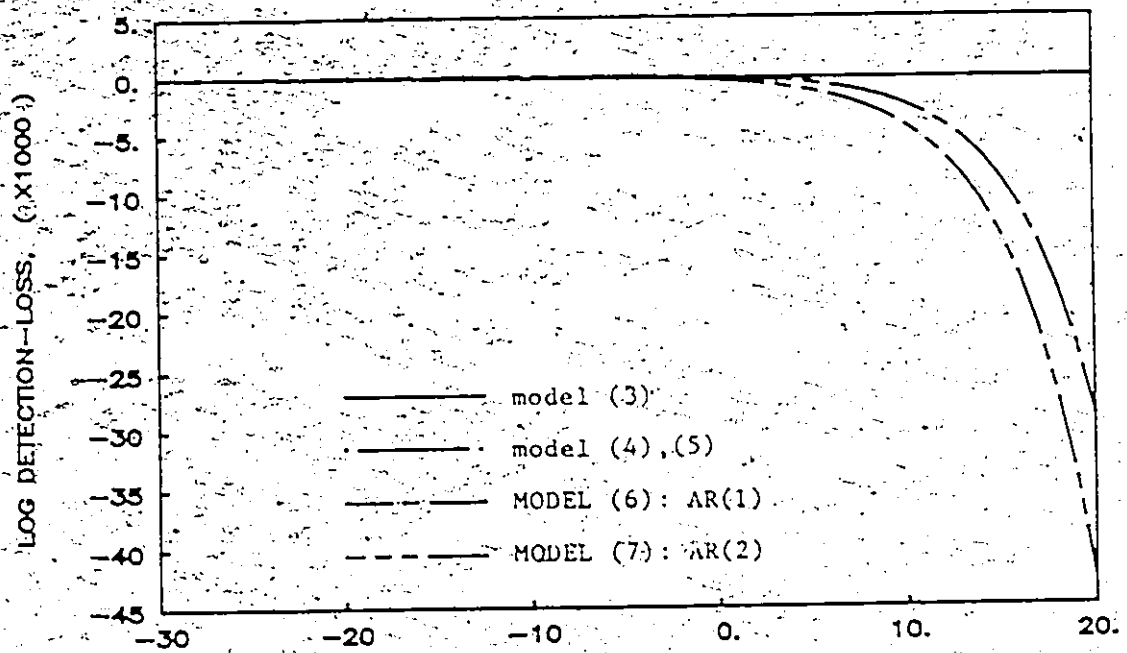


Fig. 4.5. χ^2 -loss test for H_0 -model candidates

the η -loss. As such, we conclude that the autoregressive model is not a good model for the received signal under hypothesis H_0 if an LLR receiver of high quality is to be constructed.

Summarizing the above discussion, the models suited for building a good radar receiver are:

Model I

$$H_1\text{-Model: } y(k) = [a_1, a_2, a_3, a_4, a_5] \begin{bmatrix} y(k-1) \\ y(k-2) \\ y(k-3) \\ x_{r+1}(k) \\ x_{r-1}(k) \end{bmatrix} + e_1(k)$$

$$H_0\text{-Model: } y(k) = [b_1, b_2] \begin{bmatrix} x_{r+1}(k) \\ x_{r-1}(k) \end{bmatrix} + e_0(k)$$

Model II

$$H_1\text{-Model: } y(k) = [a_1, a_2, a_3, a_4, a_5] \begin{bmatrix} y(k-1) \\ y(k-2) \\ y(k-3) \\ x_{r+1}(k) \\ x_{r-1}(k) \end{bmatrix} + e_1(k)$$

$$H_0\text{-Model: } y(k) = [b_1, b_2, b_3, b_4] \begin{bmatrix} x_{r+1}(k) \\ x_{r+2}(k) \\ x_{r-1}(k) \\ x_{r-2}(k) \end{bmatrix} + e_0(k)$$

$$H_1\text{-Model: } y(k) = [a_1, a_2, a_3, a_4, a_5]$$

$$\begin{bmatrix} y(k-1) \\ y(k-2) \\ y(k-3) \\ x_{r+1}(k) \\ x_{r-1}(k) \end{bmatrix} + e_1(k)$$

Model III

$$H_0\text{-Model: } y(k) = [b_1, b_2, b_3, b_4]$$

$$\begin{bmatrix} x_{r+1}(k) \\ x_{r+1}(k-1) \\ x_{r-1}(k) \\ x_{r-1}(k-1) \end{bmatrix} + e_0(k)$$

As a matter of fact, a receiver of same quality level should be achieved if we replace the H_1 -model used above by an AR(5) model.

Now we show in Figure 4.6 the curves of normalized LLR versus target Doppler frequency given for the LLR detection receiver based on linear modelling I and II. As a comparison, ARAR modelling (H_0 -model: AR(1); H_1 -model: AR(5)) is also included. In these experiments the SCR is set to be 0dB. From the experimental results shown in Figure 4.6, we observe that:

- (1) Irrespective of the Doppler frequency of the target, the normalized LLR based on models I and II is nearly always better than that based on ARAR modelling. This confirms the validity of η -loss test as an aid in model building.

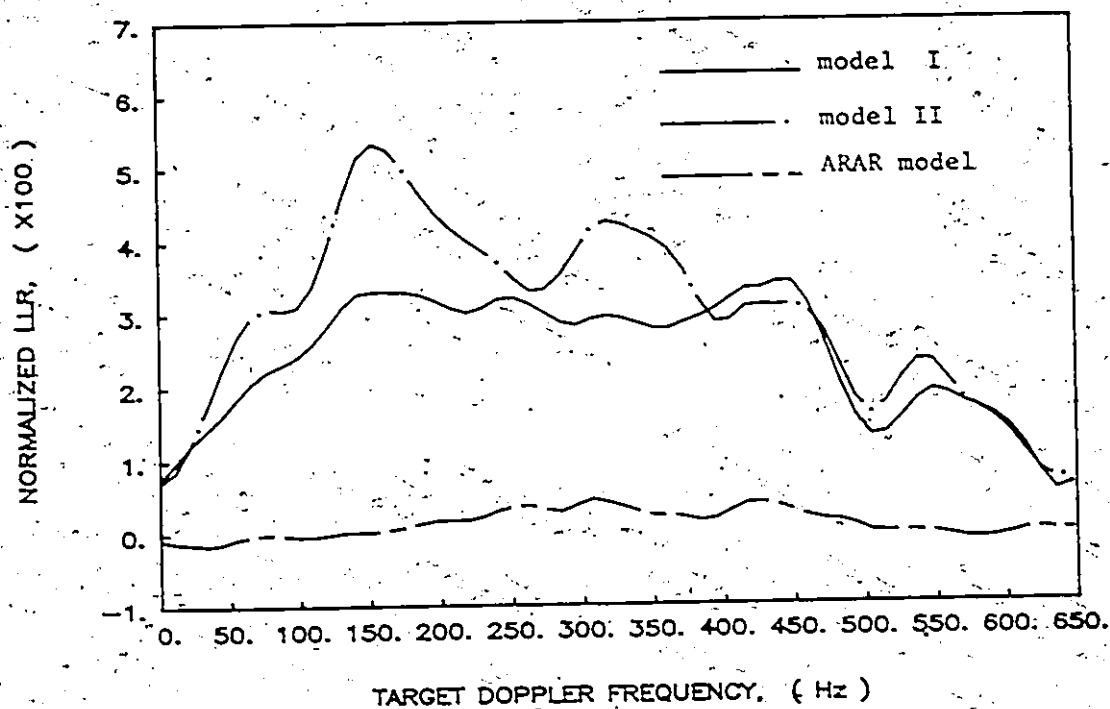


Fig. 4.6 Comparison of LLR v.s. target Doppler curves for different schemes.

(a) Clutter data taken from azimuths 1400-1420 on ring 18 of record 'WTRNW1'.

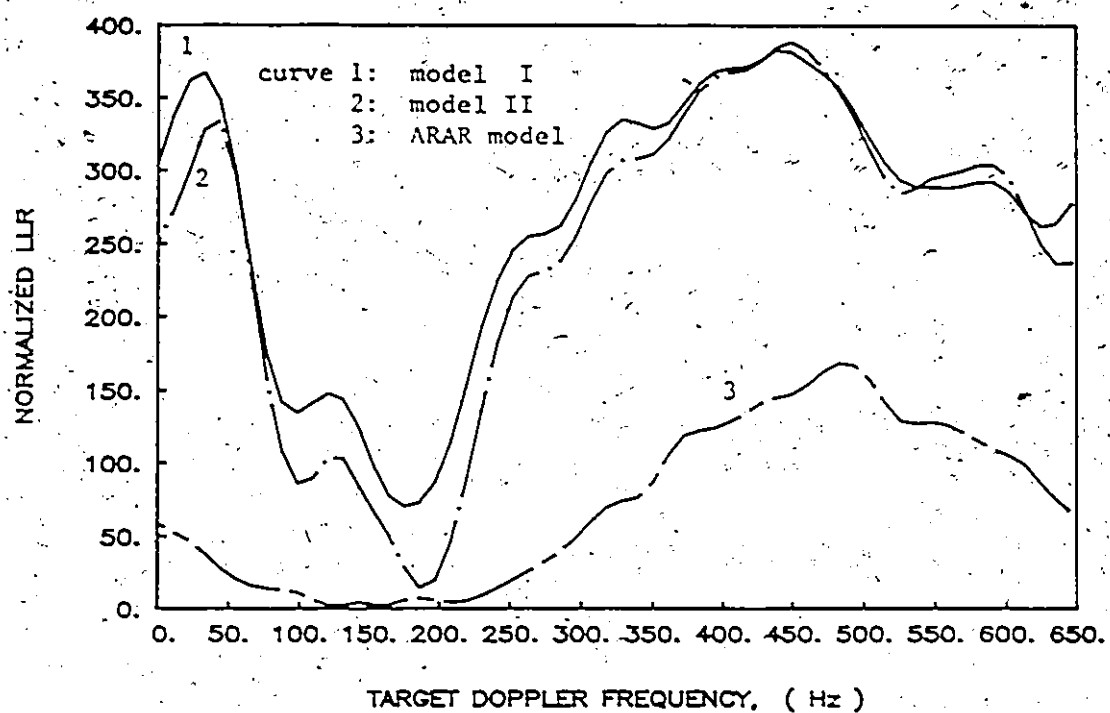


Fig. 4.6 Comparison of LLR v.s. target Doppler curves for different schemes.

(b) Weather clutter data: azimuth 3150-3170, ring 17: 'WTRNWI'

(2) Modelling II seems somewhat better than model I; however, considering computational complexity, Model I is the preferred one.

4.4 SUMMARIZING REMARKS

In the last section, it was shown that radar data are well fitted by a regressive model. Such a linear model may be viewed as a linear filter with the received signal as input and the output as a white error sequence. In some situations, however, the linear filter so obtained does not orthogonalize the input process under hypothesis H_0 . This violates the basic assumption on which the test statistic (3.24) was established and therefore suffers from a loss in detection performance. As pointed out by T. Kailath [33]: "The mathematical model may only be an idealization to varying degrees of the physical problem,.... The major aim of most mathematical analyses is to give insight into the structure of the solution; such insight can then be used to intelligently modify and adapt the mathematical solution to a particular physical solution".

Experiments show that a regressive model can be successfully used as the H_0 -model and is acceptable if the error process passes the whitening test. However, if this

is not the case, the model may be modified, using the following method:

- (1) Project the newly received sample, $y(k)$, onto the space spanned by the reference data:

$$x(k), x(k-1), \dots, x(k-q_1+1)$$

and compute the power spectrum of the error process $e_0(k)$ of the linear model:

$$y(k) = \underline{x}(k)^t \underline{b}_{q_1} + e_0(k)$$

where $\underline{x}(k) = (x(k), x(k-1), \dots, x(k-q_1+1))^t$

$$\underline{b}_{q_1} = [b_1, b_2, \dots, b_{q_1}]^t$$

- (2) If the error process is not white, try one by one, the H_0 -models with vector $\underline{x}(k)$ composed of a mixture as follows:

$$x(k), x(k-1), \dots, x(k-q_1+1), y(k-q_2)$$

$$x(k), x(k-1), \dots, x(k-q_1+1), y(k-q_2), y(k-q_2+1)$$

$$x(k), x(k-1), \dots, x(k-q_1+1), y(k-q_2), y(k-q_2+1), \dots, y(k-1)$$

The model thus constructed is called a mixed model.

Since the reference data are correlated with the clutter component of received sequence, their use results in a small η -loss. Hence, we try to make use of them first in building the model. Then, we add some past observation data in the vector $x(k)$ and apply the whitening test to the data. We repeat the procedure until the whitening test indicates the pertinent error is white.

As mentioned, an autoregressive model of high enough order can always be found to decorrelate the input signal; thus by adding past observation data continually, it is possible to find a model which satisfies the whitening condition. Adding more past data, improves the whitening capability of the linear filter, but in so doing, the η -loss and computational complexity will increase. There is a compromise between the whitening capability of the filter, on the one hand, and the η -loss and the computational burden on the other. The parameter q_1 and q_2 can be properly adjusted for this purpose.

The ARAR and REAR modelling approaches addressed in chapter 3 are apparently two special cases of the theory developed in this chapter. The key point of modelling design for solving the detection problem is to make full use of the information available to the designer. The final choice completely depends on the given detection

environment. When a reference sequence exists, the first choice for H_0 -model is a regressive model, followed by a mixed model. When no reference is available, however, ARAR modelling is the only and final choice.

CHAPTER 5

APPLICATION TO RADAR DETECTION

In previous chapters, we have developed the adaptive detection algorithm and procedures for model building. In this chapter, we apply this theory to solve the detection of a moving target echo in an airport surveillance radar environment. The radar detection problem may be described by:

$$(5.1a) \quad H_0\text{-Model: } y(k) = [b_1, b_2] \begin{bmatrix} x_{r-1}(k) \\ x_{r+1}(k) \end{bmatrix} + e_0(k)$$

$$(5.1b) \quad H_1\text{-Model: } y(k) = [a_1, a_2, a_3, a_4, a_5] \begin{bmatrix} y(k-1) \\ y(k-2) \\ y(k-3) \\ x_{r-1}(k) \\ x_{r+1}(k) \end{bmatrix} + e_1(k)$$

It is the purpose of this chapter to perform a statistical evaluation of the new scheme and discuss the related implementation problems.

Section 5.1 discusses the extraction of radar

target information involved by radar clutter. Section 5.2 summarizes the adaptive detection algorithm. The choice of initial values, computational complexity and thresholding problem are discussed in Sections 5.3, 5.4 and 5.5, respectively. Finally, the detection performances obtained by using actual and simulated radar data are described in Sections 5.6 and 5.7.

5.1 INFORMATION INVOLVED IN RADAR CLUTTER

In previous chapters, we mentioned about the spatial correlation of radar data and used it to set up regressive and autoregressive models for detection problems. In this section, we will discuss the problem of extracting useful information from radar clutter in more detail.

Radar clutter is defined [31] as "confused unwanted echoes on the radar display". The name is descriptive of the fact that such echoes tend to "clutter" the radar display and make detection of wanted (target) echoes more difficult. In a radar surveillance environment, clutter arises due to reflections from a variety of objects: (a) fixed or slowly moving objects such as trees, vegetation, hills and man-made structures;

(b) storm clouds, precipitation and other meteorological phenomena, and (c), migrating flocks of birds.

Radar Clutter is treated in different ways for different problems. For radar detection problems, the radar clutter is considered as a harmful signal and must be suppressed as much as possible. For radar clutter classification problems, on the other hand, clutter must be studied rather than simply cancelled, so that a good classifier is obtained. In either case, an important step is to analyze the characteristics of different types of clutter. Clutter consists of radar returns from scatterers and the echo waveform is strongly influenced by the geometric and physical characteristics of the scatterers. Thus, clutter echoes carry some kind of information about the scatterers. From the point of view of signal transmission, this process is equivalent to the modulation of the transmitted radar pulses by the scatterers in some unknown fashion. Exploitation of the information content of clutter is very useful for designing a radar signal detector or clutter classifier. A widely used technique for extracting this information is the analysis of the second-order statistics of clutter. In particular, power spectrum analysis has been successfully used in radar clutter classification [11, 27, 28, 39]. The peak

location and spectral band width are intimately connected with the velocities of the scatterers and can be used as signatures to classify storm cloud, migrating flocks of birds, and ground clutter, etc. [28, 39]. Under the autoregressive process assumption of radar echoes, autoregressive model coefficients (or correspondingly, partial coefficients) are equivalent to the power spectrum, as such they may be used for classification of radar clutters [70, 71].

Another form of the second-order statistic is autocorrelation function. According to Wiener-Khinchine relations [55], the autocorrelation function contains the same information as the power spectrum for a wide sense stationary process, since they are linearly related. The spatial correlation of radar clutter has been investigated experimentally by Miller [50]. In [50], it is shown that:

- (1) A large degree of correlation is present in radar data, and much of this correlation is inherent in the radar clutter.
- (2) The autocorrelation function of the clutter on a single scan may be decomposed into a static spatial component and a fluctuating temporal component. The spatial component is dominant and reflects the correlation between the feature of the scatterers, and the temporal component

is directly attributable to the radar beamshape.

Similar results are observed by Kesler [39], who shows that the power spectrum of the radar clutter data taken in range is nearly the same as that of the radar sequence taken in azimuth. This is easily understood since both sets of radar data reflect an identical spatial feature of the scatterers.

We are concerned with the correlation between data sequences taken from adjacent rings in a radar scan. Data sets sampled on range rings r and $r+1$ are denoted by $\{x_r(k), k=1, \dots, N\}$ and $\{x_{r+1}(k), k=1, \dots, N\}$, respectively. The correlation coefficient of the two sets of data can be estimated by

$$(5.2) \quad \rho_{r,r+1}(\tau) = \frac{\frac{1}{N-\tau} \sum_{k=1}^{N-\tau} [x_r(k) - \bar{x}_r][x_{r+1}(k+\tau) - \bar{x}_{r+1}]^*}{\sqrt{\text{var}(x_r) \text{var}(x_{r+1})}}$$

where

$$(5.3a) \quad \text{var}(x_s) = \frac{1}{N} \sum_{k=1}^N |x_s(k) - \bar{x}_s|^2$$

$$(5.3b) \quad \bar{x}_s = \frac{1}{N} \sum_{k=1}^N x_s(k) \quad s = r, r+1$$

Some typical results are shown in Figures 5.1 to 5.4.

Figures 5.1 and 5.2 are shown for ground clutter and

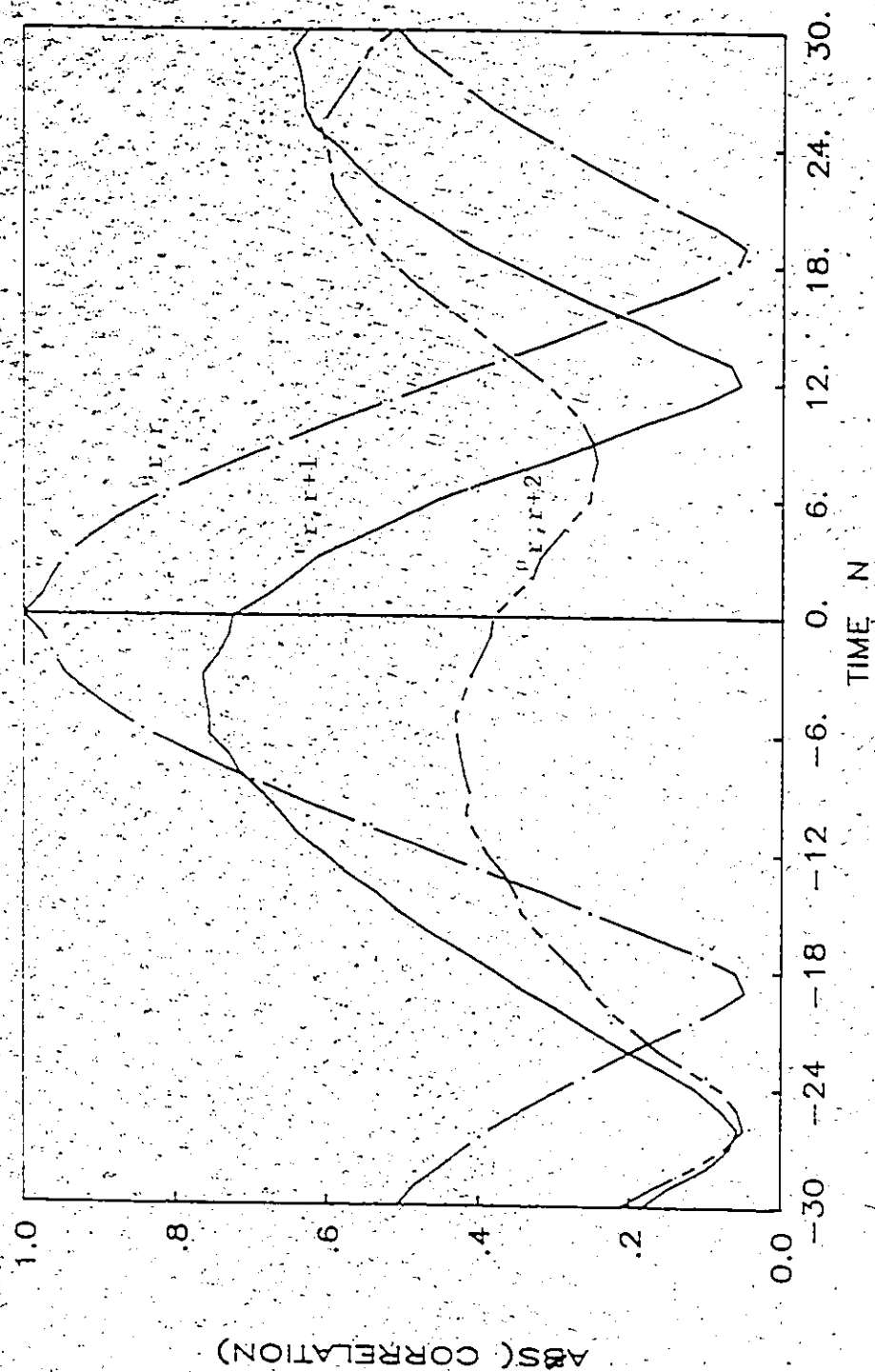


FIG. 5.1: The correlation coefficient (amplitude) between the ground clutter samples from adjacent range rings as a function of the time lag.

