Michael C. Wicks, Muralidhar Rangaswamy, Raviraj Adve, and Todd B. Hale

Space-Time Adaptive Processing

A knowledge-based perspective for airborne radar

This article provides a brief review of radar space-time adaptive processing (STAP) from its inception to state-of-the art developments. The topic is treated from both intuitive and theoretical aspects. A key requirement of STAP is knowledge of the spectral characteristics underlying the interference scenario of interest. However, these are seldom known in practice and must be estimated using training data. The collection of training data in a given scenario is limited by the scale of change of the interference phenomenon with respect to space and time as well as by system considerations such as bandwidth. Increasingly complex interference scenarios give rise to stressful conditions of training support, and the choice of training data becomes a crucial component of the adaptive process. Additional issues of importance in STAP include the computational cost of the adaptive algorithm as well as the ability to maintain a constant false alarm rate (CFAR) over widely varying interference statistics. This article addresses these topics, developing the need for a knowledge-based (KB) perspective.

Signal detection using adaptive processing in spatial and temporal domains offers significant benefits in a variety of applications including radar, sonar, satellite communications, and seismic systems [1]. The focus here is on signal processing for radar systems using multiple antenna elements that coherently process multiple pulses. An adaptive array of spatially distributed sensors, which processes multiple temporal snapshots, overcomes the directivity and resolution limitations of a single sensor. Specifically, using STAP, i.e., joint adaptive processing in the spatial and temporal domains, creates an ability to suppress interfering signals while simultaneously preserving gain on the desired signal. Using training to estimate interference statistics, this suppression is possible despite lack of a priori knowledge of the interference scenario. Training, therefore, plays a pivotal role in adaptive systems. This article focuses on several aspects of this crucial phase from a KB perspective.

Consider the operation of an airborne phased-array radar with J elements. The radar transmits a pulse in a chosen direction. The goal is to look for a target in this direction (the look angle). This transmitted pulse reflects off (possibly) a target (the desired signal) and the ground (or other clutter interfering sources). On receive, the radar samples this reflected wave at a high rate, with each of the R samples corresponding directly to reflections from a specific range. The sampled signal may also include other interfering effects of electronic counter-measures (ECM), such as jamming. This process is repeated for N pulses transmitted at a rate of the pulse repetition frequency (PRF). The entire received data can therefore be organized in a $J \times N \times R$ data cube [2], [3].

The problem at hand is to detect and locate targets, if they exist, within this data cube. This location is in terms of range (at a primary range cell) and Doppler (velocity) with the angle set to the look angle. In practice, the interference statistics and the target complex amplitude are unknown; thus the detection problem is equivalent to the problem of statistical hypothesis testing in the presence of unknown nuisance parameters. From another point of view, the Doppler-wave number or angle-Doppler spectrum per range cell provides a unique representation of a signal in a three-dimensional plane. Hence, the STAP problem may also be viewed as spectrum estimation where the two-dimensional (2-D) adaptive spectral transform of spatio-temporal data affords separation of the desired target from interference. Indeed, in spatially and temporally white noise, the 2-D Fourier transform is optimal.

[THE DOPPLER-WAVE NUMBER OR ANGLE-DOPPLER SPECTRUM PER RANGE CELL PROVIDES A UNIQUE REPRESENTATION OF A SIGNAL IN A THREE-DIMENSIONAL PLANE.]

In the context of STAP, for each range cell, the interference spectral characteristics correspond to the spatio-temporal covariance matrix of the $JN \times 1$ complex data vector under the target-free condition. The presence of these unknown parameters precludes use of a uniformly most powerful test for hypothesis testing [4]. This is because joint maximization of a likelihood ratio over the domain of unknown parameters becomes mathematically intractable and computationally expensive. Hence, ad hoc approaches have been proposed to overcome this problem. Present day computing power permits the use of well-known tools from statistical detection and estimation theory in the radar problem.

