Bistatic STAP Training Without Navigation Data

LIM Chin Heng, Elias Aboutanios and Bernard Mulgrew
Institute for Digital Communications
School of Engineering & Electronics
University of Edinburgh, King's Buildings
Edinburgh EH9 3JL, United Kingdom

C.H.Lim@ed.ac.uk, elias@ieee.org, B.Mulgrew@ed.ac.uk

ABSTRACT

Optimum STAP requires knowledge of the true interference covariance matrix. In practice, this matrix is not known and must be estimated from training data, which must be target free and statistically homogeneous with respect to the range gate under test. These conditions are often not satisfied, which degrades the detection performance. Particularly for bistatic ground moving target indication (GMTI) radar, the clutter Doppler frequency depends on range for all array geometries. This range dependency leads to problems in clutter suppression through STAP techniques. In this paper, we study issues associated with applying two novel STAP techniques, which minimize the amount of navigation data associated with both the transmitter and receiver. Performance results against existing techniques are given in terms of improvement factor loss plots and receiver operating curves.

1 Introduction

Space-time adaptive processing (STAP) is a well-established technique for detection of moving targets by an airborne radar. Interest in bistatic STAP, where the transmitter and receiver are separated, has picked up in recent years. Bistatic radar offers several advantages over its monostatic counterpart, such as the higher possibility of detecting stealth targets. Training and updating of the clutter covariance matrix is a key step in the implementation and effectiveness of any STAP system. In a bistatic or multi-static environment, the usual impediment and possible clutter in-homogeneity experienced in the linear monostatic side-looking case is further complicated by the range-dependent nature of the clutter ridge in the angle-Doppler plane induced by the physical geometry of the two (or more) aircraft [1]. Variations in the clutter covariance matrix with range in the bistatic case are thus due to a combination of these two factors, that is: (i) in-homogeneity of ground scatterers; (ii) the physical geometry of the aircraft.

Lim, C.H.; Aboutanios, E.; Mulgrew, B. (2006) Bistatic STAP Training Without Navigation Data. In *Bistatic-Multistatic Radar and Sonar Systems* (pp. 29-1 – 29-26). Meeting Proceedings RTO-MP-SET-095, Paper 29. Neuilly-sur-Seine, France: RTO. Available from: http://www.rto.nato.int/abstracts.asp.

Thus the bistatic range-dependent clutter spectrum complicates the clutter suppression problem and leads to significant degradation in performance. A mismatch in the clutter statistics between the training range cells and the test range cell will result in the widening of the STAP filter clutter notch. This will cause target returns from relatively slow-velocity or low-flying targets to be suppressed or even go undetected.

A number of compensation approaches exist to mitigate the impact of this range dependency on clutter suppression, such as Doppler compensation in the angle-Doppler domain [2] and angle-Doppler compensation [3]. However, these techniques effectively only manage to map the mainlobe clutter peak from one range gate to another and they take little account of sidelobe clutter. Other techniques have been proposed that attempt to map the sidelobe clutter, like angle and separate Doppler compensation [4], registration-based methods [5] and estimation of clutter covariance by configuration system parameter estimation [6].

The major draw-back with these techniques is that they require knowledge of the radar system navigational data, which has to be constantly updated adaptively. Although the data can be estimated, like in [6] and [7], an initial estimate or knowledge of the theoretical direction-Doppler curves are required respectively. A new technique was proposed in [8] to predict the range-dependent inverse covariance matrix using least squares estimation. No navigational data or parameters estimation has to be performed as only the clutter data is required. Moreover, the technique is not restricted to uniform linear array (ULA) applications.

Conventional STAP techniques rely on the availability of target-free training data, which must satisfy the homogeneity assumption. Quite often, the training data is not target-free nor is it homogeneous with the test data, which leads to a degradation in the detection performance. In order to alleviate the high target density or heterogeneity problem and therefore improve the detection performance, research had been carried out in knowledge-based (KB-STAP) techniques, [9], [10] and [11], which incorporate apriori knowledge into the processor.

On the other hand, in-the-gate processing or single data set detection algorithms, namely the generalised maximum likelihood estimation detector (GMLED) and the maximum likelihood estimation detector (MLED), have been proposed in [12] and [13] to circumvent the problem where such training data are not available. This approach forgoes the training data requirement and operates solely on the test data.

Our work in this area is directed towards what might be called “small STAP”, where the number of spatial channels is small and the array is non-uniform and by a desire to minimize the amount of navigation data associated with both the transmitter and receiver. Further, it is directed towards environments where all range gates contain targets. This paper presents the two techniques that we have recently developed to address the problems within a bistatic airborne radar framework. These techniques will be compared and contrasted with existing ones in terms of both performance and complexity.

In what follows, bold lower case and upper case letters are used to denote column vectors and matrices respectively. The superscripts $()^*$, $()^T$ and $()^H$ denote the complex conjugate, transpose

and Hermitian operations respectively.

2 Problem Statement

Consider a radar system utilizing an N_s-element array with inter-element spacing d. The radar transmits an M_t-pulse waveform in its coherent processing interval (CPI). The received data can then be partitioned in both space and time, by using a sliding window, into an $(N \times M)$ space-time snapshot $\mathbf{X}'$. This partitioning will result in $K_T = (N_s - N + 1)(M_t - M + 1)$ snapshot matrices being generated for processing.

The columns of these space-time snapshots are then stacked into inter-leaved column vectors $\mathbf{x}_k$ of size $(NM \times 1)$. The K_T columns are then arranged as the columns of the $(NM \times K_T)$ matrix $\mathbf{X}$. The signal model used is then given as:

$$\mathbf{X} = \alpha \mathbf{s}\mathbf{t}^T + \mathbf{N} \tag{1}$$

where both $\mathbf{s}$ and $\mathbf{t}$ are space-time vectors and α is a complex amplitude.

$\mathbf{N}$ is the $(NM \times K_T)$ zero-mean Gaussian clutter-plus-noise (interference) matrix with independent and identically distributed (iid) columns $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$. The space-time clutter-plus-noise covariance matrix is defined as $\mathbf{C}$, where $\mathbf{C} = E[\mathbf{N}\mathbf{N}^H]$ and $E[*]$ is the expectation operator. Generally, the detection problem is treated as a hypothesis test with the null and alternative hypotheses:

$$H_0 : \quad \mathbf{X} = \mathbf{N} \tag{2}$$

$$H_1 : \mathbf{X} = \alpha \mathbf{s}\mathbf{t}^T + \mathbf{N} \tag{3}$$

The optimum processor (Wiener filter) weights, of the Neyman-Pearson's [14] likelihood ratio test, are given as $\mathbf{w}_{opt} = \beta \mathbf{C}^{-1}\mathbf{s}$ in [15]. For a single data snapshot $\mathbf{x}$, which can be any one columns of $\mathbf{X}$, the resulting filter output y is:

$$y = \mathbf{w}^H \mathbf{x} = \beta^* \mathbf{s}^H \mathbf{C}^{-1} \mathbf{x} \tag{4}$$

and β is a complex arbitrary constant. The optimum processor can be viewed as a whitening filter stage followed by a matched filter stage [1]. The filter output power, $Y = |y|^2$, is then compared to a threshold γ, for a certain probability of false alarm P_{fa}. The detection test becomes

$$Y \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{5}$$

For any signal detection algorithm, a highly desirable property is the constant false alarm rate (CFAR) property. By selecting a suitable value of β, the optimum processor can be made to possess the CFAR property. The CFAR matched filter (MF) is obtained by setting $\beta = \left(\mathbf{s}^H\mathbf{C}^{-1}\mathbf{s}\right)^{-\frac{1}{2}}$ in the hypothesis test of equation (5).

Traditional detection algorithms, such as the generalised likelihood ratio test (GLRT) [16] or the adaptive matched filter (AMF) [17], require target-free training data that must be homogeneous with the test data. To implement the detection test, an estimate of the interference covariance matrix, $\widehat{\mathbf{C}} = \frac{1}{K_t}\sum_{k=0}^{K_t}\mathbf{z}_k\mathbf{z}_k^H$, is obtained from the secondary data $\mathbf{z}_k$ of size K_t. This gives the sample matrix inversion (SMI) algorithm [15].

These K_t snapshots are usually obtained from other range cells [18], thus making the algorithms vulnerable to heterogeneity problems and resulting in a degradation in detection performance [19]. Two alternative approaches, the generalised maximum likelihood estimation detector (GMLED) and maximum likelihood estimation detector (MLED) were proposed in [13] to circumvent the problem where such training data are not available. They only require the set of test data and thus can be called the single data set detection (SDSD) algorithms.