Azimuths. 1350-1480

Range ring 27.

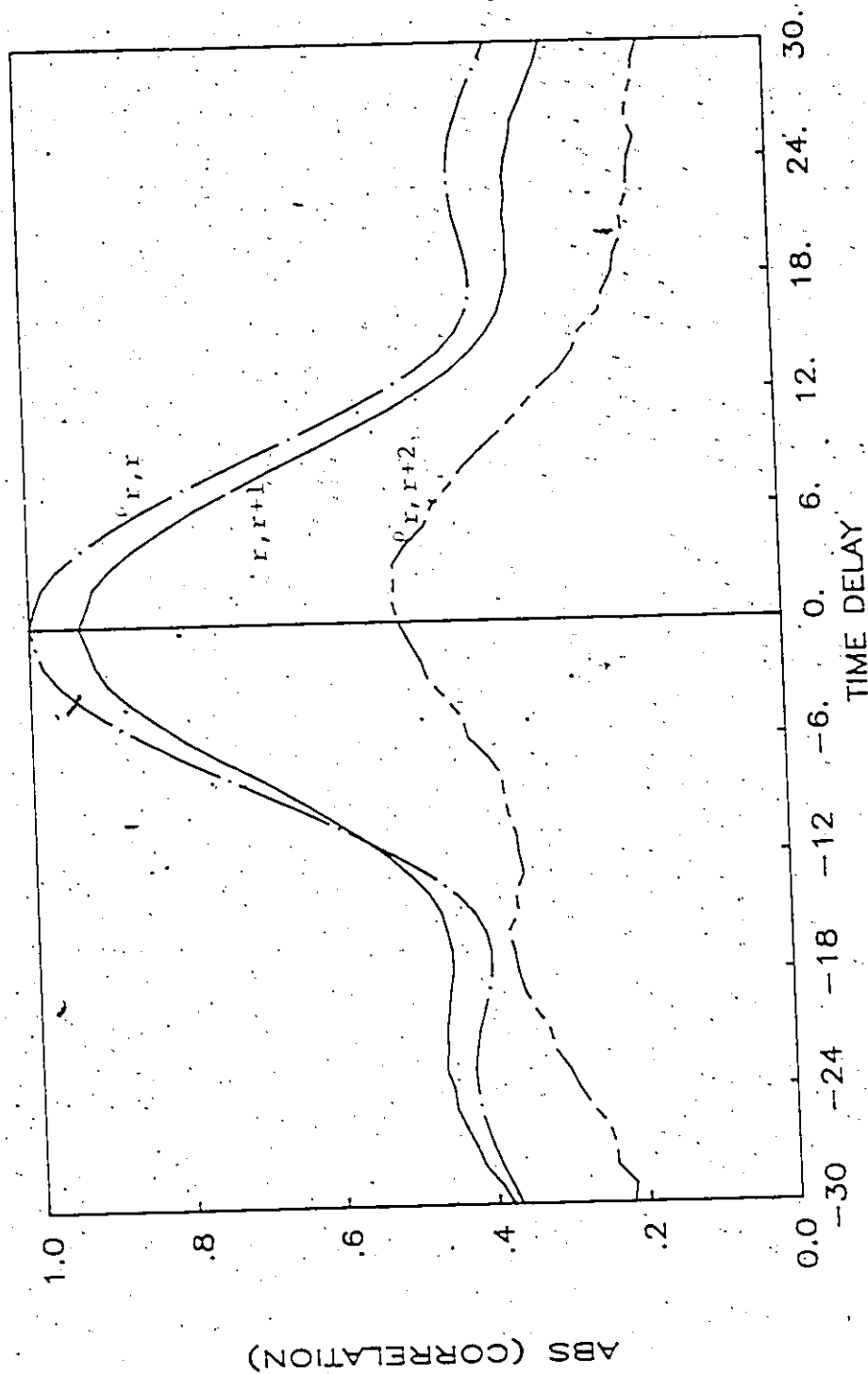


Fig. 5.2: The correlation coefficient (amplitude) between the ground clutter samples from adjacent range rings as a function of the time lag.

Azimutha 2550-2680

Range ring 11

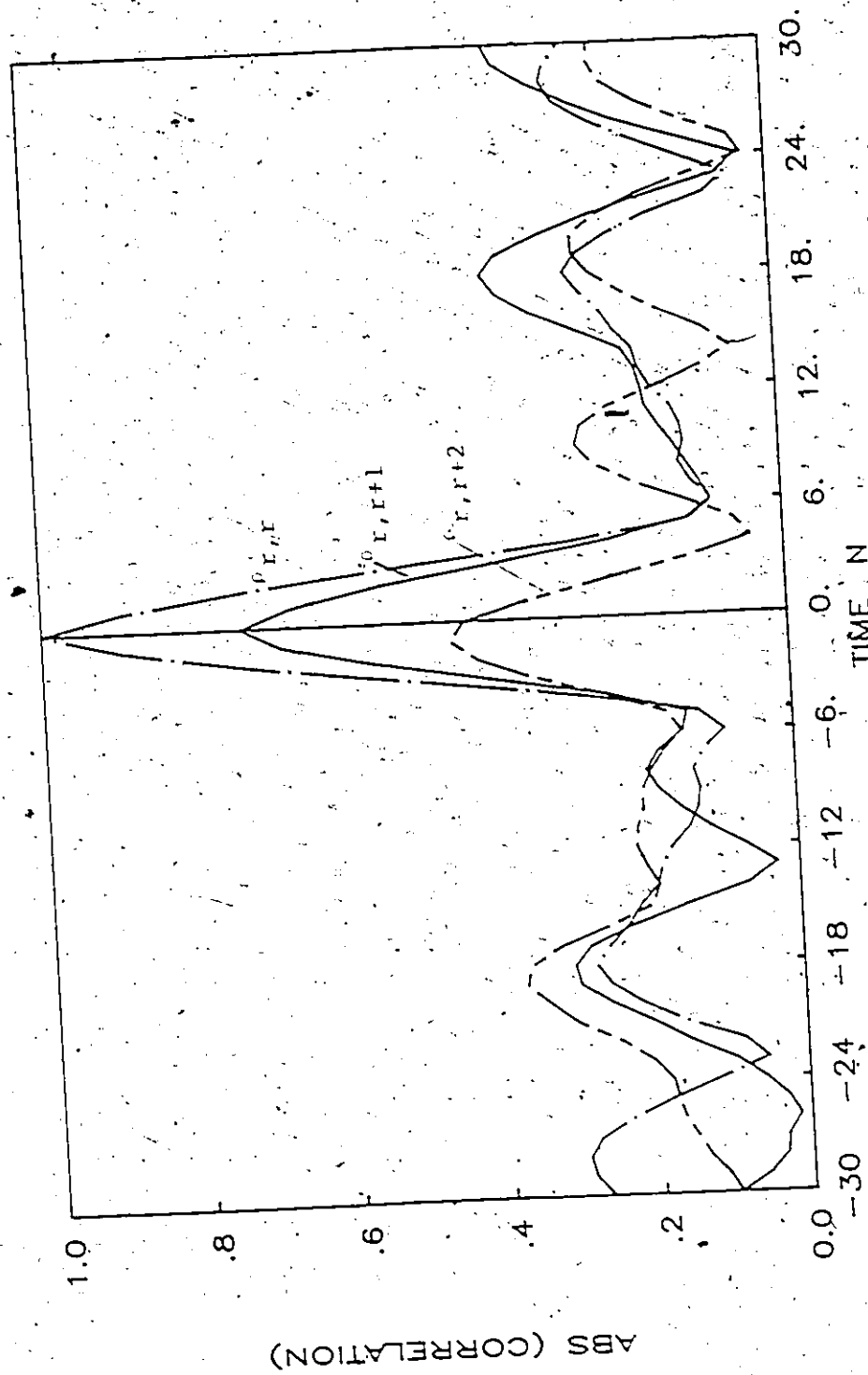


Fig. 3.3: The correlation coefficient (amplitude) between the weather clutter samples from adjacent range rings as a function of time lag.

Azimuths 3100-3230
Range ring 30

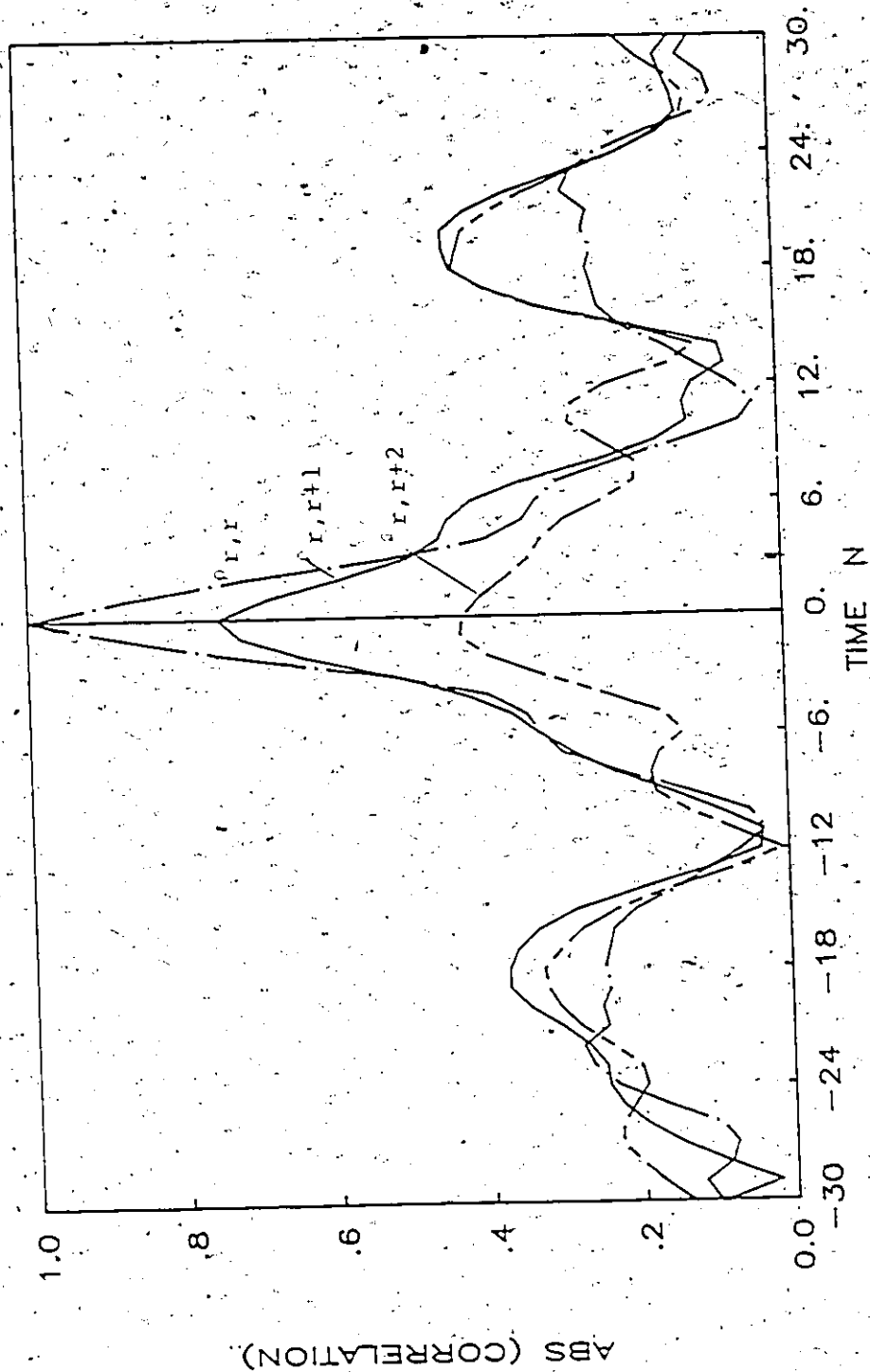


Fig. 5.4: The correlation coefficient (amplitude) between the weather clutter samples from adjacent range rings as a function of the time lag.

Azimuths 2450-2580,
Range ring 12

and Figures 5.3 and 5.4 for weather clutter. The sample size used is $N = 130$, and the data sampling interval (0.74 $\mu\text{sec.}$) is somewhat greater than the radar pulse width (0.5 $\mu\text{sec.}$). These diagrams show that the width of the correlation function for ground clutter is markedly larger than that for weather clutter. This fact implies that weather clutter has a wider spectrum than ground clutter, which agrees with well-known experimental results [28, 39].

From the above discussion, we can make the following observations:

- (1) Radar clutter contains useful information about the target illuminated by the radar antenna.
- (2) Such information is inherent in radar clutter and characterizes the geometry and/or physics of the scatterers.
- (3) We can extract such information by using a second-order statistic in the form of power spectrum, or correlation function, or reflection coefficients, etc.

5.2 SUMMARY OF THE ADAPTIVE DETECTION ALGORITHM

The adaptive detection scheme contains two Kalman filters, a test statistic computer and a decision device.

The first filter, assuming H_0 -model, transforms the received sequence into the error process $\{e_0(k|k-1)\}$. The second filter is used to estimate the unknown parameters of H_1 -model and output the coefficient $f^1(k)$ and error process $\{e_1(k|k-1)\}$.

Upon each newly received input $y(k)$, the Kalman filters calculate $e_0(k|k-1)$, $e_1(k|k-1)$ and $f^1(k)$. Employing these results, along with $\sigma_0^2(k)$ from the clutter map, the test statistic $l(k)$ is updated by the following algorithm:

- 1) Compute $e_0(k|k-1)$ using Kalman Filter 1.

$$(5.4a) \quad \underline{v}^0(k) = \underline{P}^0(k) \underline{x}^*(k)$$

$$(5.4b) \quad f^0(k) = 1 + \underline{x}^t(k) \underline{v}^0(k)$$

$$(5.4c) \quad \underline{g}^0(k) = \underline{v}^0(k) / f^0(k)$$

$$(5.4d) \quad \underline{s}^0(k) = \underline{p}^0(k) - g^0(k) \underline{v}^0(k)^H$$

$$(5.4e) \quad \underline{p}^0(k+1) = \underline{s}^0(k) + \underline{Q}$$

$$(5.4f) \quad e_0(k|k-1) = y(k) - \underline{x}^t(k) \underline{b}_q(k|k-1)$$

$$(5.4g) \quad \underline{b}_q(k+1|k) = \underline{b}_q(k|k-1) + g^0(k) e_0(k|k-1)$$

2) Compute $f^1(k)$ and $e_1(k|k-1)$ using Kalman Filter 2.

$$(5.5a) \quad \underline{v}^1(k) = \underline{p}^1(k) \underline{v}_p^*(k-1)$$

$$(5.5b) \quad f^1(k) = 1 + \underline{v}_p^t(k-1) \underline{v}^1(k)$$

$$(5.5c) \quad \underline{g}^1(k) = \underline{v}^1(k) / f^1(k)$$

$$(5.5d) \quad \underline{s}^1(k) = \underline{p}^1(k) - \underline{g}^1(k) \underline{v}^1(k)^H$$

$$(5.5e) \quad \underline{p}^1(k+1) = \underline{s}^1(k) + \underline{Q}^1$$

$$(5.5f) \quad e_1(k|k-1) = y(k) - \underline{v}_p^t(k-1) \underline{a}_p(k|k-1)$$

$$(5.5g) \quad \underline{a}_p(k+1|k) = \underline{a}_p(k|k-1) + \underline{g}^1(k) e_1(k|k-1)$$

3) Update the Test Statistic $\lambda(k)$

$$(5.6a) \quad \alpha(k) = \alpha(k-1) + \frac{|e_0(k|k-1)|^2}{\sigma_0^2(k|k-1)} - \frac{|e_0(k-n|k-n-1)|^2}{\sigma_0^2(k-n|k-n-1)}$$

$$(5.6b) \quad \beta(k) = \beta(k-1) + \log \frac{\sigma_0^2(k|k-1)}{\sigma_0^2(k-n|k-n-1)}$$

$$(5.6c) \quad \theta(k) = \theta(k-1) + \log \frac{f^1(k)}{f^1(k-n)}$$

$$(5.6d) \quad \gamma(k) = \gamma(k-1) + \frac{1}{n} \frac{|e_1(k|k-1)|^2}{f^1(k)} - \frac{1}{n} \frac{|e_1(k-n|k-n-1)|^2}{f^1(k-n)}$$

$$(5.6e) \quad \lambda(k) = \alpha(k) + \beta(k) - \theta(k) - n \log \gamma(k)$$

4) Decision-Making

$$(5.7) \quad \lambda(k) \geq \text{threshold} \quad \begin{array}{l} H_1 \text{ is true} \\ H_0 \text{ is true} \end{array}$$

5) Initialize the Parameters

$$\underline{p}^0(0) = \text{diag}(1, 1, \dots, 1)$$

$$\underline{p}^1(0) = \text{diag}(1, 1, \dots, 1)$$

$$\underline{b}_q(0|-1) = \underline{0}$$

$$\underline{a}_p(0|-1) = \underline{0}$$

$$\underline{y}_p(0) = \underline{0}$$

$$\underline{x}(0) = \underline{0}$$

$$\alpha(0) = 0$$

$$\beta(0) = 0$$

$$\gamma(0) = 0$$

$$\theta(0) = 0$$

The block diagram for implementing this detection scheme is given in Figure 5.5.

5.3 DETERMINATION OF MATRICES $\underline{Q}^0$ AND $\underline{Q}^1$

To account for nonstationarity, the tap-weight vector is modelled as a random walk process. This incurs additional unknowns $\underline{Q}^0$ and $\underline{Q}^1$, as shown in the above algorithm. $\underline{Q}^0$ and $\underline{Q}^1$ can be estimated by using the maximum-likelihood technique, but unfortunately it leads to a complicated nonlinear optimization problem and makes the adaptive detection algorithm difficult to implement.