The optimal STAP algorithm assigns a complex weight to each degree-of-freedom (DOF) one range cell at a time. These weights are generally found in the minimum mean square error (MMSE) sense assuming Gaussian interference, requiring prewhitening (inversion of the interference covariance matrix) followed by a matched filter (MF). In the MMSE sense, the theoretically optimal and most straightforward algorithm uses all JN DOF within each range cell, estimating the unknown $JN \times JN$ interference covariance matrix using training data. Clearly, the statistics of this data must match that of the interference, i.e., the training data must be target free and homogeneous. Unfortunately, obtaining an accurate estimate requires a large number of homogeneous training samples that, generally, are not available in practice. This is mainly because the training uses data from the secondary range cells, i.e., range cells other than the primary range cell. Furthermore, even if they were available, the associated computation load makes the optimal approach impractical. This problem is worsened because the STAP process must be repeated for each Doppler and range bin of interest.

There are, therefore, two fundamental issues that limit the application of STAP algorithms in practice: the need for adequate homogeneous training data and the computation load of the algorithm. This article addresses these issues in some detail, drawing from the authors' extensive research in these areas. In the area of algorithms, the discussion covers both the authors' proposals, plus important fundamental contributions beyond. We also discuss the important role of nonhomogeneity detection, covering the basics of ranking and selection theory, the theory of spherically invariant random processes (SIRPs) and the use of a nonhomogeneity detector (NHD) tied to the STAP algorithm used for target detection. There is, unfortunately, no one best algorithm or approach. The article attempts to analyze by placing these algorithms using a KB perspective. We conclude with a preliminary algorithm wherein these issues are tied together in a combined approach that addresses all the critical issues mentioned above.

The following section discusses the STAP problem in some detail, covering early work on radar adaptive signal processing and developing a data model for the algorithms that follow. A discussion of the issue of computation load follows and then the issue of secondary data support. The final sections place the algorithms presented from a KB perspective and concludes the article.

PROBLEM STATEMENT

A radar is a sensor, in our case an antenna array on an airborne platform, that transmits and receives electromagnetic radiation. The transmitted electromagnetic signal impinges on various objects such as buildings, land, water, vegetation, and one or more targets of interest. The illuminated objects reflect the incident wave, which is received and processed by the radar receiver. The reflected signal includes desired signals (targets) but also undesired returns from extraneous objects, designated as clutter. Additionally, there could be one or more jammers, high-powered noise-like signals transmitted as ECM, masking the desired target signals. Finally, the received returns include the ubiquitous background white noise caused by the radar receiver circuitry as well as by man-made sources and machinery. Typically, if it exists, the power of the desired signal return is a very small fraction of the overall interference power (due to clutter, jamming, and noise). The problem at hand is to detect the target, if it exists, within the background of clutter and jammer returns. The key to solving this problem is the availability of suitable models for targets, clutter, and jammers [2], [3], [5]. These models account for the angular position of the target in relation to the receiving array. If moving, the target signature includes the effect of the resulting Doppler frequency.

More precisely, the radar receiver front end consists of an array of J antenna elements, which receives signals from targets, clutter, and jammers. These radiations induce a voltage at each element of the antenna array, which constitutes the measured array data at a given time instant. Snapshots of the measured data collected at N successive time epochs give rise to the spatio-temporal nature of the received radar data. The spatio-temporal product $JN = M$ is defined to be the system dimensionality. Figure 1 uses the angle-Doppler space to illustrate the need for space-time (joint domain) processing. A target at a specific angle and traveling at a specific velocity (corresponding to a Doppler frequency) occupies a single point in this space. A jammer originates from a particular angle but is temporally white (noise like). The clutter, due to the motion of the platform, occupies a ridge in this 2-D space [5]—a clutter patch in front of the moving aircraft has the highest Doppler frequency, while one at broadside has zero Doppler (no relative velocity). The clutter spectrum reflects the two-way beampattern of the transmitted signal.