Besides the detection test, the efficiency of any linear processors can also be characterised by the improvement factor (IF) [1], which is defined as the ratio of signal-to-noise power ratios at the output and input.

$$IF = \frac{\mathbf{w}^H\mathbf{s}\mathbf{s}^H\mathbf{w}\cdot\mathrm{tr}(\mathbf{C})}{\mathbf{w}^H\mathbf{C}\mathbf{w}\cdot\mathbf{s}^H\mathbf{s}} \tag{6}$$

where $\mathrm{tr}(\mathbf{C})$ is the trace of $\mathbf{C}$; $\mathbf{s} = \mathbf{s}_t \otimes \mathbf{s}_s$ and $\otimes$ is the Kronecker product. The temporal and spatial dimensions of the target steering vector are respectively:

$$\mathbf{s}_t = \left[1\, e^{j2\pi f_D}\ \ldots\ e^{j2\pi(M-1)f_D}\right]^T \tag{7}$$

$$\mathbf{s}_s = \left[1\, e^{j2\pi f_s}\ \ldots\ e^{j2\pi(N-1)f_s}\right]^T. \tag{8}$$

For a bistatic forward-looking radar configuration, as shown in Fig. 1, the normalized target Doppler and spatial frequencies are respectively:

$$f_D = \frac{v_{rad,rx} + v_{rad,tx}}{\lambda\mathrm{PRF}} \tag{9}$$

$$f_s = \frac{d\sin\varphi_r\cos\theta_r}{\lambda}. \tag{10}$$

$v_{rad,rx}$ and $v_{rad,tx}$ are the relative velocities of the target to the receiver and transmitter respectively, λ is the wavelength of the radar signal, PRF refers to the pulse repetition frequency of the radar system, φ_r and θ_r are the azimuth and elevation angle of the receiver to the target on the ground respectively.

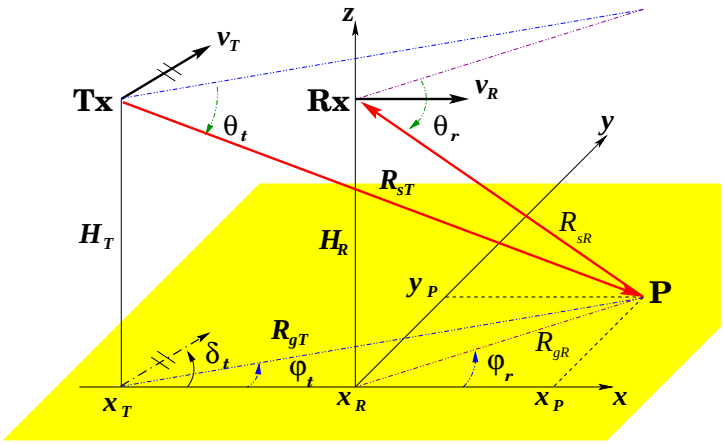

Figure 1: Geometry of an airborne bistatic radar.

Following (6), another performance metric that will be used to evaluate the performance of the processors in this paper is:

$$\begin{aligned} IF_{loss} &= \frac{IF_{opt}}{IF_{STAP}} \\ &= \frac{\mathbf{s}^H\mathbf{C}^{-1}\mathbf{s}\mathbf{s}^H\widehat{\mathbf{C}}^{-1}\mathbf{C}\widehat{\mathbf{C}}^{-1}\mathbf{s}}{\mathbf{s}^H\widehat{\mathbf{C}}^{-1}\mathbf{s}\mathbf{s}^H\widehat{\mathbf{C}}^{-1}\mathbf{s}} \end{aligned} \tag{11}$$

where IF_{STAP} is the IF of the STAP processor being looked at. The maximum attainable value of IF_{loss} is unity, indicating that the processor performance is not degraded by clutter. Optimum performance is obtained with $\mathbf{w} = \mathbf{w_{opt}}$. In practice, the processor performance is degraded by estimation losses and the bistatic clutter range-dependency problem.

3 Prediction of Inverse Covariance Matrix (PICM) Sequences

Linear prediction is widely used in coding and communications applications. In [8], a new technique was proposed to use linear prediction theory to obtain an estimate of the range-dependent inverse clutter covariance matrix. It must be noted that the linear prediction techniques are applied to the inverse covariance matrix sequence, as shown in Fig. 2, and *not* the uncorrelated data snapshots.

The test range gate is hereby referred to as the r^{th} range gate. Denote each lag λ of the stacked inverse covariance matrix of the k^{th} training range gate as $\bar{\mathbf{C}}_k^{-1}(\lambda)$, the two-taps linear prediction, shown in Fig. 3, is given as:

$$\widehat{\bar{\mathbf{C}}}_k^{-1}(\lambda) = \alpha_{+1}(\lambda)\bar{\mathbf{C}}_{k+1}^{-1}(\lambda) + \alpha_{-1}(\lambda)\bar{\mathbf{C}}_{k-1}^{-1}(\lambda) \tag{12}$$

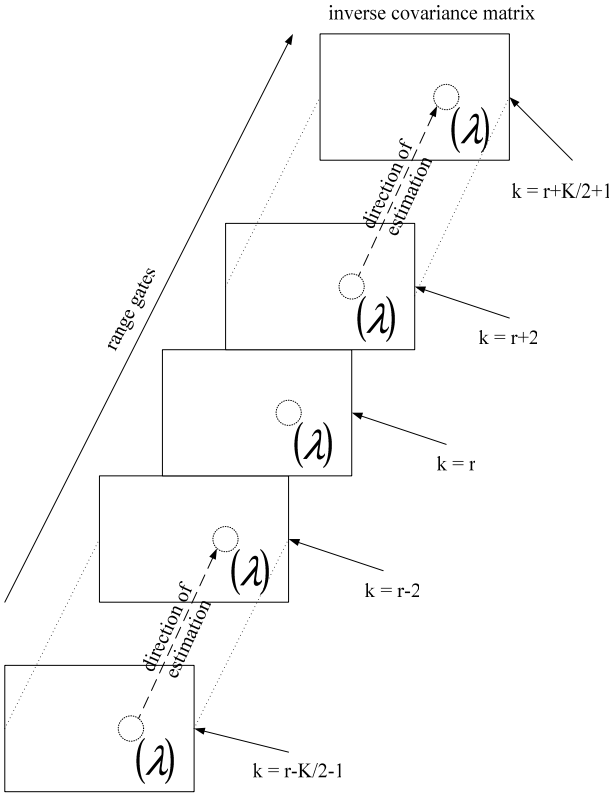

Figure 2: Inverse covariance matrix sequence

$$\text{for} \quad k = r - \frac{K}{2} - 1, \ldots \ , r-2, r+2, \ldots, r + \frac{K}{2} + 1$$

$$\lambda = 1, 2, \ldots, (NM)^2$$

where $\hat{\bar{\mathbf{C}}}_k^{-1}$, $\bar{\mathbf{C}}_{k+1}^{-1}$ and $\bar{\mathbf{C}}_{k-1}^{-1}$ are the estimated stacked inverse covariance matrix for the k^{th} range gate and the stacked inverse covariance matrices for the $(k+1)^{th}$ and $(k-1)^{th}$ range gate respectively; α_{+1} and α_{-1} are the prediction weights and K is the total number of training range gates required. The $(r-1)^{th}$ and $(r+1)^{th}$ range gates (guard-gates) are normally excluded to prevent any target signal attenuation.