The experiments using actual radar data show that generalized LR is not sensitive to deviations of $\underline{Q}^0$ and $\underline{Q}^1$ from their maximum-likelihood estimates. Thus as a

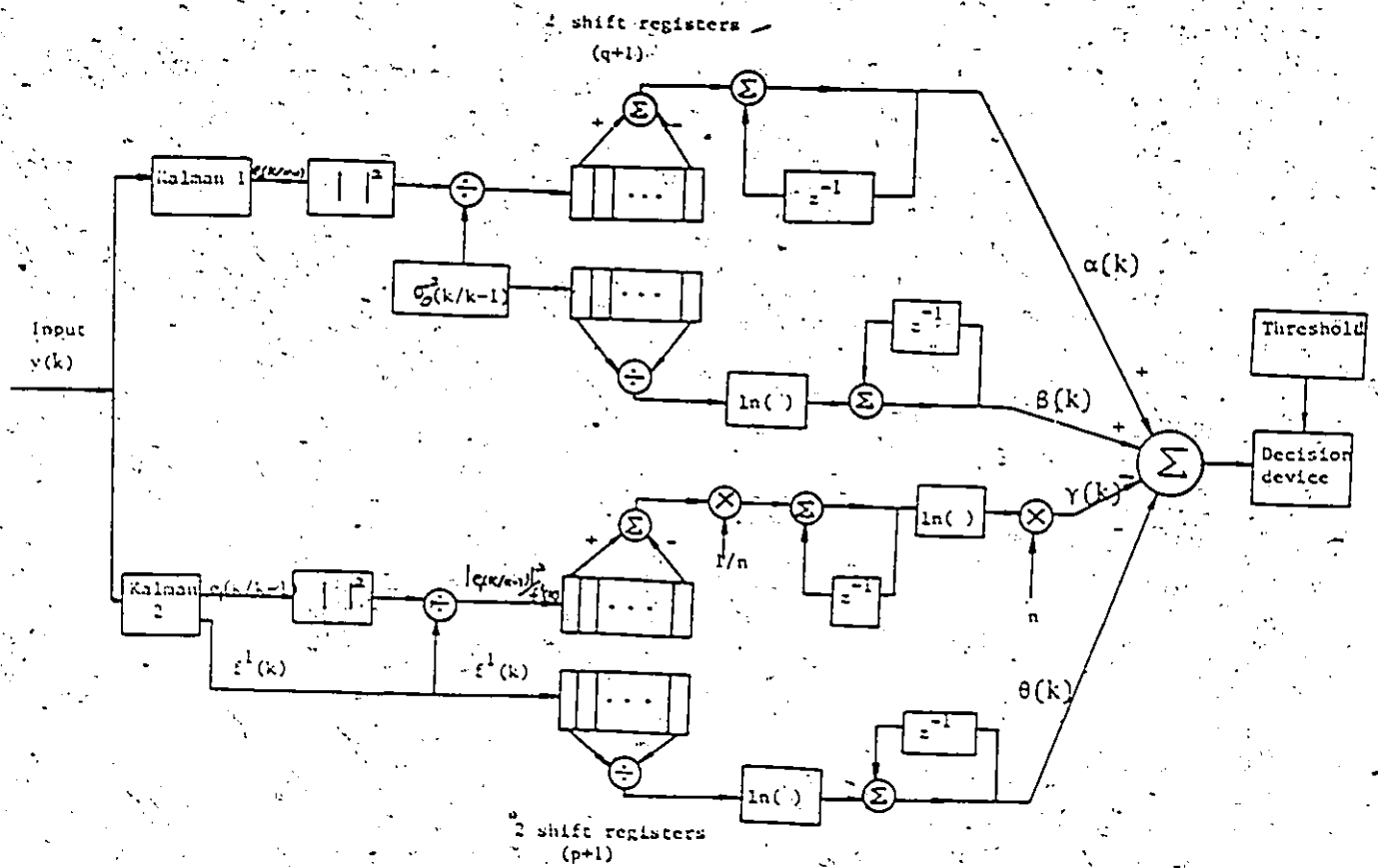


Fig. 5.5: Block diagram of the new detection scheme implementation.

tradeoff, we may train $\underline{Q}^0$ and $\underline{Q}^1$ with some real data under hypotheses H_0 and H_1 . Any optimization technique can be used during training period. Then we fix $\underline{Q}^0$ and $\underline{Q}^1$ before starting the detection algorithm.

In fact, we may further simplify the calculation by assuming that $\underline{Q}^0$ and $\underline{Q}^1$ are of diagonal form and normalized by the standard deviations of the related innovation processes, σ_0 and σ_1 , respectively.

The values of matrices $\underline{Q}^0$ and $\underline{Q}^1$ can be determined by using the maximum-likelihood method as described in chapter 2. For radar detection applications, $\underline{Q}^0$ and $\underline{Q}^1$ vary slowly with time and signal-to-clutter ratio. Moreover, experiments show that the maximum-likelihood estimates are unchanged when $\underline{Q}^0$ and $\underline{Q}^1$ deviate slightly from their optimum values. Thus, we may adopt the following strategy: Train $\underline{Q}^0$ and $\underline{Q}^1$ with a number of sets of radar data taken from different locations in a radar scan and use them in the operating mode. In our experiments, 5 sets of data were used in the training period, each set containing 100 samples. The experimental results show that the following values are suitable for ground and/or weather clutter dominated environment:

$$(\underline{Q}^0)_{ii} = 0.01 \text{ to } 0.03$$

$$(\underline{Q}^1)_{ii} = 0.05 \text{ to } 0.1$$

The fact that the elements in $\underline{Q}^1$ are larger than $\underline{Q}^0$ seems reasonable since a clutter plus target signal is more nonstationary than the clutter process itself.

5.4 COMPUTATIONAL COMPLEXITY

The computational complexity is an important consideration when implementing the detection scheme. For each iteration of the test statistic, it requires $2(p^2 + q^2) + 4(p + q) + 11$ multiplications, $2(p^2 + q^2) + 3(p + q) + 6$ additions and 2 logarithms, where p and q are the orders of H_0 - and H_1 -models, respectively. The computational burden is listed in Table 5.1; as a comparison, the computational burden required by the IBDA or ARAR scheme is also included.

Evidently the computational burden increases quickly with orders p and q ; fortunately, in radar detection application, the orders p and q are small. For instance, in the new scheme, $p = 5$, $q = 2$, and in IBDA, $p = 5$; $q = 1$; the two detection algorithms need 97 and 87 multiplications, respectively.

TABLE 5.1

Computational Complexity

	Multiplications	Additions	Logarithm
New Scheme	$2(P^2 + q^2) + 4(P + q) + 11$	$2(P^2 + q^2) + 3(P + q) + 6$	2
IBDA	$2(P^2 + q^2) + 4(P + q) + 10$	$2(P^2 + q^2) + 3(P + q) + 4$	2

5.5

ESTIMATION OF THE CFAR THRESHOLD

The estimate of the CFAR threshold is obtained from the test statistic generated for some number of, say, 8 resolution cells on either side (in range) of the cell of interest.

In order to obtain a suitable estimate of the correct threshold for the background clutter, the two immediately adjacent resolution cells are excluded from the calculation. This is to eliminate the problem of range-cell splitting when a target may be present in two adjacent range cells.

The adaptive threshold used for the implementation of the new scheme is:

$$(5.8) \quad \lambda(\ell) = k\sigma_\ell + m_\ell$$

where m_ℓ and σ_ℓ are the mean value and the standard deviation of the test statistics, ℓ 's, respectively, k is a constant and will be determined such that the given CFAR is satisfied.

A simplistic method of estimating σ_ℓ and m_ℓ is to take linear average of ℓ 's for these range cells. This approach, however, may reduce detection performance due to the possibility of spikes of ground clutter and/or targets existing in these range cells, with a subsequent increase in the estimate of the threshold. This problem is avoided by searching for the largest values of the ℓ 's from these range cells and excluding them from the calculation of the mean and variance. Denote the estimates thus obtained by $\hat{\sigma}_\ell$ and $\hat{m}_\ell$; then the estimated threshold is given by

$$(5.9) \quad \hat{\lambda}(\ell) = k\hat{\sigma}_\ell + \hat{m}_\ell$$

The CFAR threshold used in the IBDA implementation is [49]

$$(5.10) \quad \hat{\lambda}(\ell) = \hat{m}_\ell + k$$

where m_2 is the median estimate of LLRs from the range cells of interest. Constant k is fixed so as to yield the desired CFAR.

5.6 STATISTICAL EVALUATION USING ACTUAL RADAR DATA

5.6.1 Clutter Data*

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, a fully coherent radar system is needed to allow the recording of both the inphase (I) and the quadrature (Q) data channels. As the detection algorithm requires uniformly sampled data (for ease of implementation) a fixed transmitter pulse-repetition frequency (PRF) is mandatory. Following the recommendations of Gibson [18] and Taylor [73], the sampled video channel must be digitized using a minimum of 10 bits. In addition, since weather-clutter-dominated radar data is needed, a linearly polarized transmitted signal is desirable.

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 summarized in Table 5.2. Although

* This section, together with Figures 5.6-5.9 is taken from Metford and Haykin [49].

TABLE 5.2

TRACS Operating Characteristics

Transmitter frequency = 1.23 GHz
Pulse repetition rate = 657.2 Hz
Pulsewidth = 0.5×10^{-6} s
Antenna revolution rate = 12.5 rev/min (nominal)
Antenna Beamwidth = 1.35° (nominal)
Sampling rate = 1.35 MHz
Sample word size = 10 bits, both inphase and quadrature channels

the TRACS is a fully operational radar system, the air-traffic controllers and the technical support group at the site co-operated fully in configuring the TRACS to our requirements.

Referring to Figure 5.6, 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 perchannel bit-rate of 1.3 MHz. This data stream was recorded on an Ampex HBR-3000 digital recorder, in conjunction with an Ampex AR-1700 wideband airborne/mobile tape transport unit. The digital signal processing

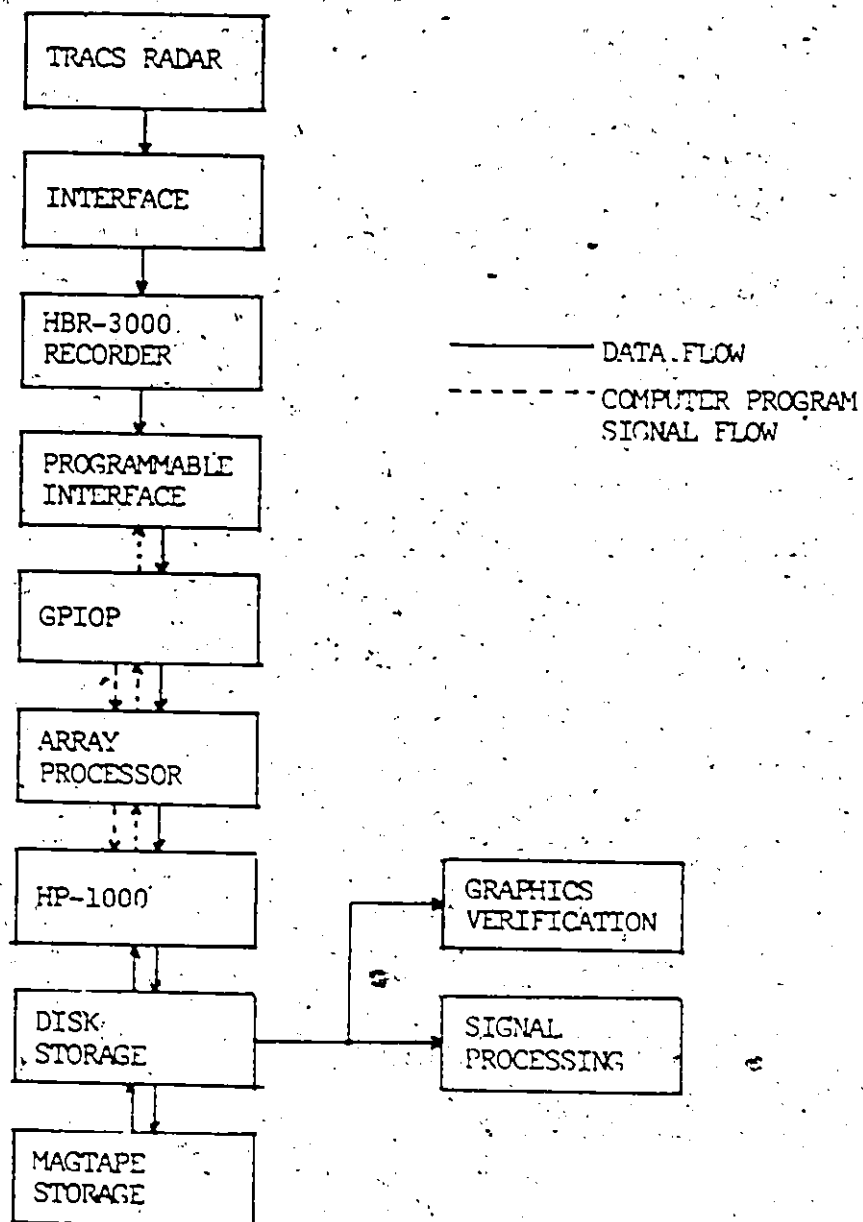


Fig. 5.6 A flowchart of the data acquisition and storage system

was performed on a HP-1000 computer.

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. Photographs of the TRACS PPI taken during these recordings are shown in Figure 5.7.

Several data sets of varying amounts of ground and/or weather clutter are needed for processing by the detection algorithms for radar data. To conform with the specifications for implementation of these algorithms, each of the data sets consist of at least 12 consecutive scans of a two-nautical-mile-width swath of radar returns. This large amount of data (approximately 3.1 M 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.

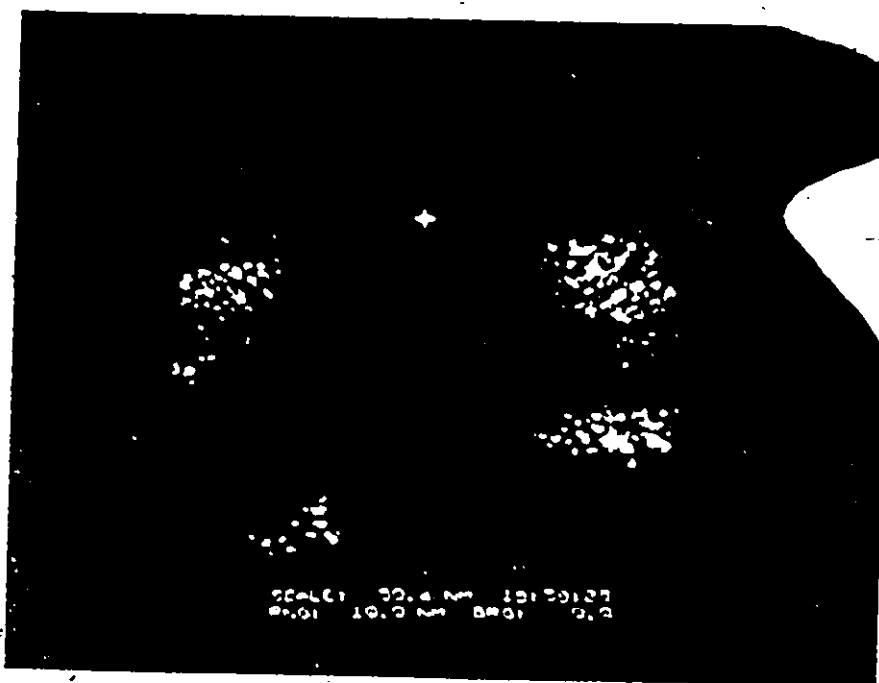
A specific data set is acquired by setting the registers within the programmable interface, so as to select the swath of interest. The data are then passed through the programmable interface into the general programmable input/output processor (GPIOP). By using the array processor (AP) as a rather expensive buffer memory, the GPIOP writes the data into the AP memory as

the HP asynchronously reads data out of AP memory onto disk. This process is continued until at least 12 scans of radar data have been acquired by the HP-1000 computer.

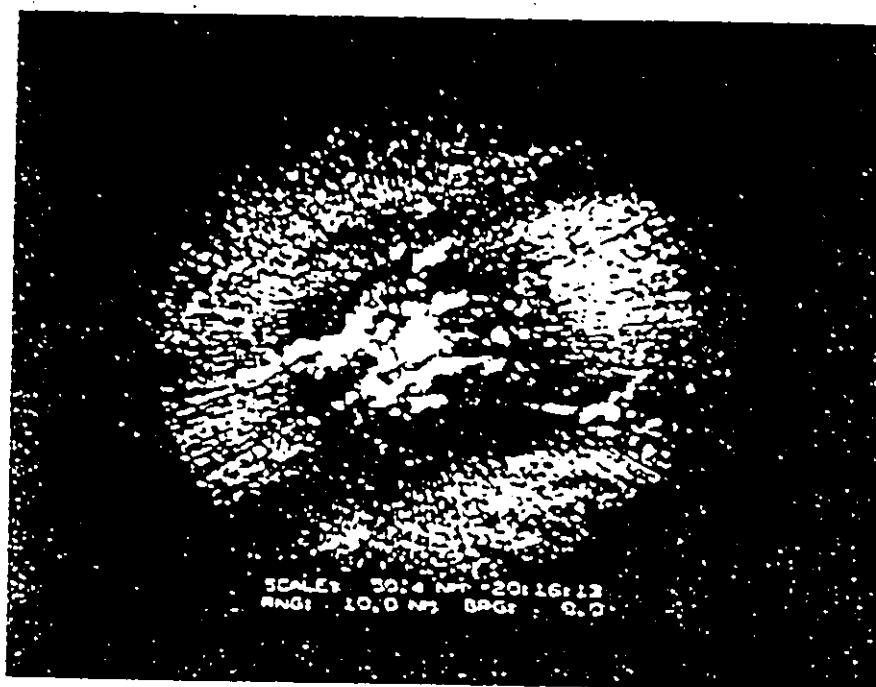
Two data sets were acquired in this manner from the first tape. The first data set is positioned in range from 12.75 to 14.75 nautical miles so as to contain the weather-clutter cells evident in the PPI display (Figure 5.7). A plot of the amplitude envelope $\sqrt{I^2 + Q^2}$ of the first scan of this data set is shown in Figure 5.8. This data set is characterized as being weather-clutter-dominated radar data. The second data set is positioned in range from 1.25 to 3.25 nautical miles and is characterized as being ground-clutter-dominated radar data. A plot of the amplitude envelopes of the first scan of this data set is shown in Figure 5.9

5.6.2 Target Signal

Since it is necessary to examine the performance of the algorithms for all SCRs and target Dopplers, artificial target signal was generated. The target response is of the form



Photograph of the MTI Processed I Channel



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

Fig. 5.7 Photographs of the TRACS PPI display

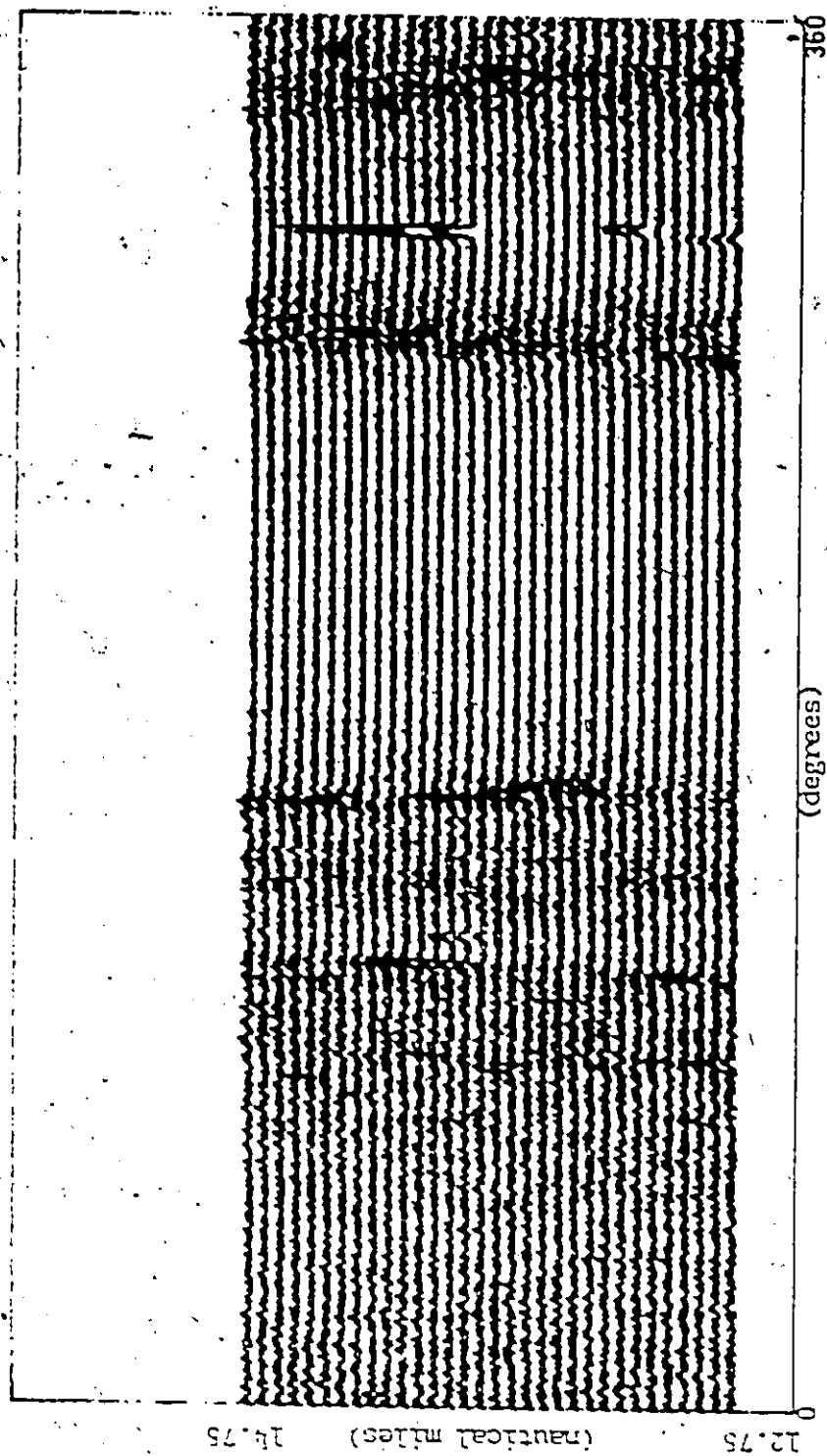


Fig. 5.8 The amplitude envelope of the first scan of the weather clutter dominated radar data