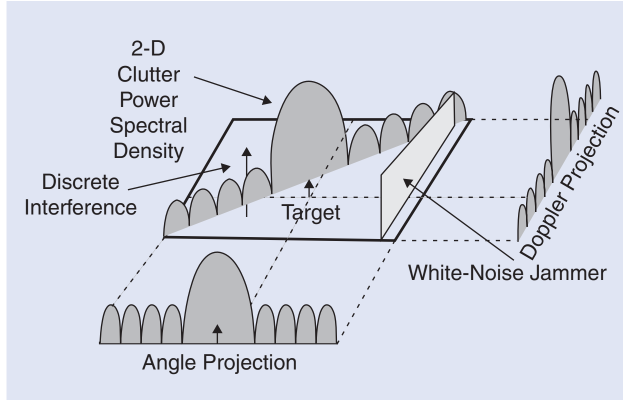
The figure also illustrates the effect of strictly temporal (Doppler) or spatial processing (in angle). The former is equivalent to a projection of the 2-D target plus interference spectrum onto the Doppler plane—however, the target signal is masked by the temporally white jamming. The latter is equivalent to a projection of the said spectrum onto an angular plane, but since the clutter power is strongest at the look angle, the target cannot be distinguished from clutter. However, joint domain processing identifies clear regions in the 2-D plane, which affords recovery of the target from the interference background.

The detection problem can be formally cast in the framework of a statistical hypothesis test of the form:

$$H_0 : \mathbf{x} = \mathbf{d} = \mathbf{c} + \mathbf{j} + \mathbf{n} \tag{1}$$

$$H_1 : \mathbf{x} = \alpha \mathbf{e}(\theta_t, f_t) + \mathbf{d} = \alpha \mathbf{e}(\theta_t, f_t) + \mathbf{c} + \mathbf{j} + \mathbf{n}, \tag{2}$$

where $\mathbf{x} \in \mathcal{C}^{JN \times 1}$ denotes the received data under either hypothesis, $\mathbf{d}$ represents the overall interference being the sum of $\mathbf{c}$ the clutter vector, $\mathbf{j}$ the jammer vector, and $\mathbf{n}$ the background white noise. Finally, $\mathbf{e}$ is a known spatio-temporal steering vector that models the target return for a specific angle-Doppler, and α is the unknown target complex amplitude. For the popular case of a linear array of equispaced elements,



[FIG1] The target and interference scenario in an airborne radar.

$$\mathbf{e} = \mathbf{e}_t \otimes \mathbf{e}_s \tag{3}$$

$$\mathbf{e}_t = \begin{bmatrix} 1 & z_t & z_t^2 & \dots & z_t^{(N-1)} \end{bmatrix}^T, \tag{4}$$

$$\mathbf{e}_s = \begin{bmatrix} 1 & z_s & z_s^2 & \dots & z_s^{(J-1)} \end{bmatrix}^T, \tag{5}$$

$$z_s = e^{j2\pi f_s} = e^{(j2\pi \frac{d}{\lambda} \sin \phi_t)}, \; z_t = e^{j2\pi f_t / f_R}, \tag{6}$$

where ϕ_t and f_t represent the look angle, measured from broadside, and Doppler frequency, respectively; $\otimes$ represents the Kronecker product of two vectors; f_R the PRF; and λ the wavelength of operation. The vectors $\mathbf{e}_t$ and $\mathbf{e}_s$ represent the temporal and spatial steering vectors, respectively. Note that from one pulse to the next and from one element to the next, the steering vectors represent a constant phase shift.

Adaptive algorithms generally determine a weight vector $\mathbf{w}$ to obtain a test statistic, Λ, i.e.,

$$\Lambda = \left| \mathbf{w}^H \mathbf{x} \right|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \lambda, \tag{7}$$

where the ^{H} represents the Hermitian transpose of a vector/matrix, and λ represents a threshold above which a target is declared present. This threshold determines the probability of false alarm, the rate at which a target is detected by mistake. For Gaussian interference statistics, the optimum processing method, corresponding to the case of a known interference covariance matrix $\mathbf{R}_d$, is the whiten-and-match filter for detecting a rank-1 signal given by [6]

$$\mathbf{w} = \frac{\mathbf{R}_d^{-1}\mathbf{e}}{\sqrt{\mathbf{e}^H \mathbf{R}_d^{-1} \mathbf{e}}} \Rightarrow \Lambda_{\mathrm{MF}} = \frac{|\mathbf{e}^H \mathbf{R}_d^{-1} \mathbf{x}|^2}{\mathbf{e}^H \mathbf{R}_d^{-1} \mathbf{e}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{\mathrm{MF}}, \tag{8}$$