The inverse covariance matrices and prediction weights are a function of λ. The solution of $\alpha_{+1}(\lambda)$ and $\alpha_{-1}(\lambda)$ are given by the minimisation of the minimum mean squared error (MMSE) of the true and estimated stacked inverse covariance matrix

$$\sum_k \left| \bar{\mathbf{C}}_k^{-1}(\lambda) - \hat{\bar{\mathbf{C}}}_k^{-1}(\lambda) \right|^2 \tag{13}$$

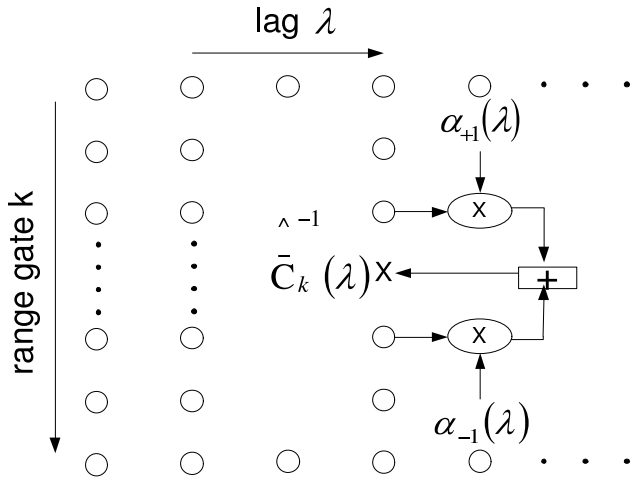

Figure 3: Two-taps linear prediction

The series of resulting linear simultaneous equations from equation (12) can be re-arranged into a matrix form and a solution can be obtained by solving the system of linear simultaneous equations. From equation (12):

$$\widehat{\mathbf{C}}_k^{-1}(\lambda) = \mathbf{\Theta}\ \mathbf{\Omega} \tag{14}$$

where

$$\mathbf{\Theta} = \left[\bar{\mathbf{C}}_{k+1}^{-1}(\lambda)\ \ \bar{\mathbf{C}}_{k-1}^{-1}(\lambda)\right] \tag{15}$$

$$\mathbf{\Omega} = \left[\alpha_{+1}(\lambda)\ \ \alpha_{-1}(\lambda)\right]^T \tag{16}$$

The linear prediction is done on each λ for the inverse covariance matrix. The prediction weights $\widehat{\mathbf{\Omega}}$ for each λ of the inverse covariance matrix is obtained from the surrounding training data (regressors). From equation (14),

$$\mathbf{\Theta}^H \widehat{\mathbf{C}}_k^{-1}(\lambda) = \mathbf{\Theta}^H \mathbf{\Theta}\ \mathbf{\Omega} \tag{17}$$

$$\left(\mathbf{\Theta}^H \mathbf{\Theta}\right)^{-1} \mathbf{\Theta}^H \widehat{\mathbf{C}}_k^{-1}(\lambda) = \widehat{\mathbf{\Omega}} \tag{18}$$

There are more equations than unknowns required, thus equation (13) is simply a standard least squares (LS) problem with a known solution. Technically, this is smoothing [20] rather than forward or backward prediction. Simple forward or backward prediction will not give as accurate an estimate as smoothing does. It should also be noted here that a LS fit allows for the smoothing out of noise in inverse covariance matrices estimates.

In practice, the exact covariance matrix $\mathbf{C}$ is unknown and needs to be estimated from the data. For each k^{th} range gate, a reasonable estimate of the covariance matrix $\mathbf{C}_k^{-1}$ is obtained by using $2NM$ snapshots [21]. For conventional STAP algorithms, the computational complexity of computing the STAP weights is given as $O(NM)^3$ for inverting the estimated covariance matrix $\widehat{\mathbf{C}}$.

For PICM, there is the additional computational complexity of $O(NM)^3$ for computing the inverse covariance matrix of each range gate. This procedure is repeated for all the range gates in the training sequence. The computational complexity of the two-taps linear prediction, of order $O(4)^3$ involves the inversion of a (2×2) matrix and the prediction weights are computed for each of the $(NM)^2$ components of $\bar{\mathbf{C}}_k^{-1}(\lambda)$. Thus the total computational complexity of the PICM technique is $\left(K \times O(NM)^3 + (NM)^2 \times O(4)^3\right)$.

By exploiting the Hermitian property of the inverse covariance matrix, the linear prediction only needs to be carried out for the upper triangular portion. The linear prediction weights for the lower triangular portion of the inverse covariance matrix are simply the complex conjugate of its upper triangular portion's counterparts.

By implementing PICM, there is no need to exploit the Toeplitz-block-Toeplitz structure of the theoretical covariance matrix, thus eliminating the requirement for a uniform linear array (ULA), as required in [22] and [6]. Without any such restrictions, PICM can be applied to arrays of arbitrary configuration.

3.1 Longer and multi-dimensional linear prediction

Longer prediction can be done by using more training range gates to obtain a better estimate of the inverse covariance matrix. The four-taps linear prediction is then simply:

$$\begin{aligned}\widehat{\bar{\mathbf{C}}}_k^{-1}(\lambda) = \alpha_{+2}(\lambda)\bar{\mathbf{C}}_{k+2}^{-1}(\lambda) + \alpha_{+1}(\lambda)\bar{\mathbf{C}}_{k+1}^{-1}(\lambda) \\ +\alpha_{-1}(\lambda)\bar{\mathbf{C}}_{k-1}^{-1}(\lambda) + \alpha_{-2}(\lambda)\bar{\mathbf{C}}_{k-2}^{-1}(\lambda)\end{aligned} \tag{19}$$

In this case, 12 training range gates are required to obtain 8 equations [21] to make a reasonable estimate of 4 parameters, α_{+2}, α_{+1}, α_{-1} and α_{-2}. The extension to even longer predictors is straight-forward.

For multi-dimensional linear prediction, more prediction coefficients are obtained from within the same range gate. From (12), the two-taps, two-dimensional linear prediction is:

$$\begin{aligned}\widehat{\bar{\mathbf{C}}}_k^{-1}(\lambda) = \alpha_{+1}(\lambda)\bar{\mathbf{C}}_{k+1}^{-1}(\lambda) + \alpha_{-1}(\lambda)\bar{\mathbf{C}}_{k-1}^{-1}(\lambda) \\ +\beta_{+1}(\lambda+1)\bar{\mathbf{C}}_{k+1}^{-1}(\lambda+1) + \beta_{-1}(\lambda+1)\bar{\mathbf{C}}_{k-1}^{-1}(\lambda+1) \\ +\gamma_{+1}(\lambda-1)\bar{\mathbf{C}}_{k+1}^{-1}(\lambda-1) + \gamma_{-1}(\lambda-1)\bar{\mathbf{C}}_{k-1}^{-1}(\lambda-1)\end{aligned} \tag{20}$$

Increasing the number of prediction weights results in a corresponding increase in the number of training range gates required. However, there is a trade-off between more accurate estimates and computational complexity. Up to a certain limit, the benefits of using more prediction coefficients will be negated by the bistatic clutter range dependency problem.

3.2 Mitigation against aliasing

A pulse Doppler radar can be ambiguous in either range or Doppler frequency [1]. For conventional pulse Doppler radar, range ambiguities occur because of the transmission of repetitive pulses. The additional ambiguous clutter returns are also known as multiple-time-around clutter. Such clutter echoes may result in additional clutter notches, thus affecting the detection of targets.

Doppler ambiguites occur when the Nyquist sampling criterion is not followed. From (9), we can see that the Doppler frequency is associated with the relative velocities of the target to the receiver and transmitter. These velocities are functions of the azimuth and elevation angles of the receiver and transmitter to the target. Thus any ambiguity in azimuth causes ambiguity in Doppler frequency and vice versa.

PICM is able to deal with the additional clutter notches resulting from range and Doppler ambiguities. In such situations, conventional compensation techniques such as angle-Doppler compensation (ADC) may not perform as well.

3.3 Performing detection in other range gates

From (12), it can be seen that the two-taps linear prediction is carried out for the whole sequence of training range gates used. Thus the prediction weights are obtained for the whole training sequence, except for the first and last range gate. As a result, the prediction weights can also be used to perform detection in other range gates. This does not lead to any additional computational costs in having to re-compute the prediction weights all over again.

4 Single Data-Set Detectors (SDSD)/ In-the-Gate Processing

Conventional STAP algorithms require the training data to be target-free and statistically homogeneous with respect to the range gate under test. However, these conditions are often not satisfied, which leads to a degradation in the detection performance. In order to alleviate the high target density or heterogeneity problem and therefore improve the detection performance, research had been carried out in knowledge-based (KB-STAP) techniques, [9], [10] and [11], which incorporate apriori knowledge into the processor.

An alternative approach that forgoes the training data requirement and operates solely on the test data set was proposed in [12]. The new SDSD, namely the generalised maximum likelihood detector (GMLED) and the maximum likelihood detector (MLED) are high-resolution CFAR spectral estimators, thus enabling them to be used for detection. They carry out a joint maximum likelihood estimate of the signal and interference subspaces.

The basic concept behind this single data set detection or in-the-gate processing is illustrated in Fig. 4 for the single channel case. In estimating the spectral content of a signal at a particular frequency f_o, it is common practice to construct some form of band-pass filter $H_o(f)$ centered at that frequency and measure the power in the signal y at the output of the filter. The frequency response of the filter might be fixed as in the fast Fourier transform (FFT) or adaptive as in the case of minimum variance spectral estimation. In common with maximum likelihood spectral estimation (APES) [23], SDSD adds a second path (lower branch in Fig. 4(a)), which consists of a unit-amplitude complex oscillator at the frequency of interest f_o and a complex gain term a.