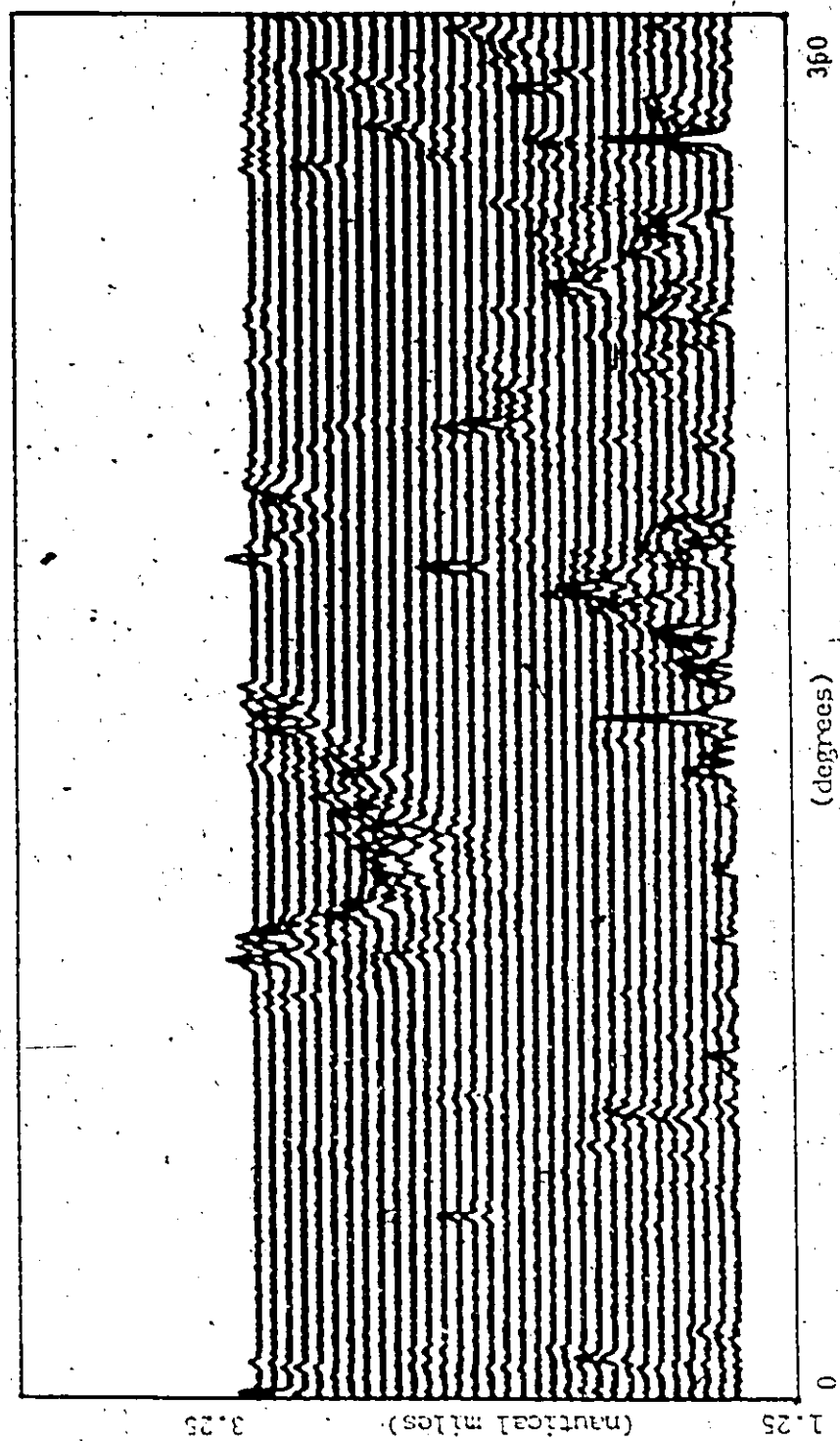


Fig. 5.9 The amplitude envelope of the first scan of the ground clutter dominated radar data

$$(5.11) \quad S(nT) = \exp[-(k-k_0)(\alpha T)^2/4\sigma^2] \exp[2\pi j f(n-n_0)T]$$

where the antenna rotation rate $\alpha = 1.254$ rad/s, the interpulse period $T = 0.0152$ s, the variance of the modelled Gaussian-shaped two-way voltage pattern $\sigma^2 = 0.0000741$ and the Doppler frequency f is in Hertz.

Based on this model and the given parameters, only 18 samples are available for each point target, which is very close to the actual case.

5.6.3 System Performance Evaluation

Each scan contains 32 range-rings and each ring has 3296 samples. 6 scans of data are used to estimate the threshold constant k for $\text{CFAR} = 10^{-5}$.

After fixing the threshold constant, we can evaluate the detection performance of the detection algorithms by using another scan of data. However, the probability density function of the target Doppler required for the evaluation is unknown. But for practical purposes, it is sufficient to assume that the target Doppler is uniformly distributed over the interval from zero to the sample frequency. Considering the possibility of the problem of range-cell splitting where a target is present in two adjacent range cells, this event is assumed to occur with probability equal 20% in our experiments. Since the target response is significantly non-zero only in around 18 samples and

clutter variance is varying spatially, thus the SCR is calculated by estimating target power and clutter power using 20 samples where target signal exists.

Experimental results are given in Figure 5.10. For comparison, the results obtained by using IBDA are also included. The experimental method by using IBDA is described in [49]. Figure 5.10(a) pertains to weather-clutter-dominated samples, and Figure 5.10(b) to ground-clutter-dominated returns. For weather clutter, the performance of the new scheme is much better than the IBDA, especially for $SCR > -3\text{dB}$. For ground clutter, however, IBDA is somewhat better than the new scheme for small SCR, but becomes inferior for $SCR \geq -3\text{dB}$. For both cases, the overall improvement of the new scheme over IBDA is significant.

5.7 TEST WITH SIMULATED DATA

In order to further evaluate the system performance of the detection algorithm, we test it with simulated data. To generate the clutter process, we must find an appropriate model. In the classic works by Barlow [7], Hansen [24], and Goldstein [20], the clutter

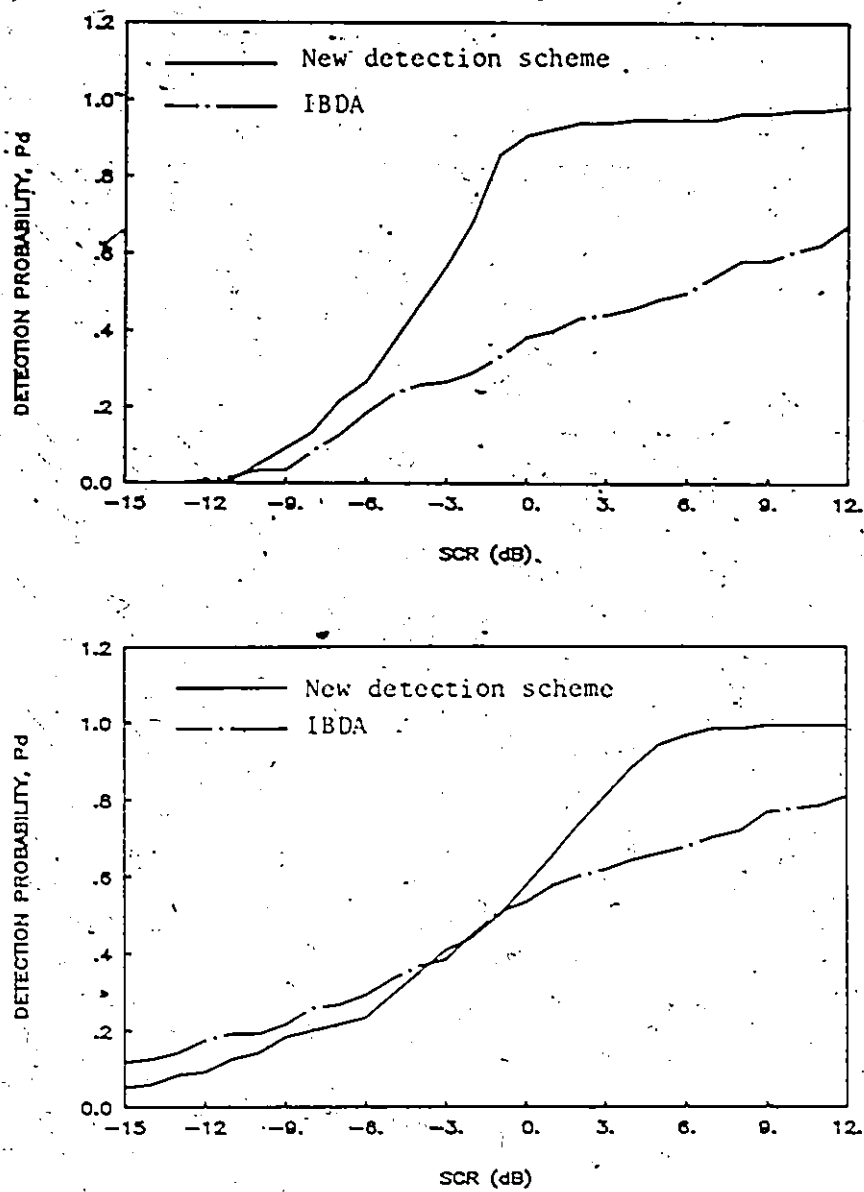


Fig. 5.10: Comparing the new detection scheme with the

IBDA using actual radar data. $P_{fa} = 10^{-5}$.

- (a) Results shown for weather clutter.
- (b) Results shown for ground clutter.

spectrum is assumed to be Gaussian for the majority of fluctuating clutter targets. This spectral shape has since been confirmed by other authors [27], [39]. A somewhat convenient form of the expression for clutter spectrum developed by Barlow [7] is

$$(5.12) \quad S(\omega) = P_c \frac{\sqrt{2\pi}}{\sigma_c} \exp \left[-\frac{(\omega - \omega_c)^2}{2\sigma_c^2} \right]$$

$$(5.13) \quad \sigma_c = 4\pi\sigma_v/(\lambda f_{PRF}) \quad -\pi \leq \omega \leq \pi$$

where ω is the normalized angular frequency, σ_v is deviation of the velocity of the scatterers, f_{PRF} is the pulse repetition frequency and λ the wavelength of transmitted wave of the radar system. P_c is the power of radar clutter process.

Two methods are commonly used to generate radar clutter with the above spectrum. The first one is to assume the clutter process is well fitted by an AR model and the model coefficients can be determined by using the spectrum matched technique [6, 15]. When using the second method, the expected spectrum is obtained by passing a white sequence through a FIR filter, the transfer function of which is given by

$$\begin{aligned}
 (5.14) \quad H(\omega) &= \sqrt{S(\omega)} \\
 &= \sqrt{\frac{P_C}{\sigma_C}} \left(\frac{2}{\pi}\right)^{1/4} \exp \left[-\frac{(\omega - \omega_C)^2}{4\sigma_C^2} \right]
 \end{aligned}$$

or equivalently

$$(5.15) \quad h(\tau) = \left[\sqrt{\frac{P_C}{\sigma_C}} \left(\frac{2}{\pi}\right)^{1/4} \exp(-\sigma_C^2 \tau^2) \right] e^{-j\omega_C \tau}$$

$$(5.16) \quad \sigma_C = 4\pi\sigma_V / (\lambda f_{PRF})$$

In our simulations, the second method was used and the pertinent parameters are

$$P_C = -1$$

$$\sigma_V = 3 \text{ M/sec}$$

$$f_{PRF} = 657.2 \text{ Hz}$$

$$\lambda = 0.244 \text{ M}$$

The way of generating target signal is described in Section 5.6.2.

Three sequences of data were produced. The

second sequence represents the received signal under test, whereas the other two simulate the data samples from the immediately adjacent range rings.

The correlation coefficients between sequences, according to the experimental results using actual radar data, were adjusted to satisfy the following conditions:

$$|\rho_{21}(0)| = 0.7$$

$$|\rho_{23}(0)| = 0.6$$

$$|\rho_{13}(0)| = 0.2$$

where $\rho_{ij}(0)$ denotes the correlation coefficient between the i^{th} and j^{th} sequences with zero time lag.

The results are shown in Figure 5.11. From these curves, we make the following observations:

(1) For $\text{SCR} \leq -5\text{dB}$, IBDA has a better performance than the new scheme, where P_d is less than 0.7. After $\text{SCR} > -5\text{dB}$ the new scheme is better.

(2) Detection probability of the new scheme tends to one quickly as SCR increases. For IBDA, however, P_d rises slowly with SCR. Even when $\text{SCR} = 10\text{dB}$, P_d only attains the value of 0.8.

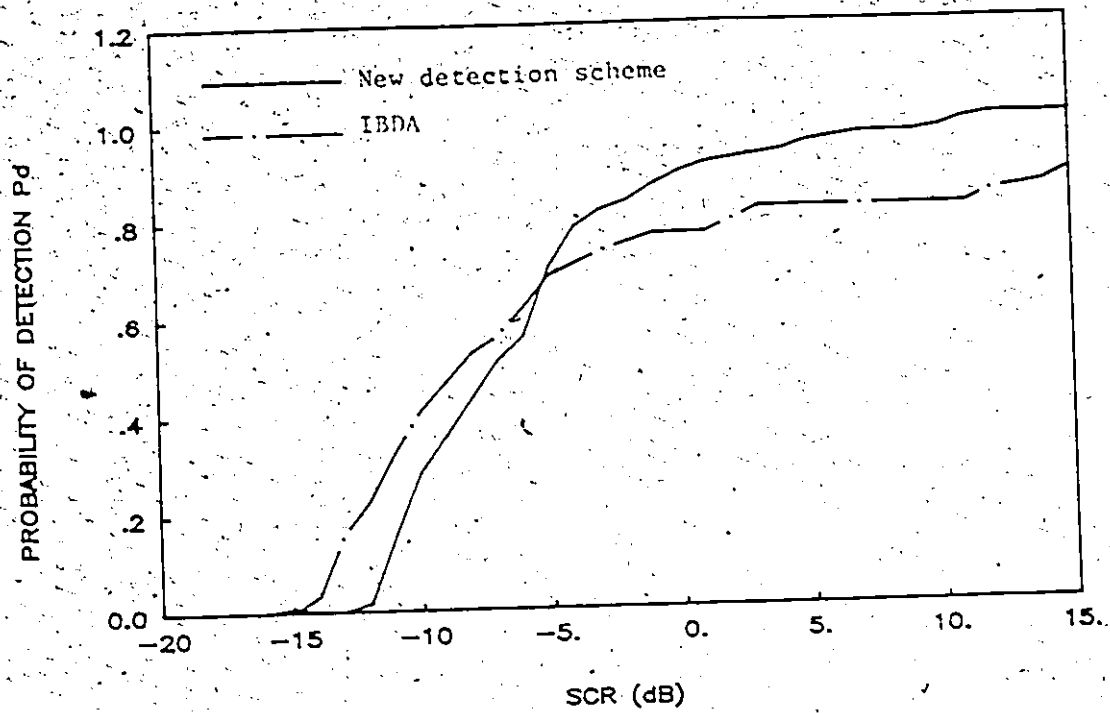


Fig. 5.11 Testing the new detection scheme using simulated data.

Comparing the results for simulated and actual data, it seems that the IBDA works better for simulated data. This is mainly due to the fact that the actual radar returns are highly nonstationary, but in simulations stationary clutter processes were constructed. Nevertheless, for both simulated and real radar data, the detection capability of IBDA increases slowly with SCR. This phenomenon was partly illustrated in Section 3.6. When target Doppler is completely separated from clutter peak, the test statistic and consequently the detection probability increases rapidly with SCR. When target Doppler gets close to, or is obscured by clutter peak, however, the test statistic remains basically unchanged with SCR, and therefore the corresponding P_d is very small.

CHAPTER 6

RADAR CLUTTER SUPPRESSION SCHEMES

In the previous chapters, the signal detection problem is tackled by using two linear models to characterize the two hypotheses, H_0 and H_1 . The relevant random processes are therefore reduced to two sets of model parameters. Correspondingly, the signal detection reduces to the problem of determining which of these two models is most likely to have generated the given observations. In order to enhance the detection performance, the major characteristics inherent in hypotheses H_0 and H_1 must be incorporated into model building. For this purpose, the cross-correlation between the reference data and the noise component of the received signal can be used to set up a linear regression model for hypothesis H_0 .

The same detection problem may be solved by following a different approach. The basic philosophy of the approach is to suppress the noise component as much as possible in order to increase the visibility of the

wanted signal. After noise cancellation, the output signal is applied to an envelope detector. This method distinguishes itself significantly from the above scheme in two aspects:

1) In this method, the noise process is viewed as harmful and must be cancelled out. This is in contrast to the model-based scheme, where the noise is considered as a random process in its own right under hypothesis H_0 ; under H_1 , it constitutes another random process together with the wanted signal.

2) The cross-correlation between the reference data and the noise component of the received signal is employed to construct a noise canceller.

Another point of interest is that in the previous chapters, the cross-correlation between the radar clutter data sampled from adjacent range rings was employed to set up a linear regressive model for hypothesis H_0 in order to solve the moving target detection problem. In this chapter we also make use of this type of characterization to construct a new radar clutter canceller.

The traditional method of clutter suppression employs a multiple-pulse canceller [69], which consists of a finite impulse response filter with binomially weighted tap-weight coefficients. However, this type of filter is only effective in cancelling clutter whose spectrum is centered at zero Doppler frequency.

In the case when covariance matrix of the clutter is known, it is possible to design an optimum linear filter that maximizes the signal/clutter ratio at its output [65]. However, in most practical situations, the statistics of clutter are seldom known in advance, and, typically, they are nonstationary by nature. Accordingly, the most effective solution to the clutter suppression problem relies on the use of adaptive filter theory. An adaptive clutter suppression scheme, based on the lattice prediction error filter (PEF), was proposed by Gibson and Haykin [17, 18]. This scheme makes use of differences between the clutter and target components of radar return; namely, clutter is generally of long duration while the target echo is of short duration. Computer simulation experiments show that an improvement factor enhancement of target echo over clutter, in the range 22-27 dB, is possible, depending on target-velocity [17]. With real radar data, however, an improvement factor of only about 7 dB is achieved [18].

The discrepancy between the results of computer simulation and tests with actual radar data can be attributed to the fact that, owing to nonstationarity, real radar clutter does not always satisfy the assumption that it has a longer stationary duration than the target. This is readily seen in the frequency domain in that the spectral width of clutter is sometimes close to that of target echoes. A clutter canceller based on the PEF tends to suppress not only the clutter component, but also the target component. This occurs when the spectral widths of the two components become close.

Widrow et. al. [75] , describe a general form of adaptive noise canceller that cancels the unwanted interference while keeping the wanted signal basically unchanged.

In this chapter we use the idea of noise cancellers to propose two new schemes for radar clutter cancellation that leave the target echo essentially unchanged. The schemes assume the availability of some reference inputs that are highly correlated with the clutter component in the received signal to be processed.

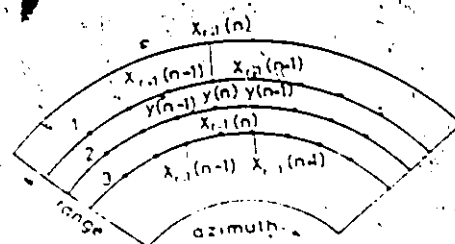
6.1

FORMULATION AND THE NORMAL EQUATION

We first consider the radar clutter suppression scheme I, as shown in Figure 6.1b. Two data sequences from range ring $r-1$ and $r+1$ are used as reference inputs when radar data in ring r are being processed.

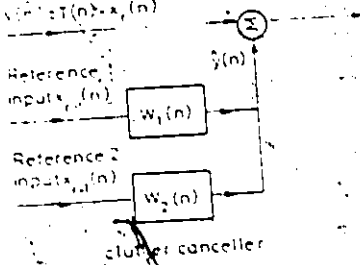
Assume the radar data to be processed, $\{y(n)\}$, are sampled from the r^{th} range ring. The reference inputs are denoted by column vectors $\underline{x}_1(n)$, $\underline{x}_2(n)$, ..., $\underline{x}_N(n)$, taken from both sides of range ring r , where subscript k denotes the ring number where data are obtained, and each vector contains M elements.

From Figure 6.1b, the system output is



1 = ring r-1
2 = ring r
3 = ring r+1

signal to be processed
 $y(n) = T(n) \cdot x_r(n)$



signal to be processed
 $y(n) = T(n) \cdot x_r(n)$

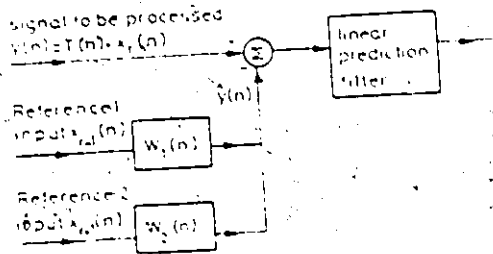


Fig. 6.1 New radar clutter rejection scheme.