which represents the matched filtering of the whitened data $\check{\mathbf{x}} = \mathbf{R}_d^{-1/2}\mathbf{x}$ and whitened steering vector $\check{\mathbf{e}} = \mathbf{R}_d^{-1/2}\mathbf{e}$. It can be readily shown that Λ_{MF} is simply the output signal-to-interference-plus-noise ratio (SINR) of the minimum variance distortionless response (MVDR) beamformer, the maximum likelihood estimate of the target complex amplitude. The relationship between Λ_{MF} and the MVDR beamformer output signal-to-noise ratio (SNR) thus provides a unified perspective of detection and estimation in the context of STAP.

In practice, the covariance matrix, $\mathbf{R}_d$, is unknown and must be estimated. Early work on antenna arrays by Widrow (least squares method) [7] and Applebaum (maximum SNR criterion) [8] suggest use of feedback loops to ensure convergence of iterative methods for calculating the weight vector. However, these methods were slow to converge to the steady-state solution. Fundamental work by Reed, Mallet, and Brennan (RMB beamformer) [9] showed that the sample matrix inverse (SMI) method offered considerably better convergence. In the SMI approach, the basis for most modern STAP algorithms, the interference covariance matrix is estimated using K data ranges for training

$$\hat{\mathbf{R}}_d = \frac{1}{K}\sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^H = \frac{1}{K}\mathbf{X}\mathbf{X}^H, \tag{9}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K]$ and the adaptive weights are obtained as $\mathbf{w} = \hat{\mathbf{R}}_d^{-1}\mathbf{e}$. A drawback of the RMB approach is the lack of a CFAR, i.e., the false alarm rate varies with the interference level, an important consideration in practical systems. Variants of the RMB beamformer to obtain CFAR, such as the Kelly generalized likelihood ratio test (GLRT) [10], the adaptive MF [6], and the adaptive coherence estimator (ACE) [11], were the focus of a number of efforts in the 1980s and early 1990s. Interestingly, the whiten-and-match filter of (8), with the true covariance matrix, $\mathbf{R}_d$, replaced with the estimated covariance matrix, $\hat{\mathbf{R}}_d$, has CFAR. There are, however, three fundamental problems with this approach when applied in the real world: the associated computation load, the need for an adequate number of training samples, and finally and most importantly, the heterogeneity of the available data.

The SMI algorithm requires the solution to a system of linear equations involving a $JN \times JN$ matrix in real time, an $\mathcal{O}(J^3N^3)$ operation. The fact that the algorithm must be executed for each range and Doppler bin of interest exacerbates the problem. Furthermore, to obtain performance within 3 dB of optimum, one requires approximately $K \geq 2JN$ training samples to estimate the $JN \times JN$ matrix $\mathbf{R}_d$. Such a large number of samples are generally not available.

Finally and most importantly, the training data must be homogeneous, i.e., statistically representative of the interference within the range cell of interest. This is generally impossible to obtain in practice due to limitations imposed by the spatio-temporal nonstationarity of the interference as well as by system considerations such as bandwidth and fast scanning arrays. For example, with $J = 11$ and $N = 32$, the parameters for the knowledge-aided sensor signal processing expert reasoning (KASSPER) data set [12], the training data support for 3 dB performance is 704. Assuming an instantaneous RF bandwidth of 500 KHz, this calls for the wide-sense stationarity (homogeneity) over a 400-km range. The scarcity of training data is exacerbated by system errors such as aircraft crabbing and internal clutter motion [5] and environmental considerations such as strong clutter discretes [13], range varying interference spectra and power levels [14], and outlier contamination of training data by target-like signals [15] occurring in dense target scenarios caused by flight formations.

[D^3 ALGORITHMS USE DATA FROM THE PRIMARY RANGE CELL ONLY, AND SO BYPASS THE PROBLEM OF THE REQUIRED HOMOGENEOUS SECONDARY DATA SUPPORT.]

These three issues are interlinked—the computation load is a function of the DOF in the adaptive process and the number of training samples are approximately twice the DOF, i.e., reducing the computation load also reduces the required training. Clearly, reducing the required training also addresses the heterogeneity problem, making it easier to acquire an adequate number of training samples.