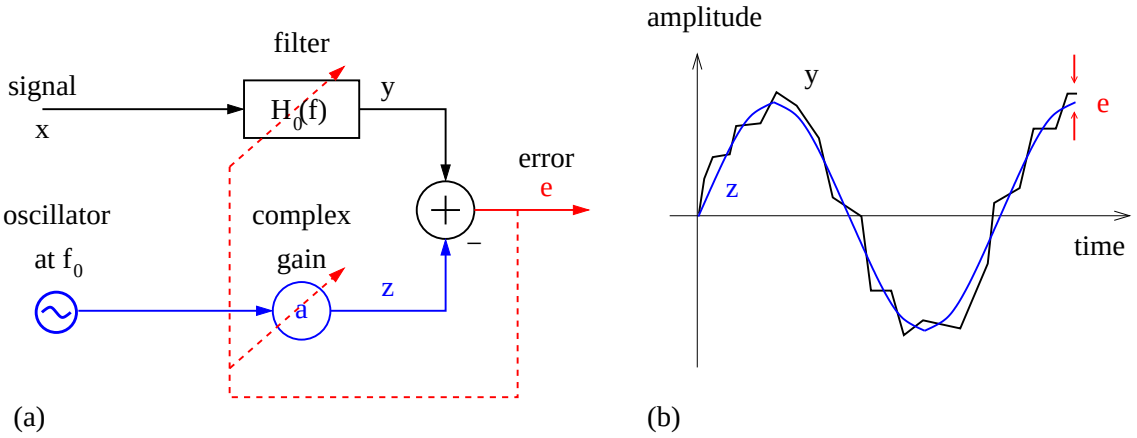

Figure 4: Single data set detecion: (a) block diagram; (b) signals.

The filter frequency response $H_o(f)$ and the complex gain a are chosen to minimize the average power in the error signal e, subject to a linear constraint that the gain of the filter at f_o is unity, that is, $|H_o(f_o)| = 1$. With this choice, an estimate of the signal or target power is given by $|a|^2$ and the clutter or coloured noise power at the output of the filter is given by the average power of the error e. The ratio of the two power estimates provides a CFAR test statistic and hence there is no requirement for secondary data.

For a full-dimension STAP system, the computational cost of directly inverting a $(NM \times NM)$ dimensional matrix, which is of order $O(NM)^3$, is high. Reducing the dimension of the STAP processor is often used to reduce the computational costs of the processor. Reduced dimension

(RD) STAP techniques require less training data and hence minimise the widening of the clutter ridge. One of them is the joint domain localized (JDL) processor proposed by WANG and CAI in [24]. It works by transforming the space-time data into the angle-Doppler domain by using a two-dimensional discrete Fourier Transformation (DFT). The angle-Doppler data is then grouped into regions called localized processing regions (LPRs) and adaptive processing is restricted to the $(N_a \times M_d)$ LPRs as shown in Fig. 5.

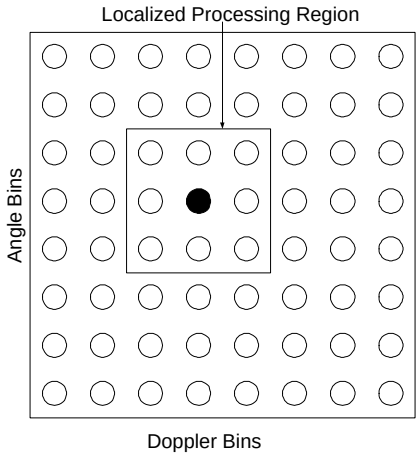

Figure 5: LPR of JDL processor.

The detection statistics of the reduced-dimension SDSD are:

$$Y_{JDL-GMLED} = \frac{|\widetilde{\mathbf{s}}^H \widetilde{\mathbf{Q}}^{-1} \widetilde{\mathbf{g}}|^2}{\widetilde{\mathbf{s}}^H \widetilde{\mathbf{Q}}^{-1} \widetilde{\mathbf{s}} \left(1 + \widetilde{\mathbf{g}}^H \widetilde{\mathbf{Q}}^{-1} \widetilde{\mathbf{g}}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{21}$$

$$Y_{JDL-MLED} = \frac{|\widetilde{\mathbf{s}}^H \widetilde{\mathbf{Q}}^{-1} \widetilde{\mathbf{g}}|^2}{\widetilde{\mathbf{s}}^H \widetilde{\mathbf{Q}}^{-1} \widetilde{\mathbf{s}}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{22}$$

$$\begin{aligned} \widetilde{\mathbf{g}} &= \frac{1}{|\mathbf{t}|} \widetilde{\mathbf{X}} \mathbf{t}^* \\ &= \frac{1}{|\mathbf{t}|} \sum_{k=1}^{K_T} \widetilde{\mathbf{x}}_k \mathbf{t}^*(k) \\ &= \begin{bmatrix} \widetilde{g}'_A \\ \widetilde{\mathbf{g}}'_B \end{bmatrix} \end{aligned} \tag{23}$$

$$\begin{aligned}\widetilde{\mathbf{Q}} &= \left(\widetilde{\mathbf{X}} - \frac{1}{|\mathbf{t}|}\widetilde{\mathbf{g}}\mathbf{t}^T\right)\left(\widetilde{\mathbf{X}} - \frac{1}{|\mathbf{t}|}\widetilde{\mathbf{g}}\mathbf{t}^T\right)^H \\ &= \sum_{k=1}^{K_T}\widetilde{\mathbf{x}}_k\widetilde{\mathbf{x}}_k^H - \widetilde{\mathbf{g}}\widetilde{\mathbf{g}}^H \\ &= \begin{bmatrix} \widetilde{q}_{AA} & \widetilde{\mathbf{q}}_{BA}^H \\ \widetilde{\mathbf{q}}_{BA} & \widetilde{\mathbf{Q}}_{BB} \end{bmatrix} \end{aligned} \tag{24}$$

where $\widetilde{\mathbf{g}}$ is the angle-Doppler 'sample mean' vector, obtained by coherently combining the K_T test data snapshots along the space-time vector $\mathbf{s}$. The centered covariance matrix $\widetilde{\mathbf{Q}}$ is the product of the original test data snapshot matrix with the sample mean subtracted $\left(\widetilde{\mathbf{X}} - \frac{1}{|\mathbf{t}|}\widetilde{\mathbf{g}}\mathbf{t}^T\right)$ and the Hermitian counterpart of this quantity. Hence $\widetilde{\mathbf{g}} \sim \mathcal{CN}(\alpha\widetilde{\mathbf{s}}, \widetilde{\mathbf{C}})$ and $\widetilde{\mathbf{Q}}$ has a complex Wishart distribution with scale matrix $\widetilde{\mathbf{C}}$, that is, $\widetilde{\mathbf{Q}} \sim \mathcal{C}W_{N_aM_d}(\widetilde{\mathbf{C}}, K_T - 1)$. The $\widetilde{@}$ refers to the quantities transformed into the angle-Doppler domain.

The conventional STAP approaches can be referred to as two-data-sets detection (TDSD), since they require one training data set to estimate the interference covariance matrix and a separate test data set. Two such detection statistics are the generalised likelihood ratio test (GLRT) [16] and the adaptive matched filter (AMF) [17]. Consequently the reduced-dimension counterparts (RD-TDSD) of these detection statistics are the JDL-GLRT [24] and the JDL-AMF.

$$Y_{JDL-GLRT} = \frac{|\widetilde{\mathbf{s}}^H\widetilde{\widehat{\mathbf{C}}}^{-1}\widetilde{\mathbf{g}}|^2}{\widetilde{\mathbf{s}}^H\widetilde{\widehat{\mathbf{C}}}^{-1}\widetilde{\mathbf{s}}\left(1 + \frac{1}{K_t}\widetilde{\mathbf{g}}^H\widetilde{\widehat{\mathbf{C}}}^{-1}\widetilde{\mathbf{g}}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{25}$$

$$Y_{JDL-AMF} = \frac{|\widetilde{\mathbf{s}}^H\widetilde{\widehat{\mathbf{C}}}^{-1}\widetilde{\mathbf{g}}|^2}{\widetilde{\mathbf{s}}^H\widetilde{\widehat{\mathbf{C}}}^{-1}\widetilde{\mathbf{s}}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{26}$$

From equations (21) and (25), it can be seen that the JDL-GMLED has a similar detection statistic as the JDL-GLRT except that $\widetilde{\mathbf{Q}}^{-1}$ (from the test data set) is used for the former and $\widetilde{\widehat{\mathbf{C}}}^{-1}$ (from the training data set) is used for the latter. It can also be shown that they have the same probabilities of false alarm and detection, with one less complex degree of freedom (DoF) for the JDL-GMLED. This loss of one complex DoF, which is used in the estimation of the signal subspace of dimension one, is a result of the dual estimation of the signal and noise subspaces from the same data set.