- a Data taken from adjacent range rings:
 $y(n)$ is signal to be processed
- b Scheme I
- c Scheme II

$$\begin{aligned}
 (6.1) \quad e(n) &= y(n) - \sum_{\ell=1}^N \underline{x}_{\ell}(n) t_{\underline{w}_{\ell}} \\
 &= y(n) - \underline{x}(n)^t \underline{w}
 \end{aligned}$$

where

$$\underline{x}(n) = \begin{bmatrix} \underline{x}_1(n) \\ \underline{x}_2(n) \\ \vdots \\ \underline{x}_N(n) \end{bmatrix}$$

$$\underline{w} = \begin{bmatrix} \underline{w}_1 \\ \underline{w}_2 \\ \vdots \\ \underline{w}_N \end{bmatrix}$$

The optimum tap-weight solution for $\underline{w}$ is obtained by applying Wiener filtering theory to equation (6.1) [29, 35]:

$$(6.2) \quad \underline{w}_0 = \underline{R}_{xx}^{-1} \underline{d}_{yx}$$

where

$$\underline{R}_{xx} = E[\underline{x}(n) \underline{x}(n)^*]$$

$$\underline{d}_{yx} = E[\underline{y}(n) \underline{x}(n)^*]$$

6.2 CLUTTER-SUPPRESSION SYSTEM I

After finding the optimum tap-weight vector $\underline{w}_0$ the response of the canceller to the input signal $y(n)$ is

$$(6.3) \quad e(n) = y(n) - \underline{w}_0^t \underline{x}(n)$$

Now we consider two cases separately, namely,

- (1) Only clutter data are applied to the system.
- (2) Input signal consists of target echo plus clutter.

Case I.

Since the input consists of clutter only, $y(n) = c(n)$, the system output equals

$$(6.4) \quad e(n) = c(n) - (\underline{R}_{xx}^{-1} \underline{d}_{cx})^t \underline{x}(n)$$

Squaring the absolute value of this equation and taking expectation, the output power is obtained as

$$\begin{aligned}
 (6.5) \quad \sigma_e^2 &= E|e(n)|^2 \\
 &= E[|c(n) - (\underline{R}_{xx}^{-1} \underline{d}_{cx})^T \underline{x}(n)|^2] \\
 &= \sigma_c^2 - \underline{d}_{cx}^H \underline{R}_{xx}^{-1} \underline{d}_{cx}
 \end{aligned}$$

where σ_c^2 denotes the power of clutter process $\{c(n)\}$.

Case II.

A target signal is present, $y(n) = T(n) + c(n)$, and the system output equals

$$(6.6) \quad e(n) = y(n) - (\underline{R}_{xx}^{-1} \underline{d}_{yx})^T \underline{x}(n)$$

Since the target signal is assumed independent of the clutter process, we have

$$\begin{aligned}
 (6.7) \quad \underline{d}_{yx} &= E[y(n) \underline{x}^*] = E[(T(n) + c(n)) \underline{x}(n)^*] \\
 &= E[c(n) \underline{x}(n)^*] = \underline{d}_{cx}
 \end{aligned}$$

Thus, equation (6.6) becomes

$$(6.8) \quad e(n) = T(n) + [c(n) - (R_{xx}^{-1} d_{cx})^t x(n)]$$

We recognize that the first term on the right is the output target signal, which is exactly the same as the input; in other words, the canceller does not suppress the target component, and the second term is the same as given in equation (6.5), representing the clutter residual in the system output.

From equation (6.8), we readily find the output power is given by:

$$(6.9) \quad \sigma_e^2 = \sigma_T^2 + (\sigma_c^2 - d_{cx}^H R_{xx}^{-1} d_{cx})$$

where σ_T^2 denotes the target signal power. The second term on the right is the power of the residual clutter; consequently, SCR at the system output equals:

$$(6.10) \quad SCR_0 = \frac{\sigma_T^2}{\sigma_c^2 - d_{cx}^H R_{xx}^{-1} d_{cx}}$$

But the SCR at the canceller input equals

$$(6.11) \quad SCR_1 = \sigma_T^2 / \sigma_c^2$$

Thus, the SCR gain obtained by using clutter suppression I

is given by

$$(6.12) \quad I = \frac{SCR_0}{SCR_1} = \frac{\sigma_c^2}{\sigma_c^2 - \frac{d^H R^{-1} d}{cx^H xx cx}}$$

This ratio may be viewed as the system improvement factor.

6.3 ANALYSIS OF SYSTEM IMPROVEMENT FACTOR

Inspecting equation (6.10), we may draw the following conclusions:

(1) The system suppresses the clutter component while keeping the target power unchanged.

(2) Improvement factor of the system is always larger than one, i.e.,

$$I = \frac{SCR_0}{SCR_1} \geq 1$$

(3) If the improvement factor, given N reference inputs, $x_1, x_2, \dots, x_N$, is denoted by I_N , then we have

$$I_{N+1} > I_N$$

Proof. Assume $\hat{c}_N(n)$ is the estimate of $c(n)$, given the N reference inputs $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N$, and $\hat{c}_{N+1}(n)$ is the estimate of $c(n)$ given $N+1$ reference inputs $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N, \underline{x}_{N+1}$. Let $e_N^c(n)$ and $e_{N+1}^c(n)$ denote the corresponding errors, respectively. Thus,

$$(6.13) \quad e_N^c(n) = c(n) - \hat{c}_N(n)$$

$$e_{N+1}^c(n) = c(n) - \hat{c}_{N+1}(n)$$

Now we decompose $\underline{x}_{N+1}$ into two orthogonal terms by projecting vector $\underline{x}_{N+1}$ onto space ϕ_N , spanned by vectors $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N$:

$$(6.14) \quad \underline{x}_{N+1} = \underline{\hat{x}}_{N+1} + \underline{v}_{N+1}$$

where $\underline{\hat{x}}_{N+1}$ is the projection of $\underline{x}_{N+1}$ and $\underline{v}_{N+1}$, the error vector. Apparently, the estimate $\hat{c}_{N+1}(n)$ of $c(n)$ is the sum of the projections of $c(n)$ onto space ϕ_N and onto $\underline{v}_{N+1}$. The projection of $\hat{c}(n)$ onto $\underline{v}_{N+1}$ is denoted by $\Delta \hat{c}_{N+1}(n)$, and that onto ϕ_N is recognized to be $\hat{c}_N(n)$. Thus,

$$(6.15) \quad \hat{c}_{N+1}(n) = \hat{c}_N(n) + \Delta \hat{c}_{N+1}(n)$$

and

$$\begin{aligned}
 (6.16) \quad e_{N+1}^c(n) &= c(n) - \hat{c}_{N+1}(n) \\
 &= [c(n) - \hat{c}_N(n)] - \Delta c_{N+1}(n) \\
 &= e_N^c(n) - \Delta c_{N+1}(n)
 \end{aligned}$$

According to the principle of orthogonality, $e_{N+1}^c(n)$ is orthogonal to $\Delta c_{N+1}(n)$, and $\Delta c_{N+1}(n)$ can be viewed as the projection of $e_N^c(n)$ onto $\underline{v}_{N+1}$. Thus

$$(6.17) \quad E|e_{N+1}^c(n)|^2 = E|e_N^c(n)|^2 - \underline{R}_{ev}^H \underline{R}_{vv}^{-1} \underline{R}_{ev}$$

where

$$\begin{aligned}
 (6.18) \quad \underline{R}_{vv} &= E[\underline{v}_{N+1}^* \underline{v}_{N+1}^t] \\
 \underline{R}_{ev} &= E[e_N^c(n) \underline{v}_{N+1}^*]
 \end{aligned}$$

Since $\underline{R}_{vv}$ is positive definite, the second term in the right is negative, we have:

$$(6.19) \quad E|e_{N+1}^c(n)|^2 < E|e_N^c(n)|^2$$

Consequently,

$$\frac{\sigma_c^2}{E|e_{N+1}^c(n)|^2} > \frac{\sigma_c^2}{E|e_N^c(n)|^2}$$

That is,

$$(6.20) \quad I_{N+1} > I_N$$

From the above discussion, two possible ways to enhance the system performance are:

- (1) Use sequences that have high correlation with the clutter component of the received signal under considerations as reference inputs to the canceller.
- (2) Increase the number of reference channels.

However, the first method is often limited by a given environment and the latter will increase the computational complexity.

6.4 IMPLEMENTATION OF CLUTTER SUPPRESSION SCHEME I

In radar suppression scheme I, we use two reference channels to construct a practical clutter canceller and the filter length for each channel is chosen to be two. Those choices are justified for two reasons. First, the cross-correlation coefficient between two rings separated

in range is generally small. Second, the correlation function decreases rapidly with the lag.

A Kalman filter assuming random walk model, as described in chapter 2, can be employed to calculate the canceller output. The technique addressed there can also be used to determine the unknown parameters of a clutter canceller. Clutter suppression scheme I is to be tested by using actual radar data, including weather clutter 'WTRNWL' and ground clutter 'WTRNGl'. For the details of these radar data, see Sections 5.1 and 5.6.1.

6.5

SYSTEM OUTPUT IN THE FREQUENCY DOMAIN

The system capability of suppressing clutter is readily observed in the frequency domain. The power spectrum of the canceller output is analyzed by using the maximum-entropy method (MEM). For comparison, the spectrum of the input signal is also provided in each diagram. The number of data samples used for analysis is 20, which is approximately equal to the duration of a target echo obtained during each antenna scan. Figure 6.2 shows the experimental results for weather clutter and Figure 6.3 for ground clutter. In these diagrams, frequency values are normalized with respect to the pulse repetition

frequency which is 657.2Hz and spectral density values are normalized with respect to the maximum value, so that they extend from 0dB to the limiting value which is selected in advance.

Figure 6.2(a) shows the case when only radar clutter is applied to the canceller. We see from this diagram that the canceller substantially suppresses the clutter, with the clutter peak becoming 15 dB attenuated compared to the original input. Moreover, the band width of the output signal is wider than the input. In other words, the canceller has a tendency to whiten the clutter signal. The experimental results when a target is present are shown in Figures 6.2(b)-(f). An important result is that the canceller suppresses the clutter component while it lets the target component to pass through it with little attenuation. As a result, the target signal is greatly enhanced at the canceller output. This confirms the property theoretically developed in Sections 6.2 and 6.3. The canceller works very well, even in a weak target signal level. In Figure 6.2(b), where SCR = -13dB, the MEM spectrum of the canceller output continues to indicate the target peak around its correct position as labelled by the dashed line. As the SCR increases to -12dB, and -10 dB the target peak becomes more and more prominent. (See Figures 6.2(c), (d)). In both cases, however, the MEM

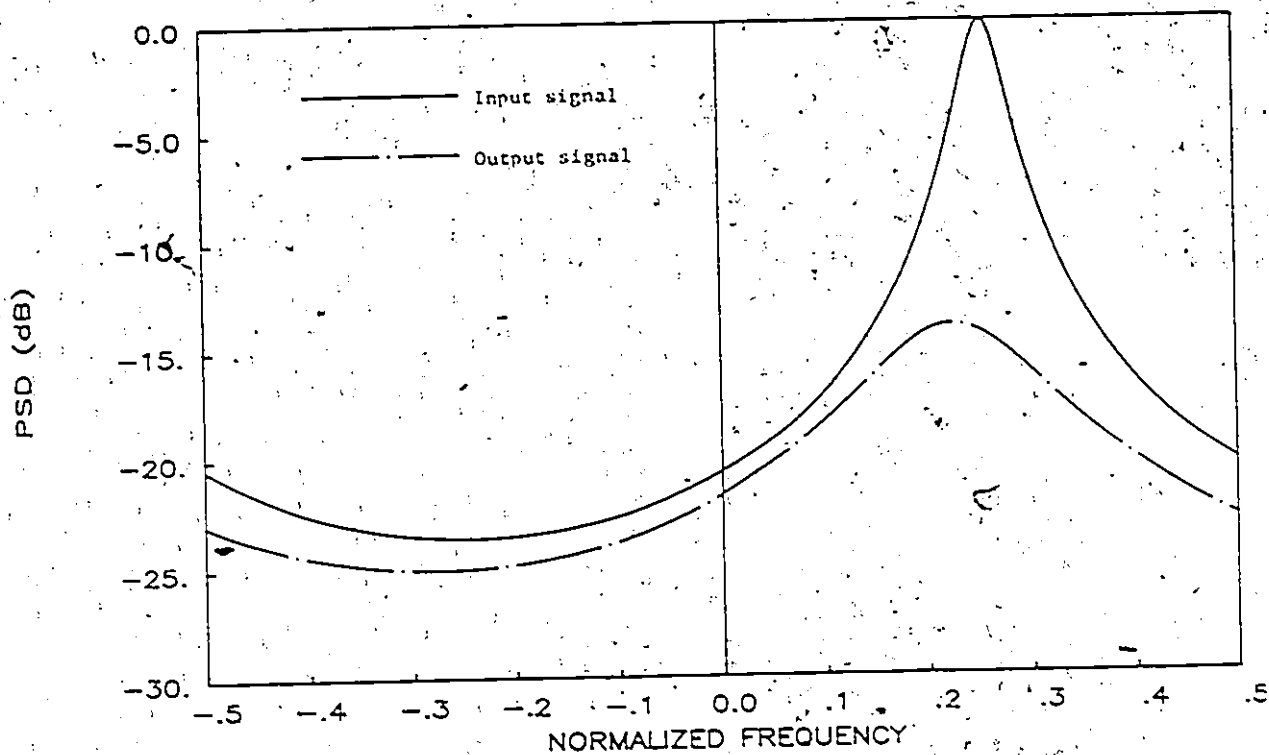


Fig. 6.2: Power spectra of the input and output signal of clutter suppression Scheme I. Weather clutter 'WTRNW1', range ring = 25, azimuth = 2520---2540.

(a) clutter only.