As is clear from the above discussion, adequately and effectively training the adaptive filter is essential. The central theme of the following discussion is the use of preexisting and the development of real-time knowledge bases to help in the training process. This knowledge base comprises many aspects—using a priori knowledge in choosing the secondary data, using real time processing to identify homogeneous data samples, and choosing the most effective algorithm based on the available information. The use of KB processing has resulted in the development of the KASSPER program [12]. Using simulated and measured data, the preliminary results, now appearing in the literature, show both the importance of and improvements from using knowledge-aided processing [15]–[19].

LOW COMPUTATION LOAD ALGORITHMS

Succinctly stated, the fully optimal STAP algorithm consists of the following steps:

1) Starting with a data cube, identify the cell under test (corresponding to the length-JN data vector $\mathbf{x}$) and form the target steering vector $\mathbf{e}$ for every Doppler bin of interest.
2) Select K representative training data from both sides of the cell under test, avoiding guard cells to account for target leakage and competing targets.
3) Form $\hat{\mathbf{R}}_d$ the estimated interference covariance matrix using the training data.
4) Calculate a weight vector $\mathbf{w} \propto \hat{\mathbf{R}}_d^{-1}\mathbf{e}$ and apply the weight vector to test cell data to obtain a test statistic, $\Lambda \propto \mathbf{w}^H\mathbf{x}$.
5) Compare the test statistic to a threshold (corresponding to a specified false alarm rate) and declare target presence when test statistic exceeds the threshold.

To overcome the problems associated with this fully adaptive algorithm, researchers have developed alternate, partially adaptive approaches that reduce the DOF with attendant reductions in required sample support and computation cost. Important works in this area include the joint domain localized

(JDL) processing algorithm [20], the parametric adaptive MF (PAMF) [21] (and references therein), the multistage Weiner filter (MWF) [22] and factored STAP methods [5]. Another important approach, not dependent on any statistical training, is the direct data domain (D^3) approach [23]. This algorithm was then extended to include statistical processing in [24]. Several studies show that there is no best algorithm but that an effective implementation would require the use of the most effective from of a library of algorithms. Other than the PAMF, D^3, and MWF (depending on the type of implementation) algorithms, all STAP methods require explicit formation and inversion of the interference covariance matrix, i.e., the issue of homogeneous training data remains.

This section develops in some detail the most popular low computation load algorithms. These algorithms are the most popular for a specific reason—they all address the issue of computation load in innovative though completely different ways. A common framework for reduced DOF process is that they all rely on a transformation of the steering vector and received data into a subspace of dimension $r < JN$. In the following, $\mathbf{T}$ denotes a general transformation matrix, and $\tilde{\mathbf{e}} = \mathbf{Te}$, $\tilde{\mathbf{x}} = \mathbf{Tx}$, and $\tilde{\mathbf{X}} = \mathbf{TX}$ denote the transformed steering vector, test cell data, and training data, respectively. The transformation matrix $\mathbf{T}$ can either be data dependent or data independent. The JDL algorithm is an example of a data independent transformation, while the PAMF, MWF, and LRNAMF [15] are instances of data dependent transformations.

JDL PROCESSING

The JDL algorithm as developed by Wang and Cai [20] maps the received data to the angle-Doppler domain. The transformation to angle-Doppler localizes the target and interference to a few angle and Doppler bins, significantly reducing the required DOF, with corresponding reductions in required sample support and computation load. The authors assume the receiving antenna to be an equispaced linear array of ideal, isotropic, point sensors. Based on this assumption, space-time data is transformed to the angle-Doppler domain using a 2-D discrete Fourier transform (DFT). This approach is only valid in the ideal case under certain restrictions. The presentation here is for the generalized JDL algorithm valid for real world antenna arrays as well [25].