4.1 Results Summary

To complete the discussion, the probability of false alarm P_{fa} and probability of detection P_d expressions for the various detection tests presented in the paper are summarised. *For the*

purpose of analysis. it is assumed that there exists a set of independent training data $\{\mathbf{z}_k\}_{k=1}^{K_t}$, *that is homogeneous with the test data. In addition, it is also assumed that although the test data snapshots are obtained using a sliding window, the columns of* $\mathbf{X}$ *from equation (1) are statistically independent.* The effects of using non-independent snapshots will be presented in the simulation results section and it must be stressed that the effects will be similar for *both* the SDSD and TDSD algorithms.

Table 1: Summary of the P_{fa} and P_d expressions

Algorithm	P_{fa}	P_d	L	τ
JDL-AMF	eq (27)	eq (29)	$K_t - N_a M_d + 1$	$\frac{1}{K_t}\widetilde{\eta}\gamma$
JDL-GLRT	eq (27)	eq (29)	$K_t - N_a M_d + 1$	$\frac{\gamma}{K_t - \gamma}$
JDL-MLED	eq (27)	eq (29)	$K_T - N_a M_d$	$\widetilde{\eta}\gamma$
JDL-GMLED	eq (27)	eq (29)	$K_T - N_a M_d$	$\frac{\gamma}{1-\gamma}$

The results summary, in Table 1, are presented in a consistent form as in [12] to provide easier comparison. The four reduced-dimension detectors share the same expressions for the P_{fa} and P_d with each detector having its specific value of the degrees of freedom L, the non-centrality parameter $\widetilde{\lambda}$ and the effective threshold τ (where τ is expressed in terms of the test threshold γ from equation (5)).

The P_{fa} of the reduced-dimension detection tests is

$$P_{fa}(\gamma) = \int_0^1 (1+\tau)^{-L} f_{\beta,L+1,N_a M_d - 1}(\widetilde{\eta}) d\widetilde{\eta} \tag{27}$$

where $f_{\beta,L+1,N_a M_d-1}(\widetilde{\eta})$ is the type I beta distribution and

$$\widetilde{\eta} = \left(1 + \widetilde{\mathbf{g}}_B'^H \widetilde{\mathbf{Q}}_{BB}'^{-1} \widetilde{\mathbf{g}}_B'\right)^{-1}. \tag{28}$$

Finally, the general expression of the P_d is

$$P_d = \int_0^1 h(\widetilde{\eta}) f_{\beta,L+1,N_a M_d - 1}(\widetilde{\eta}) d\widetilde{\eta} \tag{29}$$

where

$$h(\widetilde{\eta}) = 1 - (1+\tau)^{-L} \sum_{l=1}^{L} \binom{L}{l} \tau^l G_l\left(\frac{\widetilde{\lambda}}{1+\tau}\right) \tag{30}$$

and

$$G_l(i) = e^{-i} \sum_{n=0}^{l-1} \frac{i^n}{n!} \tag{31}$$

with $\widetilde{\lambda} = K_T N_a M_d \widetilde{\rho}$ and $\widetilde{\rho} = \frac{1}{N_a M_d K_T} |\alpha \mathbf{s}_t|^2 \widetilde{\mathbf{s}}^H \widetilde{\mathbf{C}}^{-1} \widetilde{\mathbf{s}}$.

5 Simulation Results

5.1 IF_{loss} performance

This section will show the simulation results, with respect to the IF_{loss} performance, of the PICM algorithm against conventional STAP algorithms. The radar parameters used are shown in Table 2 and the backlobe of the clutter scatterers' response was ignored.

Table 2: Radar Parameters

number of antenna elements	N	8
number of pulses delay	M	24
pulse repetition frequency	PRF	20 kHz
operating frequency	f_o	10 GHz
array geometry		bistatic forward-looking
receiver & transmitter height	H_R, H_T	1000 m
receiver & transmitter velocity	v_R, v_T	90 m/s
transmitter flight angle	δ_t	90°
receiver look angle	φ_r	45°
receiver maximum sensor pattern direction	φ_0	45°
baseline separation (along flight direction)		2000 m
clutter-to-noise ratio (CNR)		30dB

For all the algorithms shown in this sub-section, $2NM$ snapshots were used to obtain the covariance matrix $\mathbf{C}_k$ for each k^{th} range gate and a total of 16 training range gates were used to compute the STAP filter weights. The simulation results were obtained from a Monte Carlo simulation comprising of 1000 runs.

Fig. 6 shows the IF_{loss} plot of the PICM technique with two-taps linear prediction (dashed line) and the PICM technique with four-taps linear prediction (solid line). The dash-dotted line shows the conventional processor obtained from the straight averaging (SA) of the covariance matrices $\mathbf{C}_k$, which is also known as the sample matrix inversion (SMI) algorithm [15]. Lastly, the dash-dotted line with crosses shows the coventional processor obtained from the SA of the inverse covariance matrices $\mathbf{C}_k^{-1}$. The last processor was included as a reference since the linear prediction is done on the inverse covariance matrices.

It can be observed that both PICM techniques give an improvement in IF_{loss} performance over the latter two conventional processors. An important point to note is the narrower clutter notch

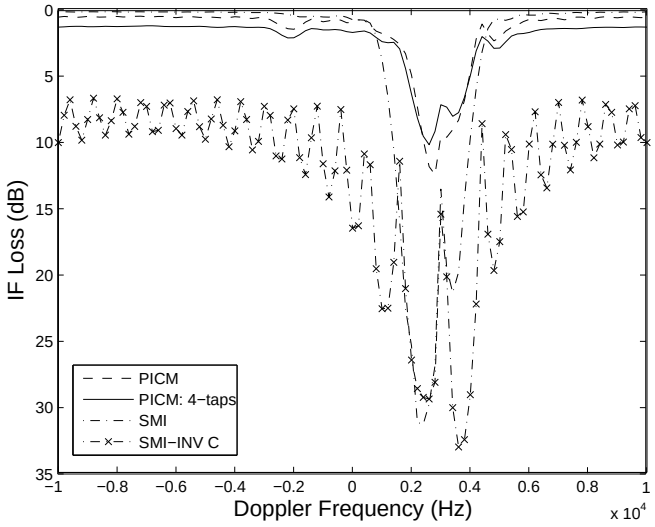

Figure 6: IF_{loss} plot: PICM (2-taps), PICM (4-taps), SMI and SMI-(INV C)

provided by both PICM techniques, hence enhancing the capability of detecting slow moving targets. By comparing the PICM with two-taps linear prediction against PICM with four-taps linear prediction, we see that the latter performs better within the clutter notch. At the critical clutter notch frequency, there is an improvement of 2.5dB.

For the rest of the simulation results, we will show PICM with two-taps linear prediction against other STAP algorithms. In Figs. 7 and 8, we show the IF_{loss} performance with range and Doppler ambiguities respectively. For the set of results with Doppler ambiguities, we change the receiver's look angle $\varphi_r = \varphi_0 = 0°$ and include both the frontlobe and backlobe of the clutter scatterers' response to illustrate the robustness of the proposed technique. The dashed line now refers to the angle-Doppler compensation (ADC) algorithm [3]. From both figures, we can see that PICM outperforms the two reference algorithms and the results highlight the benefits of using PICM in the presence of aliasing effects.

The results differ for the range and Doppler ambiguities because the effects on the clutter spectrum due to each ambiguity are different. When range ambiguity is present, the clutter echoes of a certain range gate includes the clutter contributions from other range gates. Thus the different mainlobe in ambiguous range gates move in different directions and do not coincide, even with compensation to mitigate for the bistatic clutter Doppler range dependency. Whereas for Doppler ambiguity, the clutter notch is repeated every PRF. In this case, the different mainlobes move in the same direction (constant shift).