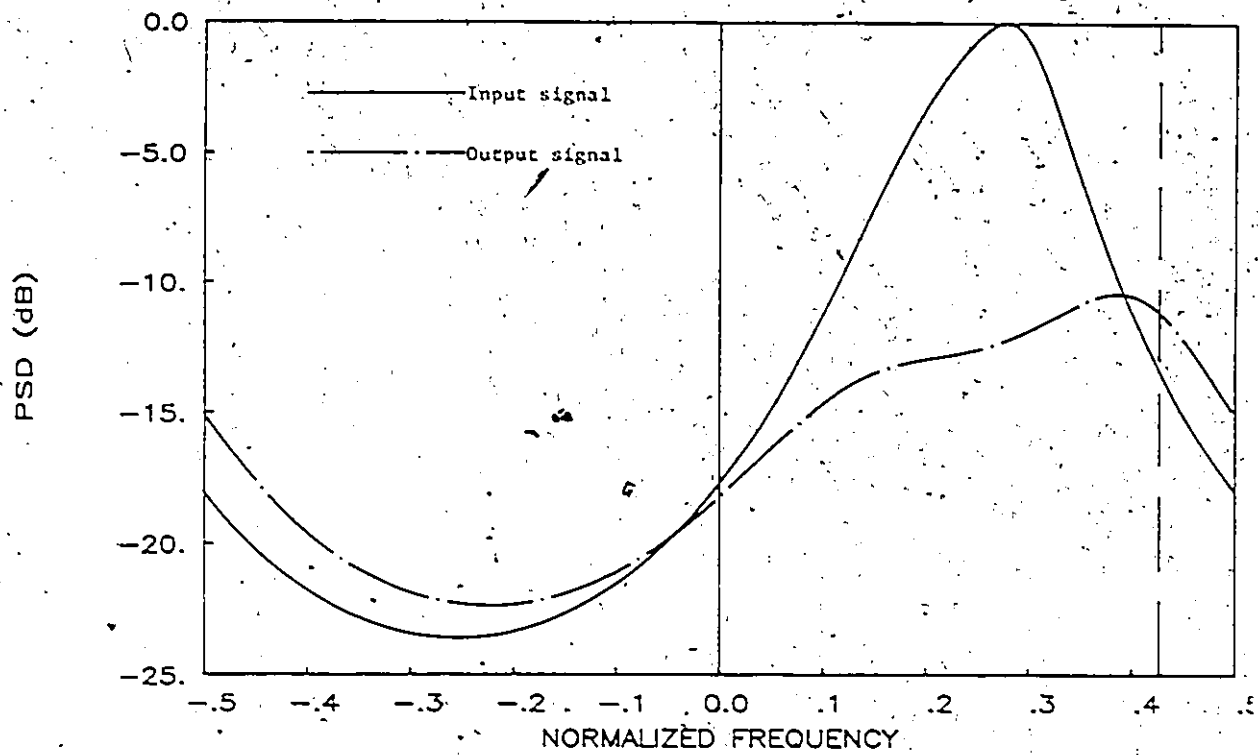


Fig. 6.2(b): SCR = -15 dB

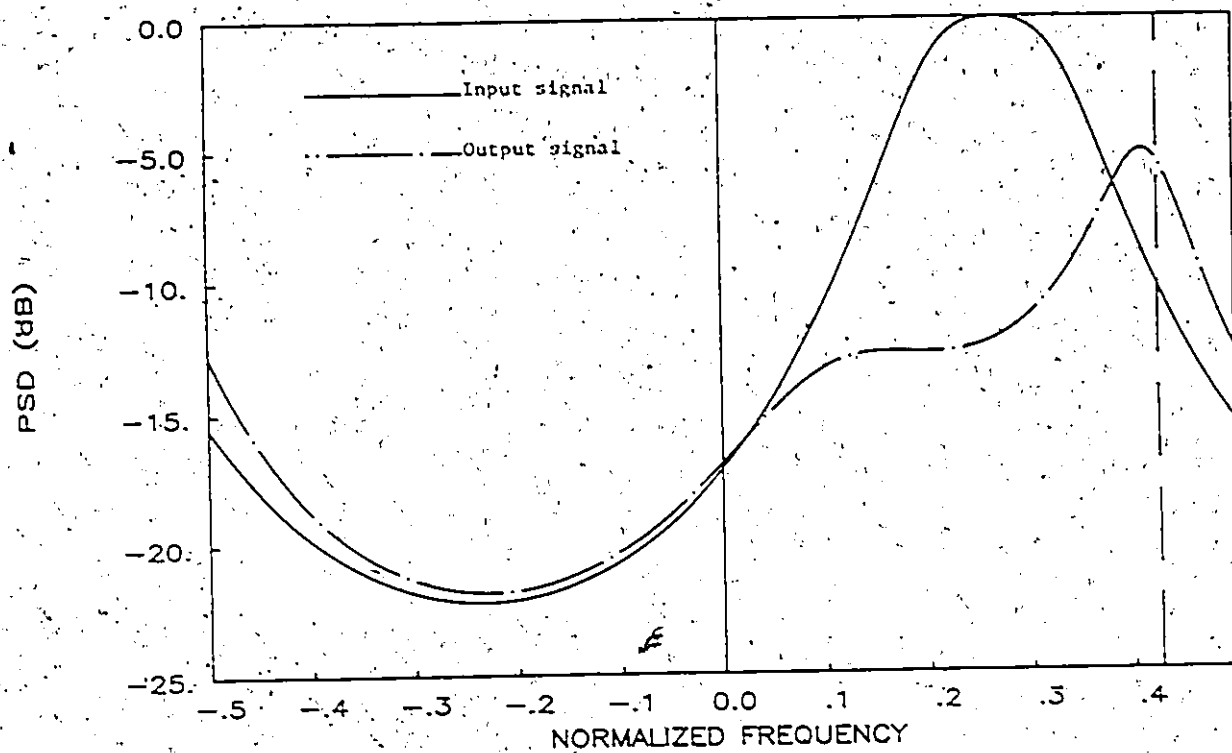


Fig. 6.2(c): SCR = -12 dB

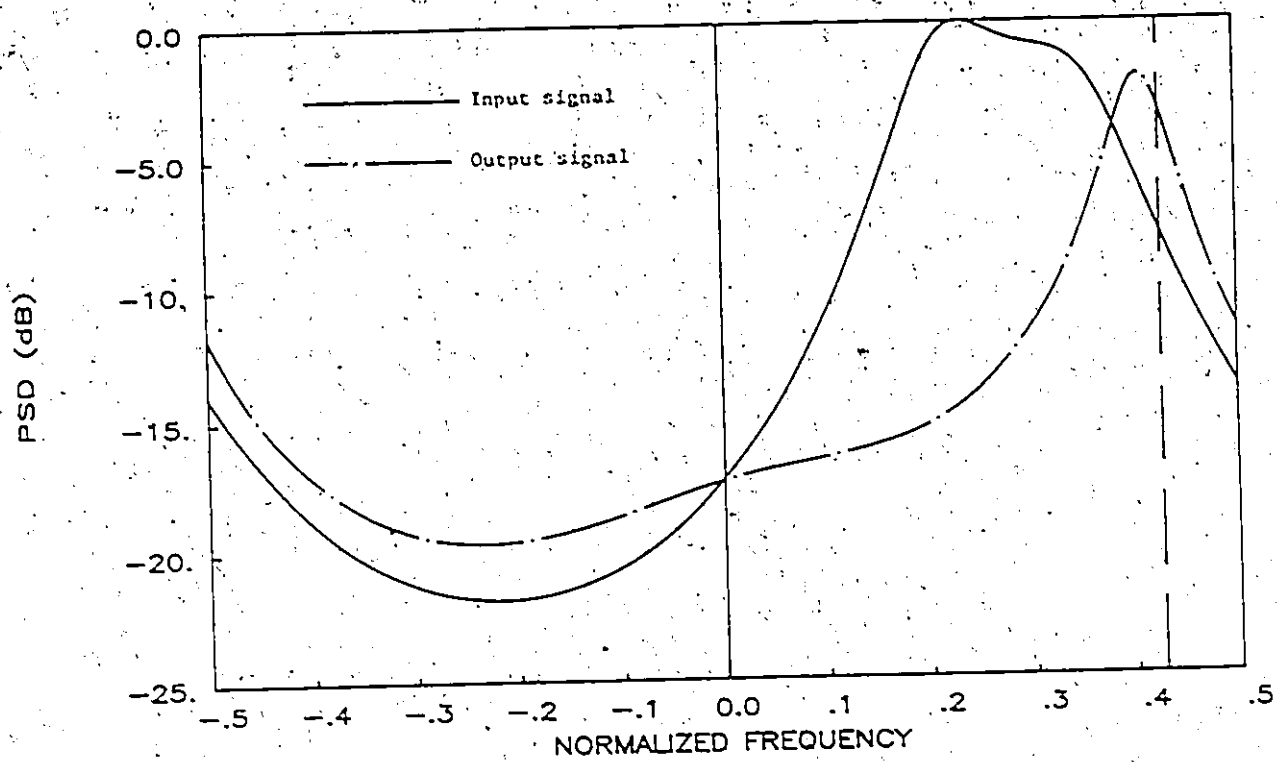


Fig. 6.2(d): SCR = -10 dB

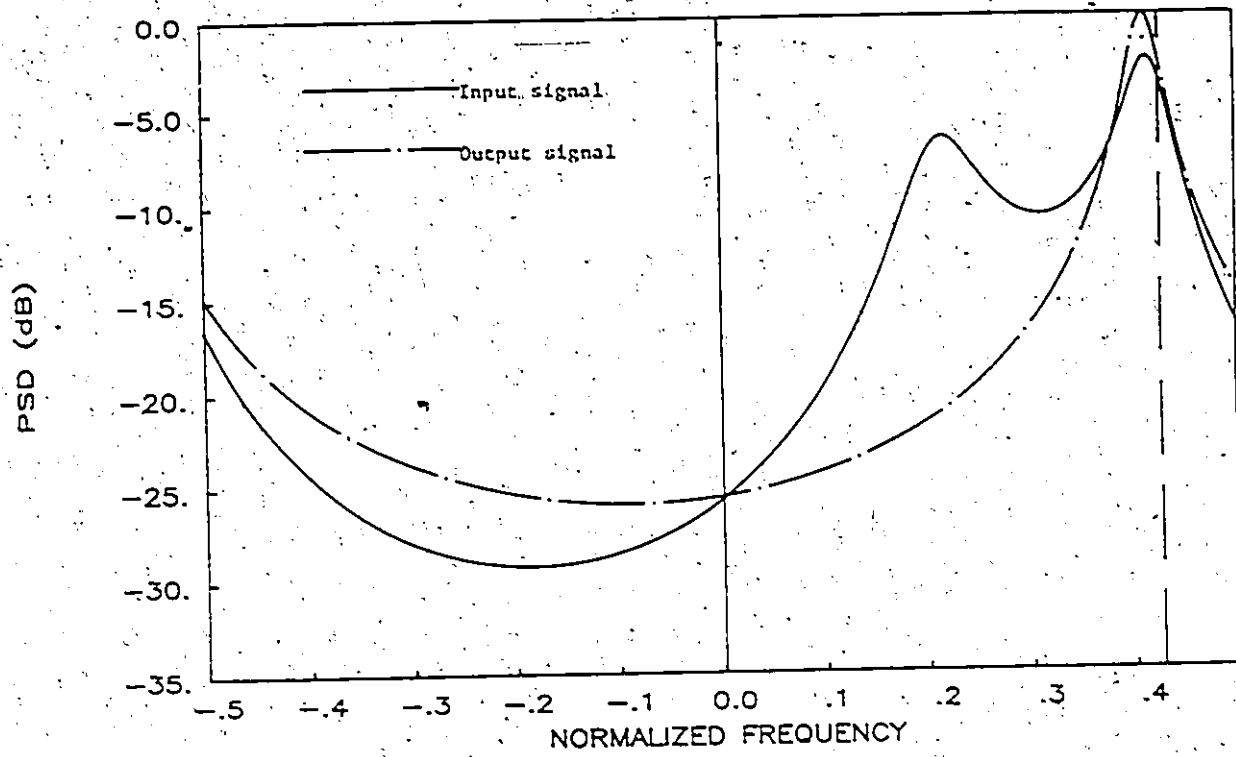


Fig. 6.2(e): SCR = -5dB

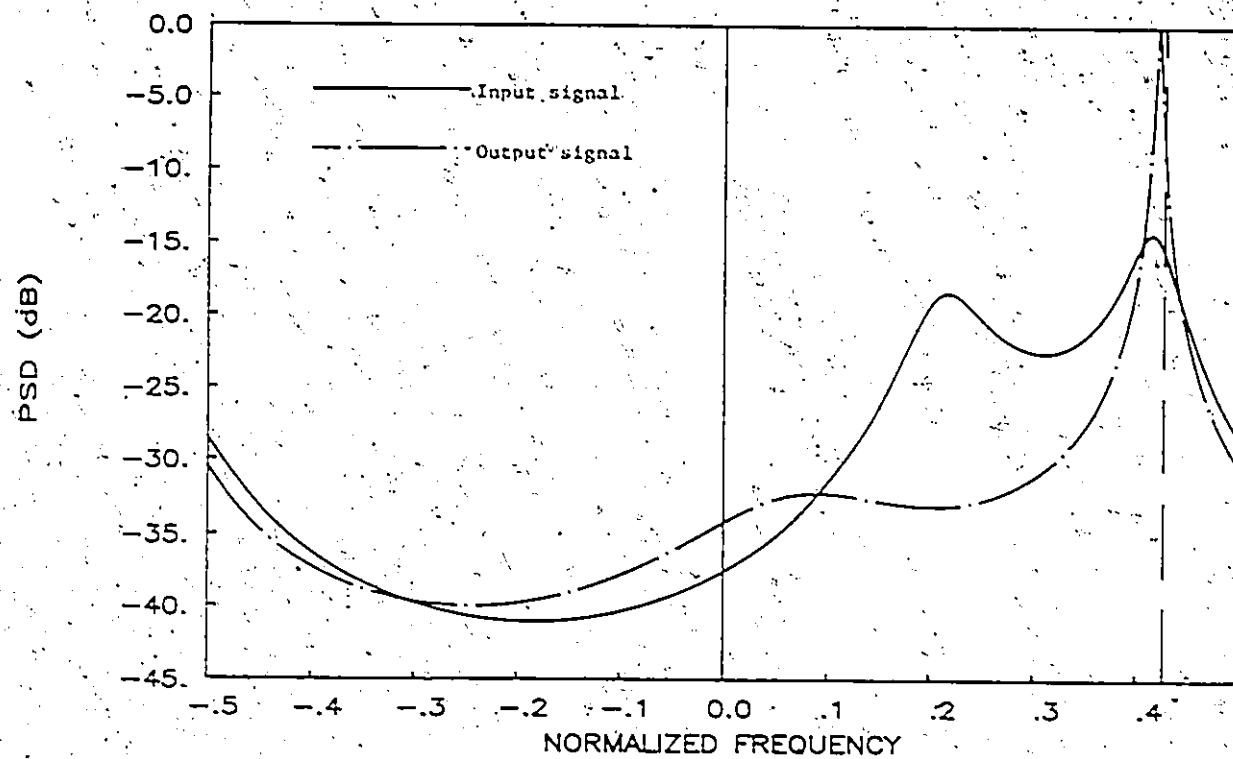


Fig. 6.2(f): SCR = -5 dB

spectra of the original input signals fail to show the presence of a target; see Figure 6.2(c). When the SCR rises further, say, -5dB in Figure 6.2(e) and -3dB in Figure 6.2(f), the power spectrum of original input starts to indicate the target peak.

Similar results are also observed for ground clutter as shown in Figure 6.3. Figure 6.3(a) shows that the canceller has an even better whitening capability for ground clutter than for weather clutter. This is mainly due to the differences in crosscorrelation characteristics for the two types of clutter. In Figures 6.3(b), (c), we put the canceller in a difficult position where the target Doppler is very close to the clutter peak. Even for a traditional fixed FIR filter, it is difficult to design a narrow transition band to meet this requirement. The adaptive suppression scheme I, however, adjusts its frequency characteristics correctly so that only the target component can pass through the filter. Excellent results are given in Figures 6.3(b), (c), where we can see the estimated target Doppler is almost exactly the same as its true value.

It is interesting to compare the frequency characteristics of clutter suppression scheme I with the adaptive prediction-error-filter (PEF) which was previously

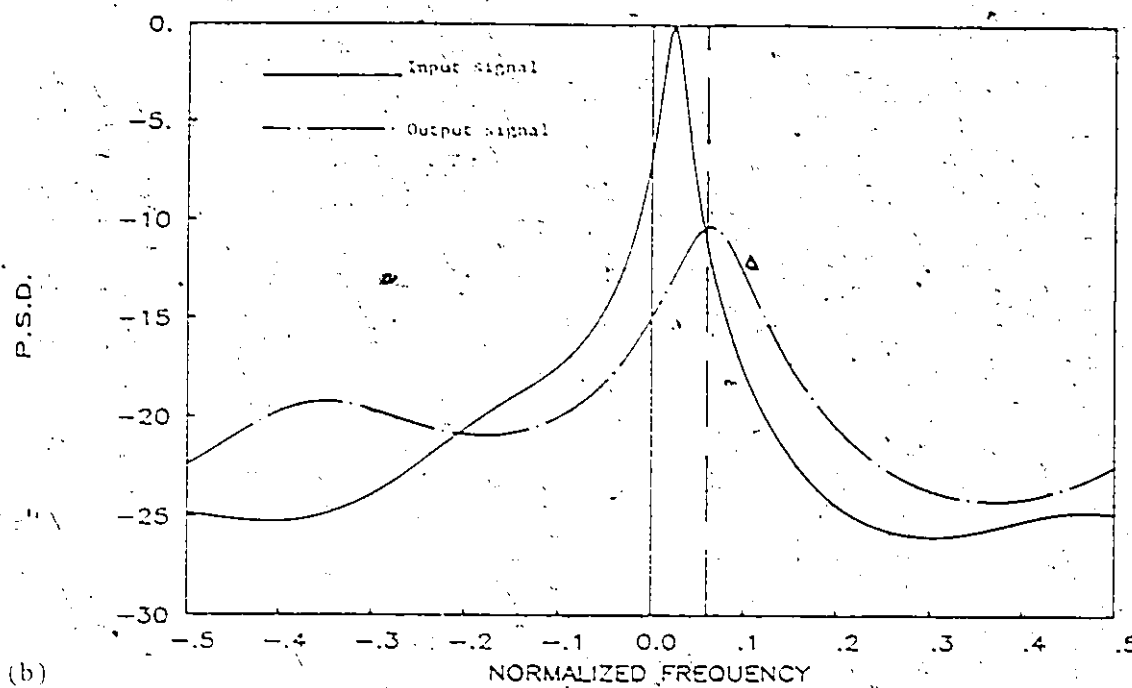
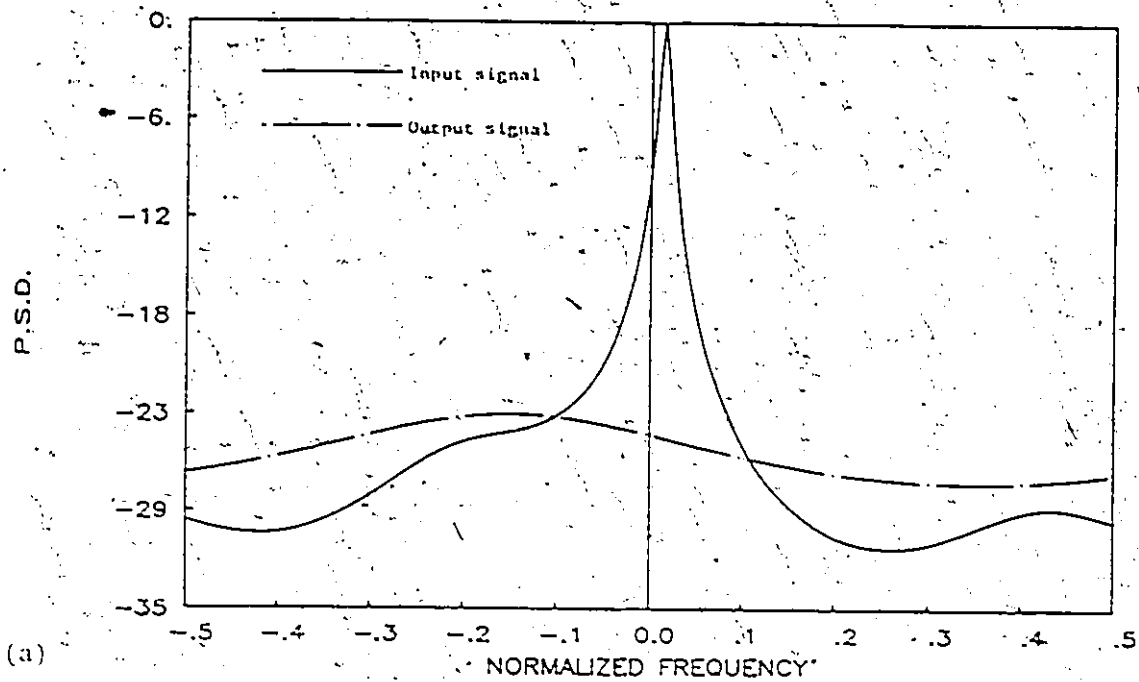


Figure 6.3: Power spectra of the input and output signal of the clutter suppression scheme 1. Ground clutter 'WTRNG1', range ring = 28, azimuth = 1500---1520.

(a): SCR = no target present

(b): SCR = -10 dB

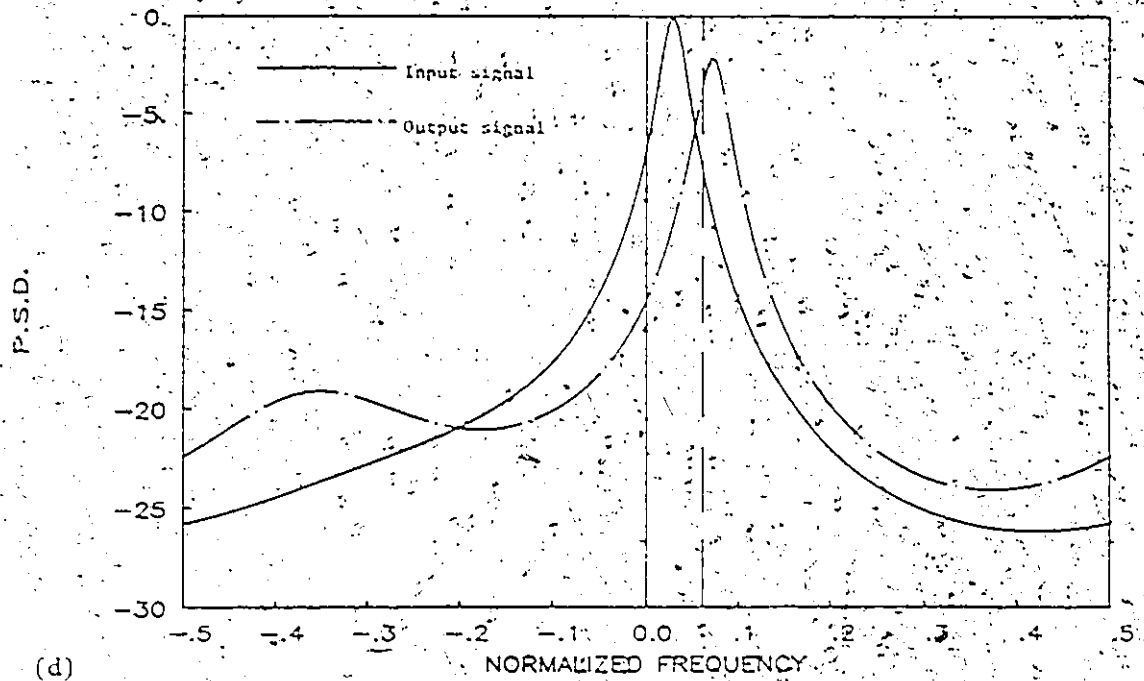
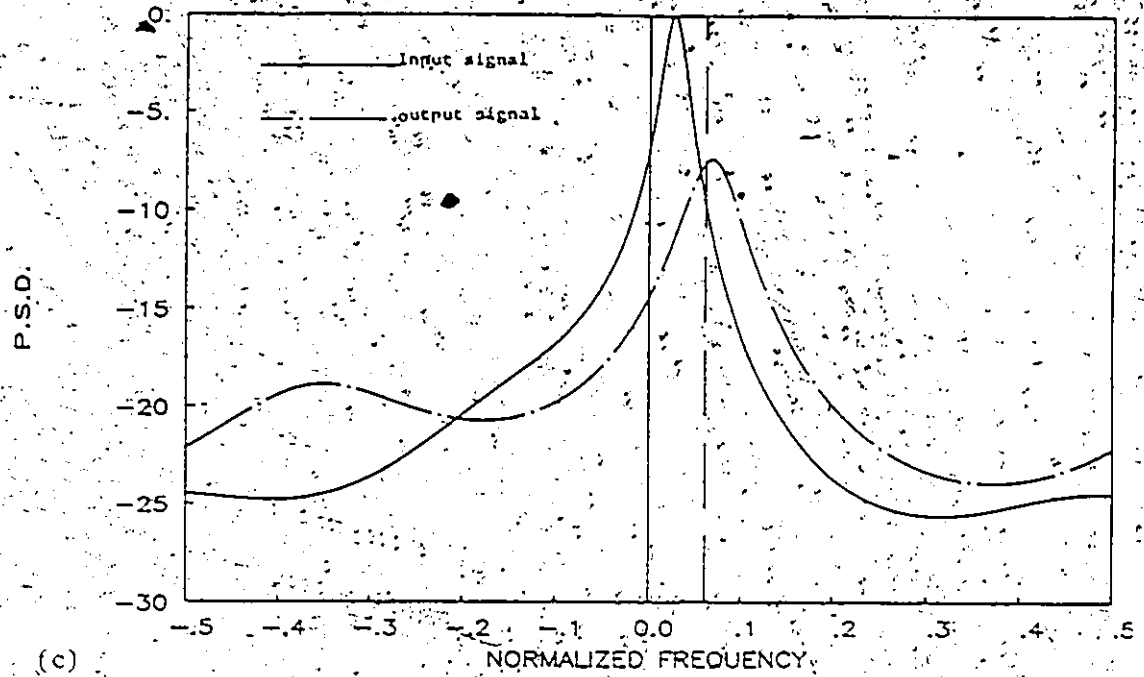


Figure 6.3: Power-spectra of the input and output signal of the clutter suppression scheme 1. Ground clutter 'WTRNG1', range ring = 28, azimuth = 1500--1520.

(c): SCR = -8 dB

(d): SCR = -5 dB

described by Gibson and Haykin [17, 18]. The results are given in Figure 6.4 and Figure 6.5 for ground clutter and weather clutter, respectively. PEF uses a parameter ω to control the time constant of system transient response. According to [17, 18], this constant should be designed such that the filter has enough time to follow the statistical characteristics of clutter component but can not adapt to sudden changes in statistics of the process due to the presence of a target. This value is experimentally determined to be 0.6 in our experiments. Figure 6.4(a) shows the case where target Doppler is completely separated from the clutter peak. The PEF cancels the target component as well as the clutter echo, but the attenuation of clutter is more than that of the target component. As a result, the target output is relatively enhanced. The canceller, on the other hand, removes the clutter peak while it keeps the target peak basically unchanged. As shown in Figure 6.4(b), when the target peak coincides with a clutter peak, the performance of the PEF becomes worse in that its output is dominated by the clutter component. This is in contrast to the canceller which still works very well.