The JDL algorithm begins with a transformation to the angle-Doppler space, i.e., the angle-Doppler response of the data is obtained at the few angle and Doppler bins within the LPR. Mathematically, the angle-Doppler response of the data vector $\mathbf{x}$ at angle ϕ and Doppler f_d is given by

$$\tilde{x}(\phi, f_d) = \mathbf{e}^H(\phi, f_d)\mathbf{x}, \tag{10}$$

where the tilde above the scalar x denotes the transform domain. Repeating this process for η_a angles and η_d Doppler bins (corresponding to $\eta_a\eta_d$ space-time steering vectors) generates a length-$\eta_a\eta_d$ vector $\tilde{\mathbf{x}}$ of angle-Doppler domain data. These η_a angles and η_d Doppler frequencies are said to comprise the localized processing region (LPR). Note that this scheme may be used in conjunction with real world arrays where the space-time steering vector would include a measured spatial steering vector. The scheme reverts to the 2-D DFT in case of an idealized linear array of isotropic point sensors. The transformation matrix $\mathbf{T}$ is given

$$\mathbf{T} = \left[\mathbf{e}_1,\ \mathbf{e}_2,\ \ldots,\ \mathbf{e}_{\eta_a\eta_d}\right], \qquad \tilde{\mathbf{x}} = \mathbf{T}^H\mathbf{x}, \tag{11}$$

where $\mathbf{e}_i$, $i = 1, \ldots, \eta_a\eta_d$ are the steering vectors corresponding to the angles and Dopplers in the LPR. In practice, the angle and Doppler points are chosen to be close to and symmetric around the look angle and Doppler. Note that the transformation matrix is independent of the data. As is usual with a Fourier transform, one could also use a taper, such as a Hamming or Kaiser window, to lower the transformation sidelobes. In the angle-Doppler domain, the adaptive weights are given by

$$\tilde{\mathbf{w}} = \tilde{\mathbf{R}}_d^{-1}\tilde{\mathbf{e}}, \tag{12}$$

$$\tilde{\mathbf{R}}_d = \frac{1}{K}\sum_{k=1}^{K}\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^H, \tag{13}$$

i.e., the JDL algorithm is basically the original SMI algorithm, but using data in the angle-Doppler space.

The steps in implementing the JDL adaptive processor are:

1) Choose the size of the LPR, i.e., η_a and η_d and the number of training data vectors that will be used to estimate the covariance matrix.

2) Set the angle bin to be the look direction and choose a set of η_a angles centered around (and including) the look angle.

3) For each Doppler bin of interest, choose a set of η_d Doppler bins centered around (and including) the look Doppler. Use the set of angles and Doppler bins to form the transformation matrix $\mathbf{T}$ using (11).

4) Transform the entire data cube to angle-Doppler space and find the transformed steering vector $\tilde{\mathbf{e}}$.

5) For each range of interest, estimate an angle-Doppler covariance matrix using (13) and obtain the angle-Doppler weights using (12) to obtain a decision statistic.

Comparing the decision statistic to the chosen threshold, as in (8), completes the detection process. The key here is that K, the number of required homogeneous samples is reduced to about $2\eta_a\eta_d - 4\eta_a\eta_d$ (as opposed to $2JN$) and the computation load to $\mathcal{O}(\eta_a^3\eta_d^3)$. Common values of η_a and η_d, usually odd, are on order of 3, 5, and 7, resulting in enormous savings in computation load and required sample support.

PAMF

The PAMF method is a case of reduced dimension processing that relies on a decomposition of the $JN \times JN$ interference covariance matrix $\mathbf{R}_d$ of the form $\mathbf{R}_d = \mathbf{LDL}^H$, where $\mathbf{L}$ is a lower block-triangular matrix with $J \times J$ identity matrices along the main block diagonal, and $\mathbf{D}$ is a block diagonal matrix with Hermitian matrices $\mathbf{D}_i \in C^{J\times J}$, $i = 1, 2, \ldots N$ [21]. Consequently, $\mathbf{L}^{-1}$ admits a representation of the form

$$\mathbf{L}^{-1} = \begin{bmatrix} \mathbf{I}_J & 0 & 0 & \dots & 0 \\ \mathbf{A}_1^H(1) & \mathbf{I}_J & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}_{N-1}^H(N-1) & \mathbf{A}_{N-1}^H(N-2) & \dots & \mathbf{A}_{N-1}^H(1) & \mathbf{I}_J \end{bmatrix}, \tag{14}$$