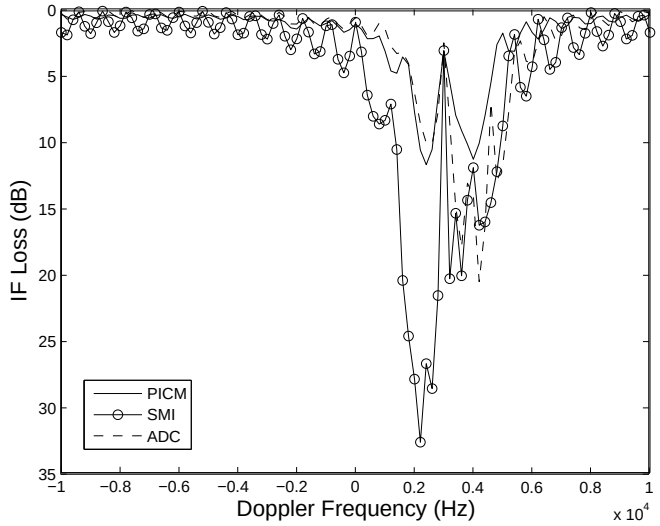

Figure 7: IF_{loss} plot with range ambiguities: PICM, SMI and ADC

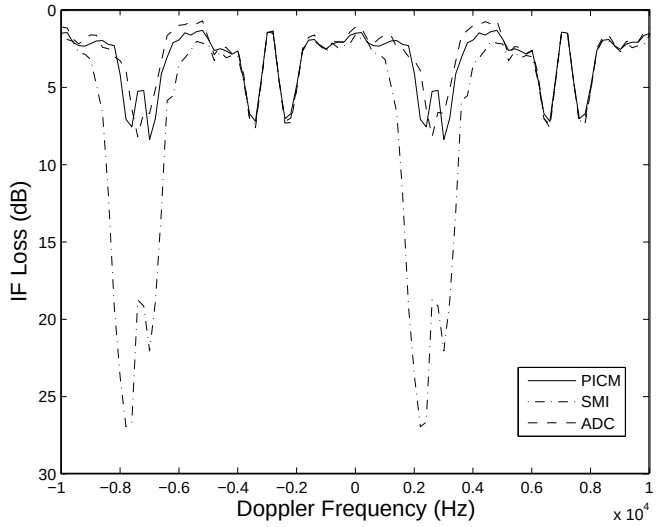

Figure 8: IF_{loss} plot with Doppler ambiguities: PICM, SMI and ADC

Specifically, we note that PICM gives a significant improvement over ADC within the clutter notches. The latter compensation method was meant to compensate with the peak angle and Doppler frequencies and thus it will perform well within the mainlobe. However, aliasing results in additional clutter notches due to the overlapping of clutter spectrums and thus the coventional compensation methods will not perform as well when aliasing occurs.

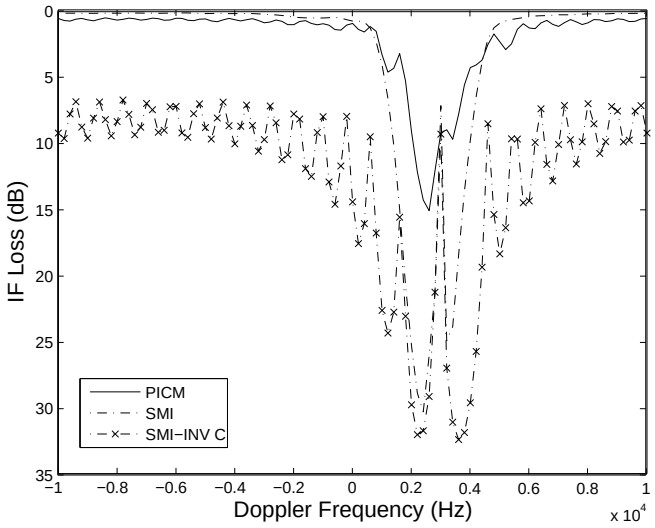

Figure 9: IF_{loss} plot of a secondary range gate

Fig. 9 shows the IF_{loss} plot of one of the secondary training range gate used in the linear prediction sequence. The superior performance of PICM over the other two conventional STAP algorithms in the secondary range gate shows the usefulness of the technique in performing clutter suppression. No additional computational costs are encountered since the prediction weights for all the range gates in the training sequence have already been determined. However, this is not the case for the other reference algorithms. Whenever the range gate under test is changed, the whole process of getting the estimated covariance matrix, used to obtain the filter weights, from neighbouring range gates, has to be repeated. Any compensation done to mitigate for the bistatic clutter Doppler range dependency also has to be re-computed, since the reference test range gate has been changed.

5.2 Receiver operating curves

Another performance metric is to compute the receiver operating curves (ROCs) of the STAP processors to evaluate the performance of the various training algorithms. The ROCs show the

probability of target detection as a function of the signal-to-clutter plus noise ratio (SCNR).

The JDL-GMLED, JDL-MLED, along with the JDL-GLRT and JDL-AMF were simulated. The radar parameters used in this section are shown in Table 2 and the following figures show the theoretical (labelled TH) and simulated P_d curves vs SCNR for $P_{fa} = 10^{-2}$. The results presented in this section were obtained from Monte Carlo simulations comprising of 10000 runs each. The target signal was injected at a frequency of 3333Hz.

In order to make a fair comparison, the number of test data snapshots K_T was taken to be the same for all the detection algorithms, since it is a parameter that affects the test data which is required for all the algorithms. Different simulation results are shown to illustrate the effects of assuming independent and identically distributed (iid) data snapshots. For the RD-TDSD approaches (JDL-AMF and JDL-GLRT), the number of training data snapshots K_t was set to be the same as K_T used for the RD-SDSD approaches.

Two sets of results are shown; the first one uses $K_T = 2NM$ test data snapshots and $K_t = 2NM$ training data snapshots while the second set of results uses $K_T = 2N_aM_d$ and $K_t = 2N_aM_d$. For the first set, the results are shown in Figs. 10-13 respectively, when iid and non-iid data snapshots are used. Similarly, for the second set, the results are shown in Figs 14-17 respectively.

For all the figures, the dashed lines with diamonds, asterisks, squares and pluses refer to the theoretical results for the JDL-MLED, JDL-GMLED, JDL-AMF and JDL-GLRT respectively. These theoretical results were obtained using equations (27) and (29), with the parameters shown in Table 1. The same theoretical results were used in both the iid and non-iid data snapshots scenarios. The simulation results for the reduced-dimension algorithms are shown by the solid lines with the same markers. In addition, the solid lines with right and left-pointing triangles refer to the JDL-AMF and JDL-GLRT for the filter-banks based JDL processor with angle and separate doppler compensation (FB-JDL-ASDC) [4]; the solid line with crosses and circles refer to the JDL-AMF and JDL-GLRT for angle-Doppler compensation (ADC) [3]. For the ADC technique, the angle-Doppler compensation was performed in the space-time domain followed by a JDL processor to obtain the STAP filter weights.

When iid data snapshots are assumed, as in Figs. 10-11 and 14-15, the bunch of lines on the left-most side includes the theoretical plots and the simulation results for the JDL-MLED and JDL-GMLED. The fact that the RD-SDSD approaches (JDL-GMLED and JDL-MLED) need to estimate both the signal and noise subspace from the same set of data results in a loss of 1 complex DoF in the noise subspace estimate. As K_T and K_t increase, the loss in the DoF becomes less significant and the various JDL-STAP P_d curves converge due to the beta-distribution,

Moreover, for large values of K_T and K_t, the additional terms in the denominator of equations (21) and (25) will tend to unity, thus making them similar to equations (22) and (26) respectively. As a result, we see that, in Figs 10 and 11, the P_d curves for the generalised and non-generalised detectors, between the JDL-MLED and JDL-GMLED or the JDL-AMF and JDL-GLRT are almost identical.

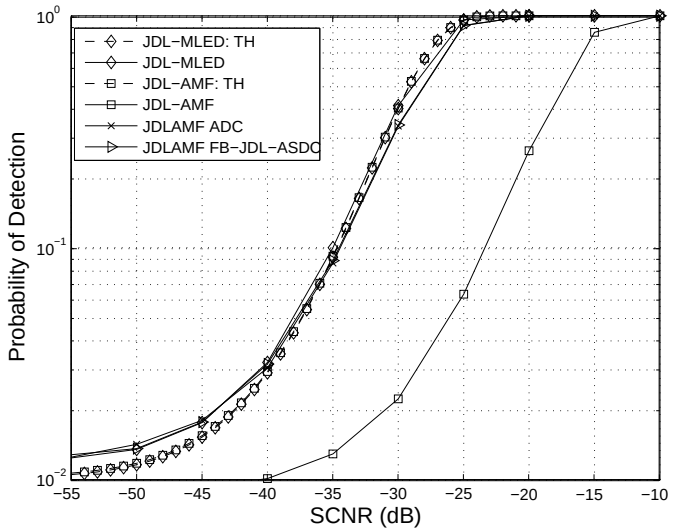

Figure 10: Probability of detection vs $SCNR$ of generalised detectors for $K_T = K_t = 2NM$, assuming iid data snapshots.