When operating in a weather clutter-dominated

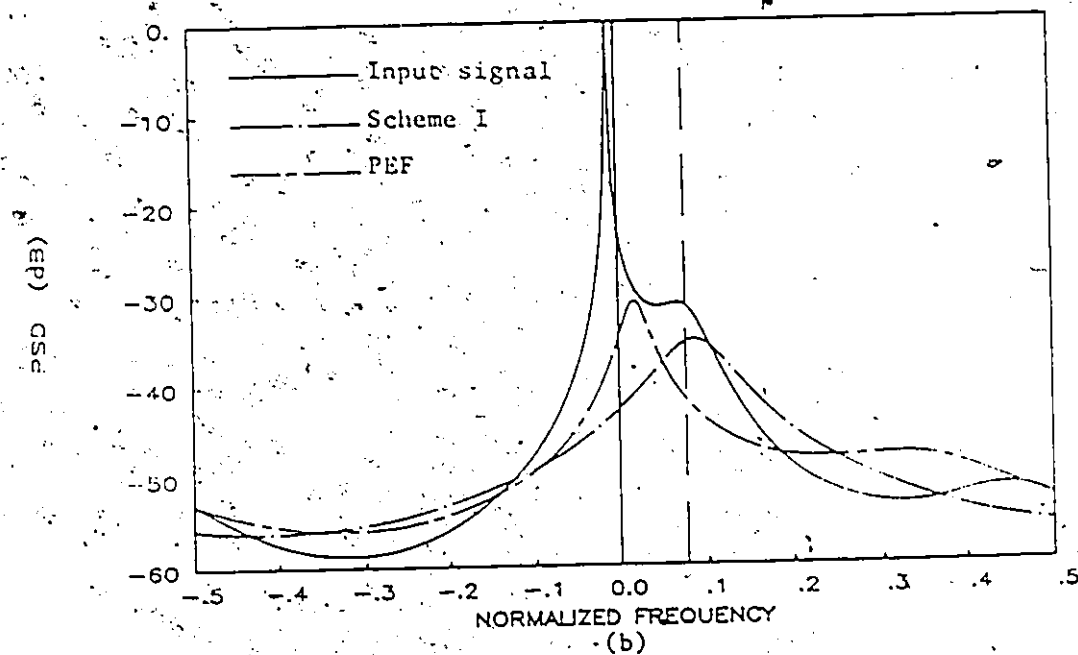
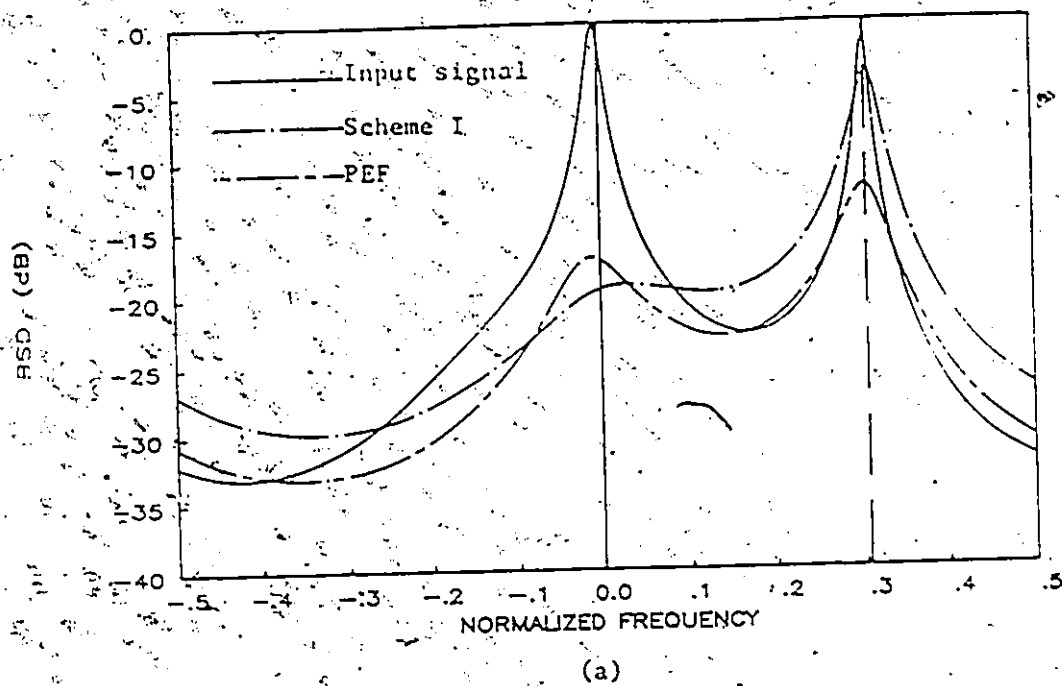
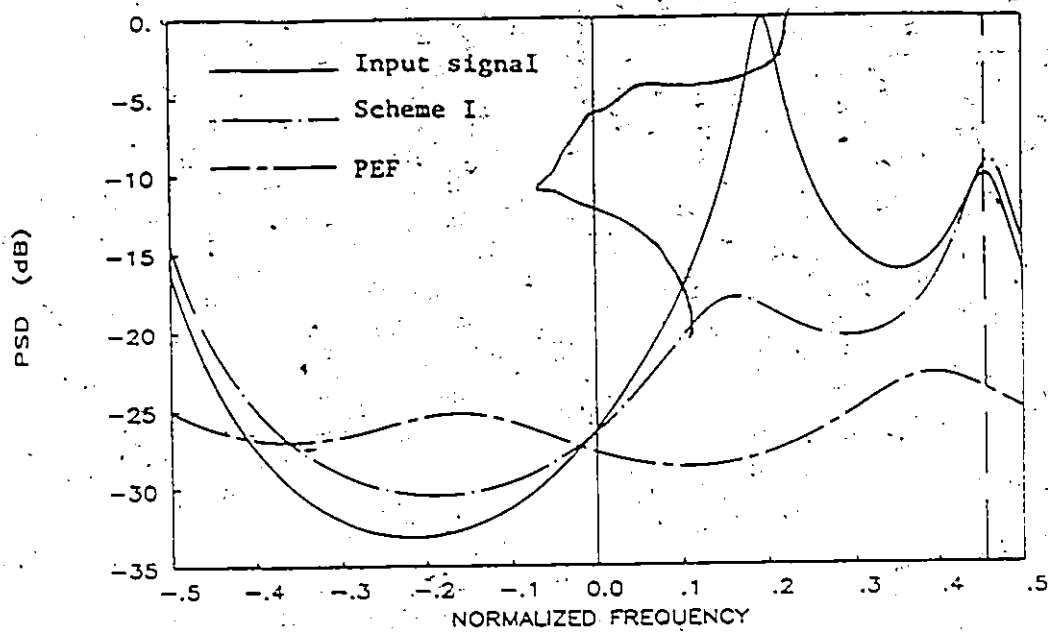


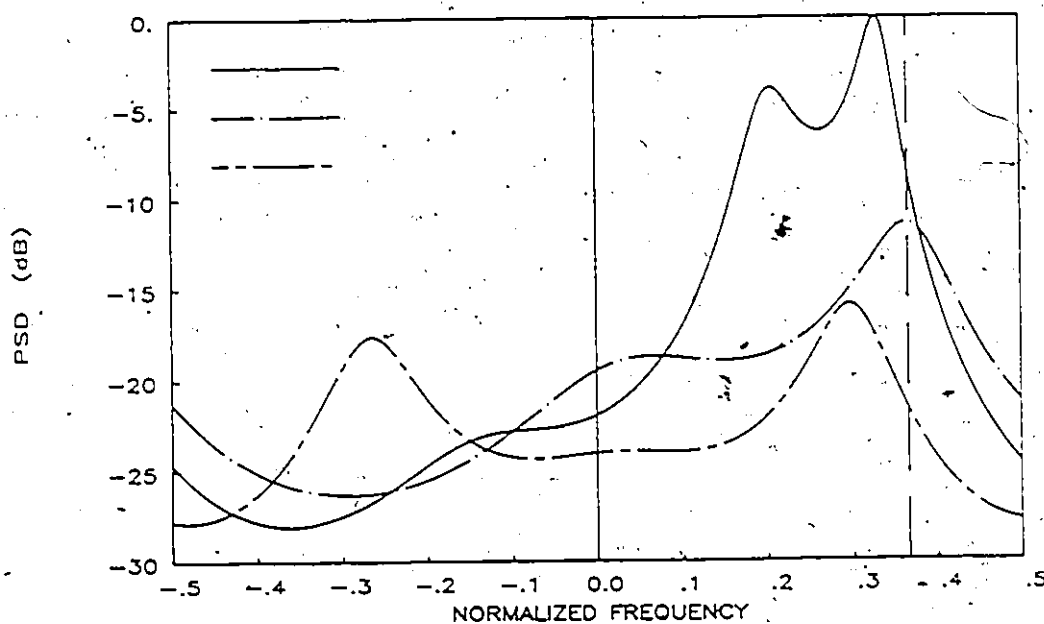
Fig. 6.4. Filter outputs in frequency in response to ground clutter.

(a) Target and clutter peaks are completely separated.

(b) Two peaks are partly overlapped.



(a)



(b)

Fig.6.5 Filter outputs in frequency in response to weather clutter.

(a) Target and clutter peak are completely separated.

(b) Two peaks are partly overlapped.

environment, however, the PEF does not work so well as in ground clutter situations, as shown in Figure 6.5. Even when the target Doppler is separated from the clutter peak, the target component at the PEF output is only slightly enhanced, compared to the clutter component. When the two peaks representing clutter and target get close, the PEF becomes very poor in performance. In both cases, the canceller works well.

6.6 SYSTEM OUTPUT IN THE TIME DOMAIN

In the previous section, we studied experimentally the canceller output in terms of power spectra. In this section, we will investigate the system response in the time domain.

In Figure 6.6 we show the plots of the amplitude response of the input and outputs using the PEF and clutter canceller for target-plus-ground clutter. The ground clutter data used for this experiment are taken from data file 'WTRNG1', located between azimuth 1275 and 1375 on range ring 20. A large number of data samples, say 100 in our own experiment, has been chosen since the PEF requires time to adapt its tap weight coefficients to suit the input clutter process. This condition, however,

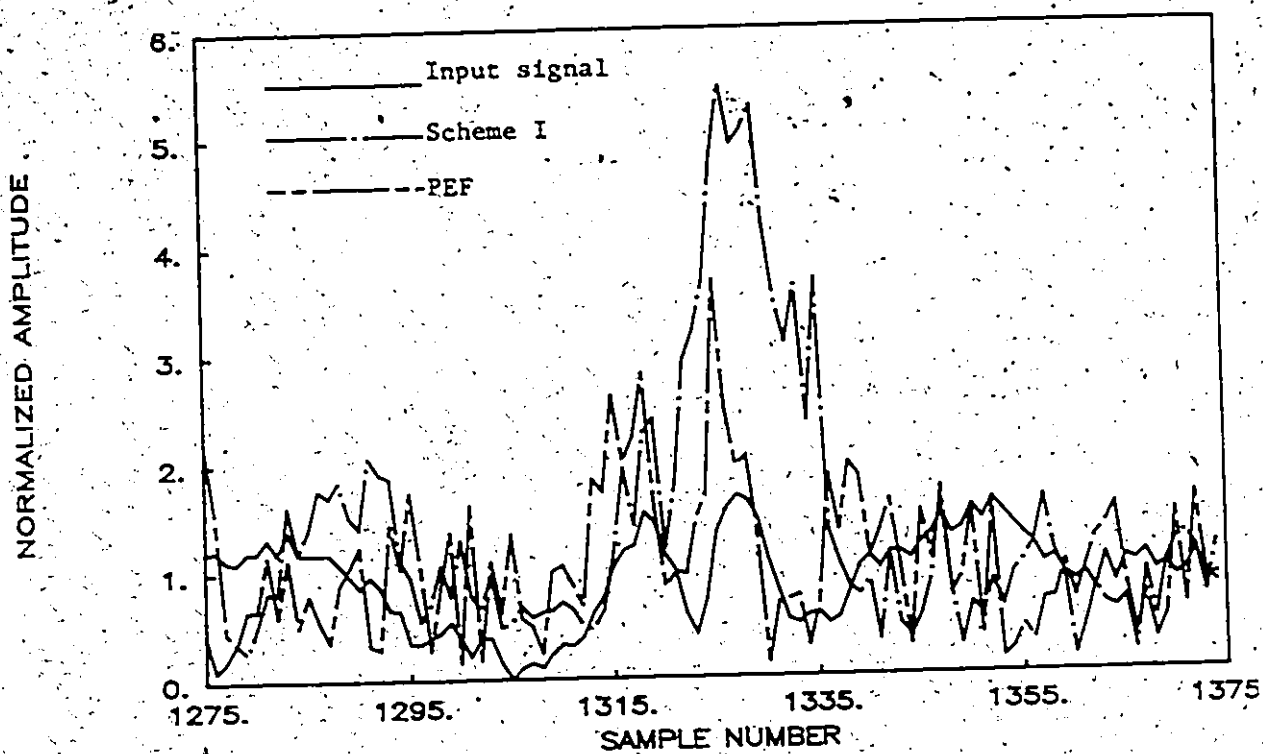


Fig. 6.6 Filter outputs in time domain in response to ground clutter.

is not necessary in the suppression scheme I. A target signal is located at the center of the clutter, that is, between azimuth 1315 and 1335. The SCR is defined as the ratio of target power averaged over 20 samples to the clutter power evaluated from 100 samples. In Figure 6.6 $SCR = -3dB$. After processing by the clutter filters, both target and clutter component are weakened to some extent. But the only thing we are concerned about is the enhancement of the target component compared to the clutter. Thus, we normalize the input and output signals of the clutter filters so that the clutter components are of unit power; this facilitates the comparison of different schemes. Based on the input amplitude response, we cannot recognize the presence of a target echo. But after signal processing by the PEF and clutter suppression scheme I, the target signals become distinctly visible. The noticeable differences between the two schemes are:

- (1) The output target component of clutter suppression Scheme I is much larger than that of the PEF in amplitude.

- (2) The target component at the output of suppression Scheme I possesses nearly the same duration as the time target signal. This is in contrast to that of the PEF which is enhanced only for a couple of sample points.

(3) The clutter suppression scheme I provides a much greater output target power than the PEF does.

Similar results can be observed for weather clutter. Figure 6.7 shows the normalized amplitudes of the output signals of the PEF and canceller in response to an input signal composed of target plus weather clutter. The clutter data samples are chosen from data file 'WTRNWL', located between azimuths 3110 and 3210 on range ring 15, and the target echo is placed between azimuth 3150 and 3170. The SCR equals -3dB. We again see that the canceller is much superior to the PEF in clutter suppression performance.

6.7 IMPROVEMENT FACTOR

Improvement factor (I.F.) is an important overall performance measure of a radar clutter-suppression filter. Theoretical analysis of improvement factor of clutter suppression scheme I was discussed in Sections 6.2 to 6.3. In this section we will present experimental results to support the theory therein.

The improvement factor is evaluated experimentally in the following way. Choose 30 blocks of clutter data,

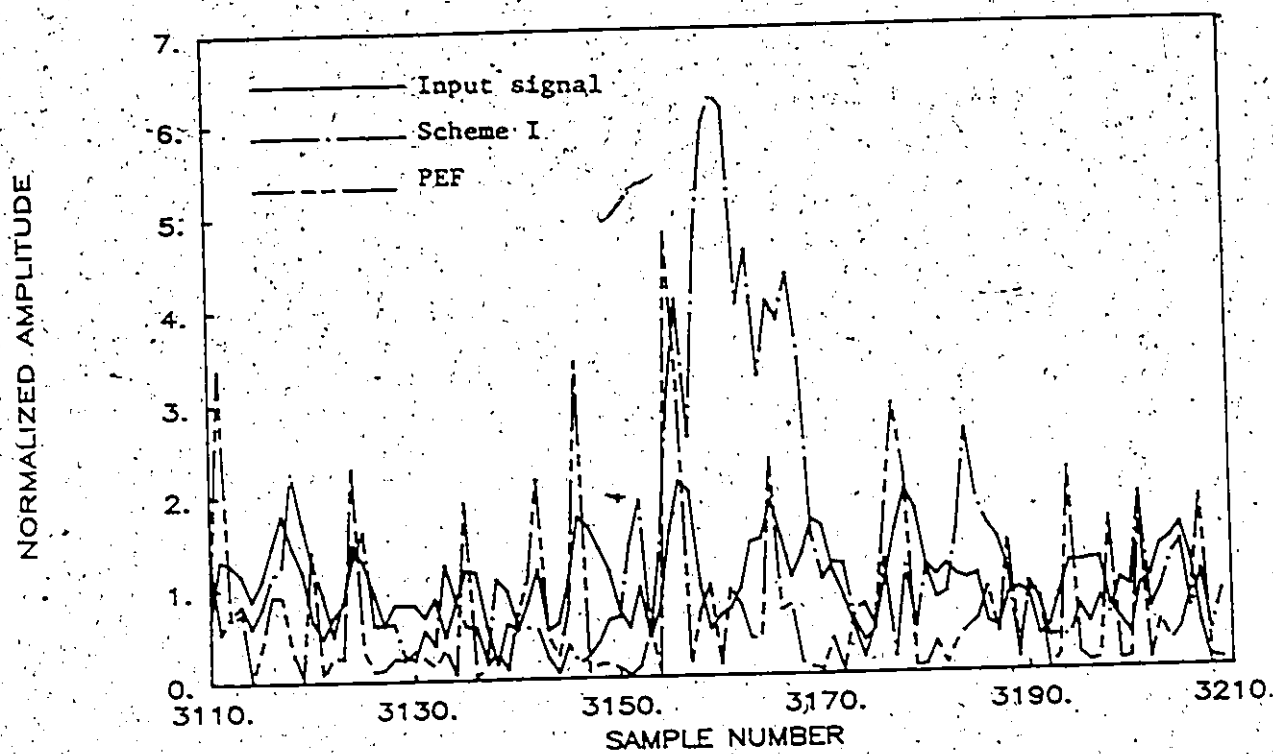


Fig. 6.7 Filter outputs in time domain in response to weather clutter

located at different azimuths and range rings. The data sets so chosen reflect general characteristics of the clutter under study. Each block of data contains 130 samples and a target will be placed on its center. Then fix the target Doppler frequency and apply clutter plus target signal to the clutter filters. We calculate the improvement factor for each block of clutter data, which we call local improvement factor. Taking the average over 30 local values, we get the improvement factor for a given target Doppler. Finally, changing the target Doppler, we obtain an improvement factor for all possible target Doppler frequencies. The results so obtained are shown in Figures 6.8 and 6.9; Figure 6.8 for weather clutter and Figure 6.9 for ground clutter. From these diagrams we make the following observations.

(1) Clutter suppression scheme I is always better than the PEF over all possible Doppler frequencies, having 5dB improvement for weather clutter and 4dB improvement for ground clutter.

(2) There are troughs in improvement factor versus target Doppler curves. In particular, the trough occurs when the target Doppler is close to a clutter peak. This is easy to understand since it is difficult to design a filter to separate signal from interference component

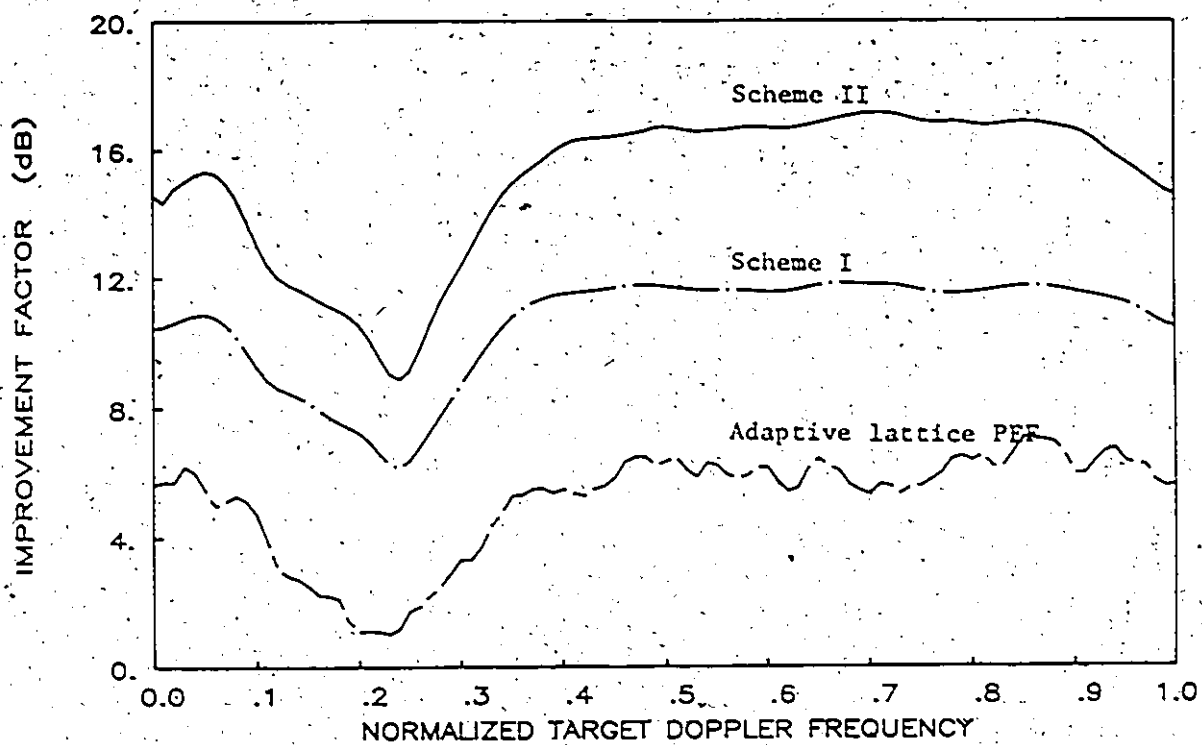


Fig.6.8: Comparison of the performance of different radar clutter filters.

Results shown are for weather clutter.