where $\mathbf{A}_l(k)$, $k = 1, 2, \dots, N$ denote the coefficients of the lth order multichannel forward linear predictor or multichannel (matrix) autoregressive (AR) linear predictor, and $\mathbf{D}_i$ is the covariance matrix of the residual from the ith order multichannel linear predictor. Thus, the transformation matrix $\mathbf{T}$ takes on the form $\mathbf{T}_{\text{PAMF}} = \mathbf{D}^{(-1/2)}\mathbf{L}^{H^{-1}}$ for the multichannel parametric method. The block form of the transformation is computationally expensive to implement due to the fact that it requires the calculation of all the matrix prediction coefficients of orders 1 to N. Consequently, a sequential method for implementing $\mathbf{T}_{\text{PAMF}}$ is developed in [21]. Furthermore, it has been found for a variety of simulated and measured radar data sets that a low-order ($p = 3$ or 4) multichannel linear predictor provides a good approximation to $\mathbf{R}_d$ [21]. When $\mathbf{R}_d$ is unknown, the adaptive algorithm directly estimates the coefficients of a pth order multichannel linear predictor and the associated block diagonal covariance matrix using training data from the data matrix $\mathbf{X}$. The PAMF test is given by

$$\Lambda_{\text{PAMF}} = \frac{|\tilde{\mathbf{e}}^H\tilde{\mathbf{x}}|^2}{\tilde{\mathbf{e}}^H\tilde{\mathbf{e}}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{\text{PAMF}}, \tag{15}$$

where $\tilde{\mathbf{e}}$ and $\tilde{\mathbf{x}}$ are the steering vector and observed data vector transformed by $\mathbf{T}_{\text{PAMF}}$. The low model order $p << JN$ enables significant reduction in the training data support requirements. Thus, explicit formation and inversion of the interference covariance matrix is avoided. Instead, the coefficients of multichannel linear prediction and the associated residual error covariance matrices, succinctly embed the information contained in $\mathbf{R}_d$. For example, with both simulated and measured data sets, it was shown in [21] that the sample support for $JN = 128$, $p = 3$, and $K = 8$, the PAMF exhibited performance to within 0.5 dB of the optimal MF. The key to obtaining enhanced PAMF performance is the use of efficient parameter estimation algorithms for calculating the multichannel coefficients of linear prediction and the block diagonal error covariance matrix. A number of approaches for this purpose have been discussed in [26]. However, the method best suited for the STAP problem is the covariance method of linear prediction also known as the method of least squares. The computational cost underlying the algorithm is $O((JN)^2p)$, which provides an order of magnitude reduction in the computational cost for $p << JN$. However, when the training data is subject to outlier contamination, the PAMF performance is severely degraded. Mitigating strategies for this problem have been discussed in [27] with other PAMF extensions and variants presented in the references therein. Unfortunately, the detection and false alarm probability for the PAMF test and its variants cannot be calculated using closed form analytical expression. Consequently, these issues are studied using Monte Carlo simulations in [21] and [27] and the references therein.

MULTISTAGE WEINER FILTER

The MWF is another reduced rank processing method, which relies on a serial decomposition of the MVDR beamformer weight vector in the form of a generalized sidelobe canceller (GSC) [22], [28]. The GSC processor relies on a projection of $\mathbf{x}$ onto the signal subspace, $d = \mathbf{e}^H\mathbf{x}$, and a projection onto an orthogonal complement subspace, $\mathbf{b} = \mathbf{B}\mathbf{x}$, where $\mathbf{B}$ is a blocking matrix with orthonormal columns that are orthogonal to $\mathbf{e}$. The GSC weight vector is given by $\mathbf{w}_{\text{GSC}} = [1 \;\; \mathbf{v}]^T$, where $\mathbf{v} = -\mathbf{R}_b^{-1}\mathbf{r}_{bd}$ is a $1 \times M - 1$ row vector, with $\mathbf{R}_b = \mathbf{B}\mathbf{R}_d\mathbf{B}^H$and $\mathbf{r}_{bd} = \mathbf{B}\mathbf{R}_d\mathbf{e}$. The error variance at the MVDR beamformer output can then be expressed as $(1/\mathbf{e}^H\mathbf{R}_d^{-1}\mathbf{e}) = \mathbf{e}^H\mathbf{R}_d\mathbf{e} - \mathbf{r}_{bd}^H\mathbf{R}_b^{-1}\mathbf{r}_{bd}$. This form of the error admits a sequential representation in terms of a Rayleigh quotient and an inverse Rayleigh quotient. More precisely, let $\delta_i = (\mathbf{v}_i^H\mathbf{R}_i\mathbf{v}_i/\mathbf{v}_i^H\mathbf{v}_i)$ and $\xi_i = (\mathbf{v}_i^H\mathbf{v}_i/\mathbf{v}_i^H\mathbf{R}_i^{-1}\mathbf{v}_i)$, where $\mathbf{v}_0 = \mathbf{e}$, $\mathbf{R}_0 = \mathbf{R}_d$, $\mathbf{v}_1 = \mathbf{R}_b^{-1}\mathbf{r}_{bd}$. Then, we have a recurrence relationship between the error residuals at the output of successive stages of the MWF. Specifically, for $i = 0, 1, \dots, M-1$,