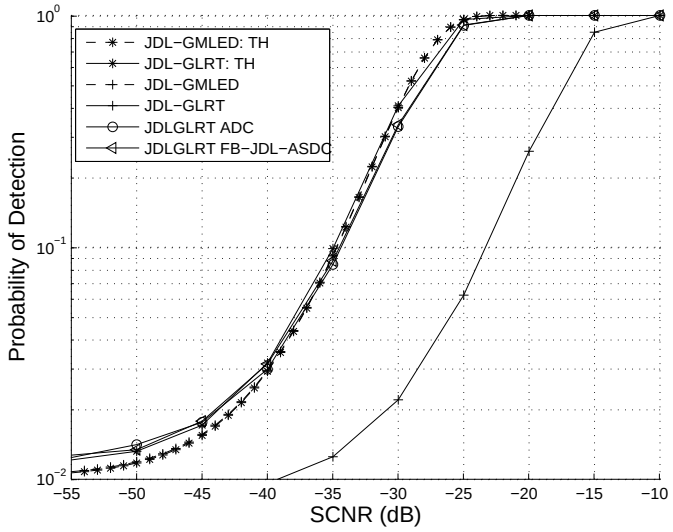

Figure 11: Probability of detection vs $SCNR$ of non-generalised detectors for $K_T = K_t = 2NM$, assuming iid data snapshots.

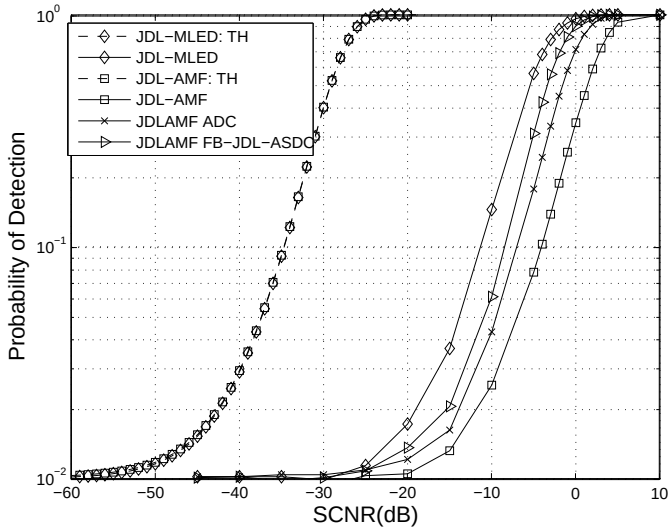

Figure 12: Probability of detection vs $SCNR$ of generalised detectors for $K_T = K_t = 2NM$.

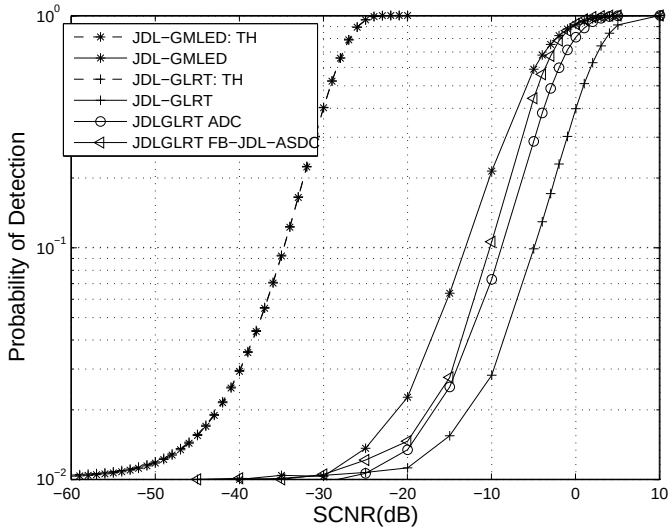

Figure 13: Probability of detection vs $SCNR$ of non-generalised detectors for $K_T = K_t = 2NM$.

For the case when $K_T = K_t = 2N_aM_d$, as in Figs. 14 and 15, it can be observed that when the values of K_T and K_t are decreased, the theoretical results can be distinguised, due to the additional terms for the JDL-GMLED and JDL-GLRT expressions. For all the figures, the order of the theoretical results for the algorithms, from left to right, is as follows: JDL-GLRT, JDL-GMLED, JDL-AMF and JDL-MLED. This observation follows closely to that in [13].

When the iid assumption of the data snapshots is not met, it can be observed that there is a general degradation in performance results for *all* the detectors. This can be seen by comparing Figs. 10-11 with Figs. 12-13 or Figs. 14-15 with Figs. 16-17. However, the SDSD algorithms (JDL-GMLED and JDL-MLED) still perform better than the TDSD algorithms (JDL-GLRT and JDL-AMF). It should be emphasised that the degradation is associated with *both* the SDSD and TDSD algorithms.

The first and second set of figures are shown to illustrate the effects of K_T and K_t on the detection performance although the second set of results is more applicable for JDL-STAP since it uses $K_T = 2N_aM_d$ and $K_t = 2N_aM_d$ [21]. Increasing K_t affects only the JDL-AMF and JDL-GLRT since this translates into an improvement in the estimation of the noise covariance matrix and hence the noise subspace. Increasing K_T on the other hand, results in an improvement in the signal subspace estimation, that is, improved SCNR. This is exhibited in the change in the non-centrality parameter of all algorithms.

Therefore, K_T also affects the noise subspace estimation and consequently the algorithms' performances through τ and L. Hence, there is an improvement in detection performance between Figs. 10-11 and 14-15 or between Figs. 12-13 and 16-17, at the expense of higher computational complexity.

When iid data snapshots are used, for the JDL-GMLED and JDL-MLED, it can be seen that the theoretical and simulation results agree well for the two different sets of parameters, even for a complicated bistatic clutter model. However, this is not the case for the RD-TDSD algorithms as we note the severe deterioration in P_d performance for the JDL-GLRT and JDL-AMF from their theoretical results. This shows the degradation of conventional clutter suppression algorithms for bistatic radar applications, an observation that has been widely reported in the literature.

Although there is an improvement in detection performance when compensation is performed to mitigate for the bistatic clutter Doppler range dependency problem, the performance is still worse off than the in-the-gate detectors. This again highlights the benefits of using the SDSD algorithms over conventional STAP algorithms in heterogeneous clutter environments. It can also be observed that the gain in P_d performance, between the SDSD algorithms and the compensation algorithms, is greater for the second set of simulation results. This provides an additional benefit of implementing the SDSD algorithms in heterogeneous environments since there are less iid data snapshots for training purposes.

In summary, for the bistatic ground moving target indication scenario considered, the SDSD algorithms give the best performance whereas the conventional STAP approaches suffer a degradation due to the bistatic clutter range dependency problem.

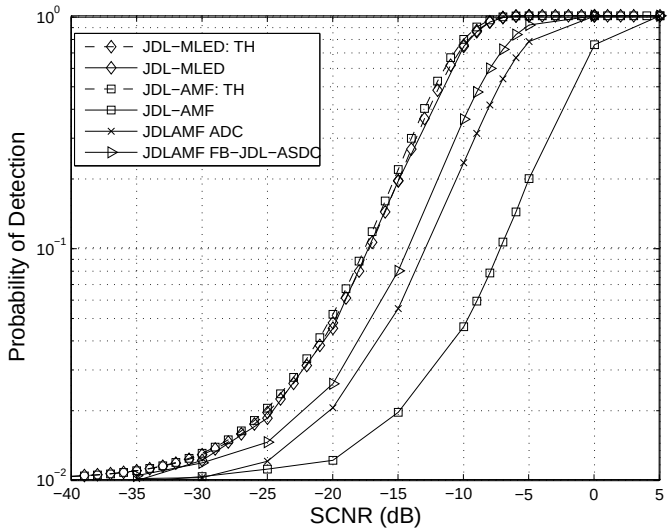

Figure 14: Probability of detection vs $SCNR$ of generalised detectors for $K_T = K_t = 2N_aM_d$, assuming iid data snapshots.