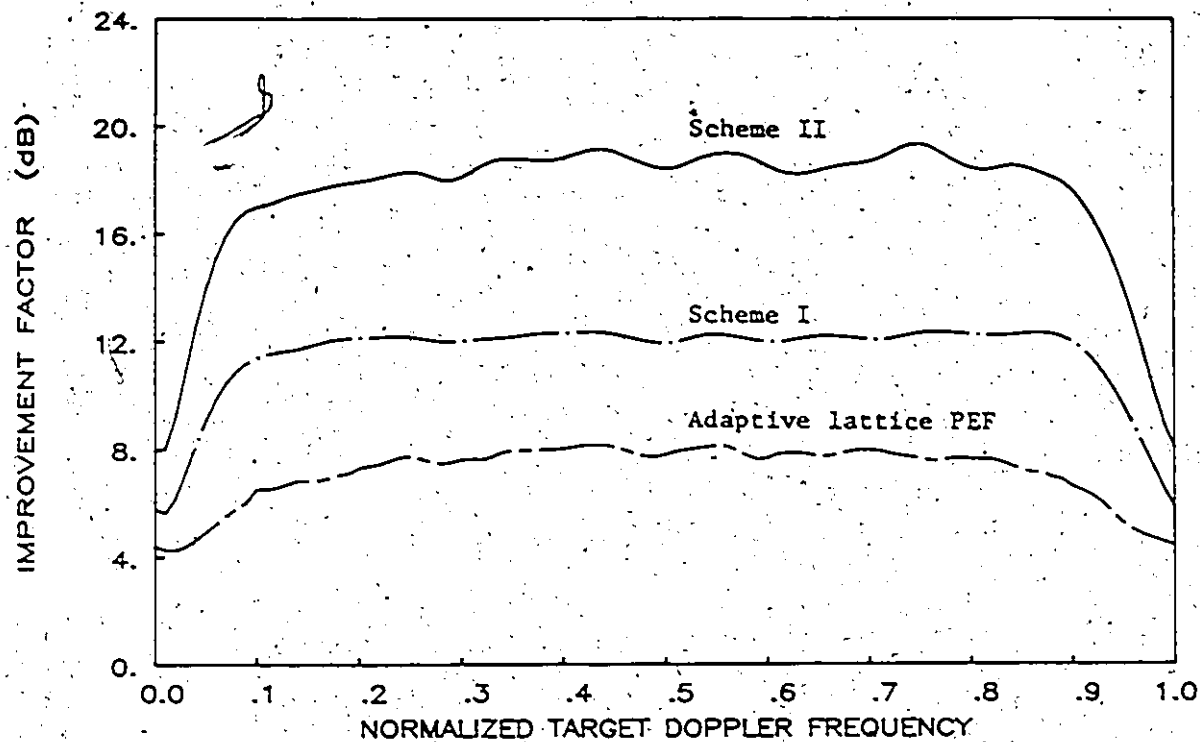


Fig. 6.9 Comparison of the performance of different radar clutter filters.

Results shown are for ground clutter.

9

if they overlap in frequency.

In practice, it is possible for a radar target to take on any value in some range and therefore, it is desirable to evaluate a clutter suppression filter in terms of its averaged improvement factor. In order to calculate the averaged improvement factor, however, one must know the probability density of the target velocity over some range. Generally, we do not have such information. To simplify the discussion, it is convenient to assume that the target Doppler is uniformly distributed over some frequency range, say, from zero Hz to pulse repetition frequency (PRF). Thus, we get an averaged improvement factor for the PEF and suppression Scheme I as follows:

<u>Suppression Filter</u>	<u>Ground Clutter</u>	<u>Weather Clutter</u>
Scheme I	11.429 dB	10.546
PEF	7.097	5.1007

These results show that scheme I produces an average performance improvement over the PEF by 4.33dB and 5.46dB for ground and weather clutter, respectively.

6.8 FURTHER IMPROVEMENT -- SUPPRESSION SCHEME II

After processing the radar data with Scheme I, the clutter component is significantly reduced. However, there is still some wideband interference in the canceller output, which is partly due to thermal noise at front end of the receiver and partly due to clutter residual. The latter component can be rationalized as below. Owing to the existence of white noise at the canceller input, the clutter/noise-density ratio will vary with frequency, assuming its maximum at the frequency where clutter peak occurs. As the canceller tries to minimize the output power, the weight vector of the canceller is adjusted such that a significant part of the clutter component passes through the reference channels and is subsequently cancelled at the output for frequencies corresponding to high clutter/noise-density ratio. On the other hand, small weight is given to those frequencies that pertain to low clutter/noise-density ratio. Accordingly, the clutter residual at the canceller output is flattened out in the spectrum.

In order to filter out this wideband interference, a one-step-forward linear prediction filter (LPF) operating in an adaptive manner is used immediately after

the canceller described above, as shown in Figure 6.1(c). We thus get another cancellation scheme, termed Scheme II.

A linear prediction filter tends to enhance narrowband signal while it suppresses wideband component. The reason is quite simple and intuitive. A signal with narrow bandwidth, according to Fourier transform, has a wide autocorrelation function. Consequently, the corresponding time series is easy to predict based on past data samples. The autocorrelation coefficient of a wideband signal, on the other hand, is very small, except when the time lag equals zero; therefore, a wideband signal is difficult to produce. As an extreme case, white noise is completely unpredictable.

In a similar way, it may be argued that a prediction-error-filter has a tendency to enhance wideband signal. The PEF, combined with an LPF, can be used to separate two signal components with noticeable differences in frequency bandwidth: the wideband signal mainly comes out from the PEF and the LPF enhances the narrowband component.

The clutter suppression scheme II can be effectively evaluated in terms of the improvement factor. This scheme has been tested with actual weather and ground clutter respectively, and the results are shown in Figures 6.8 and 6.9.

To implement suppression Scheme II, two Kalman filters with a random walk model are employed, both having order 4. For the design of this type of filter, see chapter 2 for details. For comparison, the adaptive PEF, suggested by Gibson and Haykin [17] is also included. The adaptive PEF is implemented with a gradient lattice algorithm. Its parameters are: filter length equals 6, and the memory factor is experimentally determined to be 0.6.

Performance comparison of these filters is summarized in Table 6.1 in terms of averaged improvement factor.

TABLE 6.1

Comparison of Improvement Factor (in dB) of the clutter filters

<u>Filter Type</u>	<u>Ground Clutter</u>	<u>Weather Clutter</u>
Scheme I	11.429	10.546
Scheme II	17.314	15.015
PEF	7.0966	5.101

These results show that Scheme I and Scheme II produce average performance improvement over the lattice PEF by 4.3dB and 10.2dB, respectively.

For weather clutter, the improvements obtained with the new schemes I and II are 5.4dB and 9.9dB, respectively.

The computational complexities of the new schemes I and II and the PEF are included in Table 6.2.

TABLE 6.2

Computational Complexity of Clutter Filters

Filter Type	Multiplications	Additions
Scheme I	48	44
Scheme II	96	88
PEF (order 6)	54	30

We see that Scheme I requires less computations than the PEF and yet it offers a considerably better performance. Scheme II is much superior to the PEF in performance, and superior to a slightly lesser extent than Scheme I; however, its computational complexity is the highest.

CHAPTER 7

CONCLUSIONS

The problem of detecting a random signal in a background of colored plus white Gaussian noise, both having unknown statistics, is of theoretical importance and practical interest. There is evidently no uniformly most powerful (UMP) test for such a problem. Several methods proposed previously in the literature seem to be somewhat ad hoc. The question is whether one can design a better receiver for a given detection environment. In this thesis we have developed some effective solutions to the problem for an environment for which reference data are available.

The major contributions of the thesis are summarized as follows:

(1) We have developed a general procedure for detecting a random signal in colored plus white Gaussian noise, all having unknown statistics. In particular, techniques are provided for the following:

(a) Model building for detection problem:
the concept of characteristic projection
space is introduced to choose model

candidates, and the whitening test and η -loss test are suggested for performing "initial" tests on them.

- (b) Derivation of test statistic based on the generalized LLR.
- (c) Development of an adaptive detection algorithm.

(2) The adaptive detection procedure has been successfully applied to solve the radar detection problem. Experimental results using actual as well as simulated radar data show that the new detection scheme is much better than the previously developed IBDA algorithm in performance.

(3) We have provided an intuitive geometrical interpretation for the new receiver: A generalized LLR receiver requires to construct two projection spaces for the two hypotheses, and the decision is made by comparing the distances of the newly received sample to these two spaces.

(4) We have developed two new radar clutter suppression schemes based on the idea of multiple-reference noise cancellation. We have tested their performance by using actual radar data, which shows that the new schemes are around 5dB and 10dB better than the adaptive lattice PEF, respectively, in terms of improvement factor.

In previous chapters, a general detection method based on linear modelling has been described. In chapter 5, this method has been successfully applied to solve the radar detection problem. However, it can be applied, at least in principle, to a variety of detection problems, where reference sequences are available. In this connection potential applications may be biomedical and geophysical signal processing.

REFERENCES

- [1] Akaike, H., "Fitting Autoregressive Models for Prediction", Ann. Inst. Statist. Math., Vol. 21, pp. 243-247, 1969.
- [2] Akaike, H., "On Entropy Maximization Principle", Krishnaiah, ed., Applications of Statistics. North-Holland (1977).
- [3] Akaike, H., "A New Look at the Statistical Model Identification", IEEE Trans. Automat. Contr., Vol. AC-19, pp. 716-723, 1974.
- [4] Anderson, B.D.O., and J.B. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs, N.J., 1979.
- [5] Anderson, T.W., The Statistical Analysis of Time Series, Wiley, New York, 1971.
- [6] Anderson, T.W., "Estimation for Autoregressive Moving Average Models in Time and Frequency Domains", Ann. Stat., Vol. 5, pp. 842-865, Oct. 1977.
- [7] Barlow, E.J., "Doppler Radar", Proc. IRE, Vol. 37, pp. 340-355, April 1949.
- [8] Berkowitz, R.S., Modern Radar: Analysis, Evaluation, and System Design, Wiley, New York, 1965.
- [9] Box, G.E.P. and G.M. Jenkins, Time Series Analysis: Forecasting and Control, Revised edition, Holdey-Day, 1976.
- [10] Brillinger, D.R., Time Series: Data Analysis and Theory, Holt, Rinhart and Winston, Inc., New York, 1975.
- [11] Currie, B.W. and S. Haykin, "Radar Clutter Classifier Prototype for Use with the TRACS Airport Surveillance Radar", in Proc., 1st North American Birdstrike Prevention Workshop (Ottawa, Ontario, September 17-18, 1981). pp. 6.1-6.30.

- [12] Dickinson, B.W., "Properties and Applications of Gaussian Autoregressive Processes in Detection Theory", IEEE Trans. Information Theory, Vol. IT-27, pp. 343-347. May 1981.
- [13] Doob, J.L., Stochastic Processes, New York: Wiley and Sons, 1953.
- [14] Drape, N.R., and H. Smith, Applied Regression Analysis, 2nd edition, Wiley, New York, 1981.
- [15] Friedlander, B. and B. Porat, "A Spectral Matching Technique for ARMA Parameter Estimator", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-32, pp. 338-343. April 1984.
- [16] Gibson, C., S. Haykin and S. Kesler, "Maximum Entropy (adaptive) Filtering Applied to Radar Clutter", Proc. 1979 IEEE Int. Conf. Acoust., Speech, Signal Processing, 79, pp. 166-169.
- [17] Gibson, C. and S. Haykin, "Learning Characteristics of Adaptive Lattice Filtering Algorithms", IEEE Trans. Acoust., Speech Signal Processing, Vol. ASSP-28, pp. 681-691. December 1980.
- [18] Gibson, C., "Adaptive Lattice Filtering for Radar Applications", Ph.D. Dissertation, McMaster University, Hamilton, Ontario, Canada, 1982, pp. 131-135.
- [19] Godard, D., "Channel Equalization Using a Kalman Filter for Fast Data Transmission", IBM J. Research and Development, May 1974, pp. 267-273.
- [20] Goldstein, H., "Sea Echo, The Origins of Echo Fluctuation, and the Fluctuations of Clutter Echoes", Propagation of Short Radio Waves, D.E. Kerr, ed., MIT Radiation Laboratory Series, Vol. 13, Secs. 6.6-6.21, pp. 560-587, New York: McGraw-Hill, 1951.
- [21] Goodwin, G.C. and R.L. Payne, Dynamic System Identification: Experiment Design and Data Analysis. Academic Press, New York, 1977.
- [22] Goodwin, G.C. and K.S. Sin, Adaptive Filtering, Prediction and Control, Prentice-Hall, Englewood Cliffs, N.J., 1984.

- [23] Hannan, E.J., Multiple Time Series, Wiley, New York, 1970.
- [24] Hansen, V.G., Clutter Suppression in Search Radar, pp. 397, Radars Vol. 7, Ed., D. Barton, 1976.
- [25] Haykin, S., Editor, Nonlinear Methods of Spectral Analysis, 2nd edition, Springer-Verlag, New York, 1983.
- [26] Haykin, S., Introduction to Adaptive Filters, Macmillan, New York, 1984.
- [27] Haykin, S., B.W. Currie and S.B. Kesler, "Maximum-Entropy Spectral Analysis of Radar Clutter", Proc. IEEE, Vol. 70, pp. 953-962, September 1982.
- [28] Haykin, S., S. Kesler and B. Currie, "An Experimental Classification of Radar Clutter", Proc. IEEE, Vol. 67, No. 2, pp. 332-333, Feb. 1979.
- [29] Haykin, S., Adaptive Filter Theory, Prentice Hall, 1986.
- [30] Hsu, F.M., "Square Root Filtering for High Speed Data Received Over Fading Dispersive H.F. Channels", IEEE Trans. Information Theory, Vol. IT-28, pp. 753-763, September 1982.
- [31] IEEE Standard Dictionary Electrical and Electronics Terms, Wiley, New York, 1977.
- [32] Jazwinski, A.H., Stochastic Processes and Filtering Theory, Academic Press, 1970.
- [33] Kailath, T., "A General Likelihood-Ratio Formula for Random Signals", IEEE Trans. Information Theory, Vol. IT-15, pp. 350-361, May 1969.
- [34] Kailath, T., "Likelihood Ratios for Gaussian Processes", IEEE Trans. Information Theory, Vol. IT-16, pp. 276-288, May 1970.
- [35] Kailath, T., Lectures in Wiener and Kalman Filtering, Springer-Verlag, New York, 1981.

- [36] Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems", Trans.-ASME, J. Basic Eng., Vol. 82, pp. 34-45, 1960.
- [37] Kendall, M. and A. Stuart, The Advanced Theory of Statistics, Vol. 2, 4th edition, Macmillan Publishing Co. Inc., New York, 1979.
- [38] Kendall, M.G. and A. Stuart, The Advanced Theory of Statistics, Vol. 2, p. 249, New York: Hafner Publishing Co., 1961.
- [39] Kesler, S., "Nonlinear Spectral Analysis", Ph.D. Thesis, McMaster University, Hamilton, Ontario, Canada, 1978.
- [40] Kitagarwa, G., "A Nonstationary Time Series Model and Its Fitting by a Recursive Filter", Journal of Time Series Analysis, Vol. 2, No. 2, pp. 103-116, 1981.
- [41] Landau, I.D., "A Feedback System Approach to Adaptive Filtering", IEEE Trans. Information Theory, Vol. IT-30, pp. 251-262, 1984.
- [42] Lancaster, P. Theory of Matrices, Academic Press, New York, 1969.
- [43] Ledolter, J., "A Recursive Approach to Parameter Estimation in Regression and Time Series Models", Commun. Statist.-Theor. Meth., A8(12), 1227-1245, (1979).
- [44] Lehmann, E.L. Testing Statistical Hypothesis, New York: Wiley and Sons, 1959.
- [45] Ljung, L. and T. Soderstrom, Theory and Practice of Recursive Identification, MIT Press, 1983.
- [46] Makhoul, J., "Linear Prediction: A Tutorial Review", Proc. IEEE, Vol. 63, 1975, pp. 561-580.
- [47] Messerschmitt, D.G., "A Geometric Theory of Intersymbol Interference, Part I: Zero-Forcing and Decision-Feedback Equalization", BSTJ, Vol. 52, No. 9, pp. 1483-1519, Nov. 1973.
- [48] Metford, P.A.S., S. Haykin and D.P. Taylor, "An Innovation Approach to Discrete-Time Detection Theory", IEEE Trans. Information Theory, Vol. IT-28, pp. 376-380, 1982.

- [49] Metford, P.A.S., and S. Haykin, "Experimental Analysis of an Innovations-Based Detection Algorithm for Surveillance Radar", Proc. IEE, Vol. 132, pp. 18-26, Feb. 1985.
- [50] Miller, R.J., "Spatial Correlation in Ground Clutter", GEC Journal of Research, Vol. 1, No. 2, pp. 122-126, 1983.
- [51] Mueller, M.S., "Least-Squares Algorithms for Adaptive Equalizers", BSTJ, Vol. 60, No. 8, 1981, pp. 1905-1925.
- [52] Nau, R.F. and R.M. Oliver, "Adaptive Filtering Revisited", Journal of the Operational Research Society, Vol. 30, No. 9, pp. 825-31, 1979.
- [53] Nicholson, G. and J.P. Norton, "Kalman Filter Equalization for a Time-Varying Communication Channel", A.T.R., Vol. 13, pp. 3-11, No. 1.
- [54] Noble, B., Applied Linear Algebra, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1969.
- [55] Papoulis, A., Probability, Random Variable and Stochastic Processes, McGraw-Hill, 1965.
- [56] Parzen, E., "Some Recent Advances in Time Series Modelling", IEEE Trans. Control, Vol. AC-19, pp. 723-729, December 1974.
- [57] Parzen, E., Multiple Time Series Modelling: Determining the Order of Approximating Autoregressive Schemes, in Multivariate Analysis, IV, Ed., P. Krishnalah, pp. 283-295, North-Holland, 1977.
- [58] Parzen, E., "An Approach to Time Series Modelling and Forecasting Illustrated by Hourly Electricity Demands", Statistical Science Division, State University of New York at Buffalo, Tech. Rep. 37, Jan. 1976.
- [59] Priestley, M.B., Spectral Analysis and Time Series, Vols. 1 and 2, Academic Press, New York, 1981.
- [60] Porat, B. and B. Friedlander, "Parametric Techniques for Adaptive Detection of Gaussian Signals", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-32, pp. 780-790, Aug. 1984.

- [61] Rissanen, J., "Modelling by Shortest Data Description", Automatica, Vol. 14, pp. 467-471, 1978.
- [62] Rissanen, J., "Consistent Order Estimation of Autoregressive Processes by Shortest Description of Data", Analysis and Optimization of Stochastic Systems, Jacobs et. al., Eds., New York: Academic Press, 1980.
- [63] Rissanen, J., "A Universal Prior for the Integers and Estimation by Minimum Description Length", Ann. Stat., Vol. 11, pp. 417-431, 1983.
- [64] Root, W.L., "An Introduction to the Theory of the Detection of Signals in Noise", Proc. IEEE, Vol. 58, pp. 610-623, May 1970.
- [65] Rummler, W.D., "Clutter Suppression by Complex Weighting of Coherent Pulse Trains", IEEE Trans. Aerospace and Electronic Systems, Vol. AES-2, pp. 689-699, November 1966.
- [66] Schwartz, G., "Estimating the Dimension of a Model", Ann. Stat., Vol. 6, pp. 461-464, 1978.
- [67] Searle, S.R., Linear Models, Wiley, New York, 1971.
- [68] Silverman, R.A., "Locally Stationary Random Processes", IRE Trans. Information Theory, Vol. IT-3, pp. 182-187, Sept. 1957.
- [69] Skolnik, M.I., Introduction to Radar Systems, McGraw-Hill, 1980, 2nd edition, Chapt. 4.
- [70] Stehwien, W. and S. Haykin, "Classification of Radar Clutter Based on Spectral Parameters", Int. Conf. on Radar, pp. 126-131, Paris, 1984.
- [71] Stehwien, W., "A Clutter Classifier-Preliminary Results", Inter. Rept. 135, Communications Res. Lab., McMaster University, Ontario, Canada, 1985.
- [72] Stewart, G.W., Introduction to Matrix Computation, Academic Press, New York, 1973.

- [73] Taylor, J.W., "Sacrifices on Radar Clutter Suppression Due to Compromises in Implementation of Digital Doppler Filters", IEEE Int. Radar Conf., 1982.
- [74] Tretter, S.A., Introduction to Discrete-Time Signal Processing, Wiley and Sons, New York, 1976.
- [75] Widrow, B., et. al., "Adaptive Noise Cancelling Principles and Applications", Proc. IEEE, Vol. 63, pp. 1692-1716, 1975
- [76] Wilks, S.S., Mathematical Statistics, Wiley, New York, 1963.
- [77] Wold, H., A Study in Analysis of Stationary Time Series, Almquist & Wiskell, Uppsala, 1938.
- [78] Zhang, Q.T. and S. Haykin, "Radar Clutter-Suppression Schemes", Electronics Letters, Vol. 20, pp. 1007-1008, Nov. 1984.
- [79] Zhang, Q.T. and S. Haykin, "Tracking Characteristics of the Kalman Filter in a Nonstationary Environment for Adaptive Filter Applications", Proceedings ICASSP-83. Vol. 2, pp. 671-674, Boston, 1983.
- [80] Zhang, Q.T. and S. Haykin, "Adaptive Radar Detection", Electronics Letters, Vol. 21, No. 18, pp. 808-810, August, 1985.