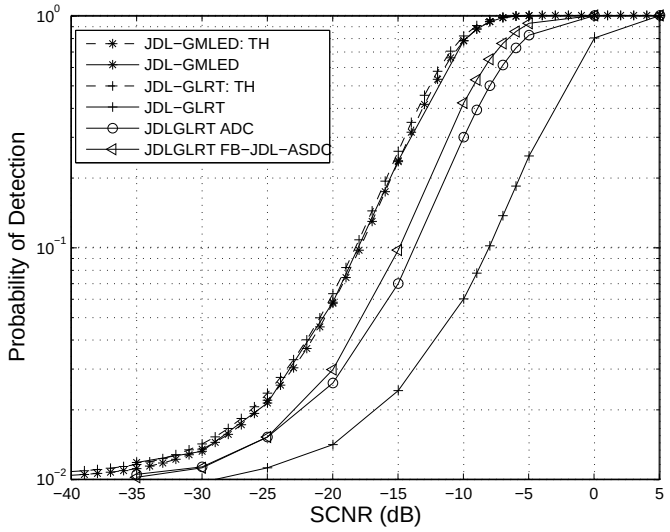

Figure 15: Probability of detection vs $SCNR$ of non-generalised detectors for $K_T = K_t = 2N_aM_d$, assuming iid data snapshots.

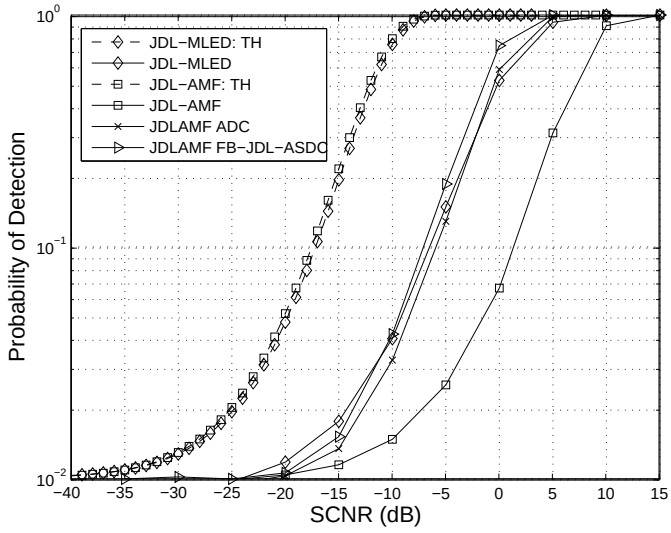

Figure 16: Probability of detection vs $SCNR$ of generalised detectors for $K_T = K_t = 2N_aM_d$.

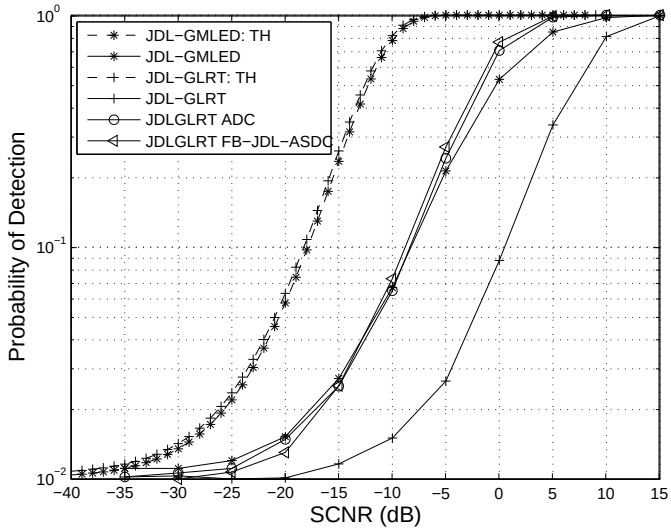

Figure 17: Probability of detection vs $SCNR$ of non-generalised detectors for $K_T = K_t = 2N_aM_d$.

6 Conclusion

In this paper, we provided the analysis of using linear prediction theory to obtain an estimate of the inverse covariance function. Simulation results show the improvement in performance over conventional STAP techniques and using longer prediction results in an improved performance. The technique is able to mitigate against aliasing and also provide clutter suppression in the secondary range gates without additional computational costs. We have also discussed a novel in-the-gate processing approach, which forgoes the training data and operates solely on the test data. This single data set detection applies particularly to heterogeneous environments where the training data are often not target-free or homogenous with the test data. Simulation results for a bistatic radar application show the improvement in detection performance of this approach over conventional STAP approaches.

7 Acknowledgment

The authors would like to thank the support of BAE Systems and SELEX Sensors and Airborne Systems Ltd for this work.

References

[1] R. Klemm, *Principles of space-time adaptive processing.* IEE Radar, Sonar, Navigation and Avionics Series, second edition, 2002.

[2] C. Lim, E. Aboutanios, and B. Mulgrew, "Modified JDL with Doppler compensation for airborne bistatic radar," *Proceedings of IEEE Radar Conference*, pp. 854–858, May 2005.

[3] B. Himed, Y. Zhang, and A. Haijari, "STAP with angle-Doppler compensation for bistatic airborne radars," *Proceedings of IEEE Radar Conference*, pp. 311–317, 2002.

[4] C. Lim and B. Mulgrew, "Filter banks based JDL with angle and separate Doppler compensation for airborne bistatic radar," *International Radar Symposium*, pp. 583–587, September 2005.

[5] F. Lapierre and J. Verley, "Registration-based solutions to the range-dependence problem in STAP radars," *Proceedings of Adaptive Sensor Array Processing (ASAP) workshop, (MIT Lincoln Laboratory)*, March 2003.

[6] A. Jaffer, B. Himed, and P. Ho, "Estimation of range-dependent clutter covariance by configuration system parameter estimation," *Proceedings of IEEE Radar Conference*, pp. 596–601, May 2005.

[7] F. Lapierre and J. Verly, "Registration-based range-dependence compensation for bistatic STAP radars," *EURASIP Journal on Applied Signal Processing, Special Issue on Advances in Sensor Array Processing Technology*, vol. 1, pp. 85–98, 2005.

[8] C. Lim and B. Mulgrew, "Prediction of inverse covariance matrix (PICM) sequences for STAP," *IEEE Signal Processing Letters*, vol. 13, no. 4, pp. 236–239, April 2006.

[9] R. Adve, P. Antonik, W. Baldygo, C. Capraro, G. Capraro, T. Hale, R. Schneible, and M. Wicks, "Knowledge-base application to ground moving target detection," *Technical Report AFRL-SN-RS-TR-2001-185, Air Force Research Laboratory*, September 2001.

[10] W. Melvin, M. Wicks, P. Antonik, Y. Salama, P. Li, and H. Schuman, "Knowledge-based space-time adaptive processing for airborne early warning radar," *IEEE Aerospace and Electronic Systems Magazine*, vol. 13, no. 4, pp. 37–42, April 1998.

[11] H. B. Rangaswamy, M. and J. Michels, "Performance analysis of the non-homogeneity detector for STAP applications," *Proceedings of IEE Radar Conference*, pp. 193–197, May 2001.

[12] E. Aboutanios and B. Mulgrew, "Maximum likelihood detection of known signals in unknown coloured Gaussian noise," *In Review for IEEE Transactions on Signal Processing.*

[13] ——, "A STAP algorithm for radar target detection in heterogeneous environments," *IEEE Workshop on Statistical Signal Processing*, July 2005.

[14] N. Goodman, "Statistical analysis based on a certain multivariate complex gaussian distribution," *Annals Mathematical Statistics*, vol. 34, pp. 152–177, March 1963.

[15] L. Brennan and I. Reed, "Theory of adaptive radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-9, no. 2, pp. 237–252, March 1973.

[16] E. Kelly, "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, no. 1, pp. 115–127, March 1986.

[17] F. Robey, D. Fuhrmann, K. E., and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, January 1992.

[18] J. Ward, "Space-time adaptive processing for airborne radar," MIT Lincoln Laboratory, Tech. Rep., 1994.

[19] M. Melvin, "Space-time adaptive radar performance in heterogeneous clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, pp. 621–633, April 2000.

[20] S. Haykin, *Adaptive Filter Theory.* Prentice Hall, fourth edition, 2002.

[21] I. Reed, J. Mallett, and L. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-10, no. 6, pp. 853–863, November 1974.

[22] F. Lapierre, M. Droogenbroeck, and J. Verley, "New solutions to the problem of range dependence in bistatic STAP radars," *Proceedings of IEEE Radar Conference*, pp. 452–459, May 2003.

[23] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and sar imaging," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1469–1483, June 1996.

[24] H. Wang and L. Cai, "On adaptive spatial-temporal processing for airborne radar systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 660–670, July 1994.