$$\xi_i = ||\mathbf{v}_i||^2\delta_i - \frac{||\mathbf{v}_{i+1}||^2}{\xi_{i+1}}, \tag{16}$$

where $||.||$ denotes the norm of a vector. This form of implementation provides a continued fraction expansion of the MVDR beamformer output variance, which results in a tridiagonal covariance structure for the transformed data [22], [28]. Such a form lends itself to an iterative calculation of the MVDR beamformer weight vector via the conjugate gradient method [29]. In sample support deficient scenarios, this method has been found to converge to the principal components inverse method [30]. Key features of this method are the absence of the formation and inversion of the full dimension covariance matrix. Additionally, the MWF implementation in [22] and [28] is computationally expensive due to the need to calculate a sequence of matrix products to recombine the error residuals from the transformed data. However, this is greatly alleviated by the conjugate gradient method which only requires a one-way computation [29]. Performance comparisons of the MWF with competing techniques can be found in [22], [28], and [29], and the references therein.

ISSUES OF DATA SUPPORT

As discussed in this article, an extremely important issue in STAP is the formation and inversion of the covariance matrix underlying the disturbance. In practice, the unknown interference covariance matrix is estimated from a set of independent identically distributed target-free training data, which is assumed to be representative of the interference statistics in a cell under test. Frequently, the training data is subject to contamination by discrete scatterers or interfering targets. In either event, the training data becomes nonhomogeneous. As a result, it is not representative of the interference in the test cell. Hence, standard estimates of the covariance matrix from nonhomogeneous training data result in severely undernulled clutter. Consequently, CFAR and detection performance suffer. Significant performance improvement can be achieved by employing preprocessing to select representative training data.

Figure 2, borrowed from [13], illustrates the importance of nonhomogeneous data. The figure plots the probability of detection (P_d) versus SINR for a false alarm rate of $P_{fa} = 0.01$ and a clutter-to-noise ratio of 40 dB. The system uses $M = JN = 64$ and $K = 128$ range cells to estimate the interference covariance matrix. MF curve is theoretical, corresponding to optimal performance in Gaussian clutter. The curve for the AMF, operating with homogeneous interference has performance within 3 dB of optimal. The curve corresponding to nonhomogeneous data is obtained using Monte Carlo simulations. The training data is corrupted using 30 high-amplitude discrete targets. As is clear, the nonhomogeneity of the data significantly worsens detection performance, by as much as 3–5 dB. Using an NHD [31], the performance of the AMF algorithm is restored.

In general, nonhomogeneity of training data is caused by environmental factors, such as the presence of strong discrete scatterers, dense target environments, nonstationary reflectivity properties of the scanned area, and radar system configurations such as conformal arrays, and bistatic geometries. A variety of robust adaptive signal processing methods to combat specific types of nonhomogeneities have been developed in [15] and [32]–[35]. In this effort, we confine ourselves to the problem of selecting representative training data, when the training data is contaminated by outliers resembling a target of interest (specifically, outliers sharing the same steering vector as a target of interest).

NONHOMOGENEITY DETECTION

The problem of outlier contamination of STAP training data assumes increased significance in dense target scenarios, where outliers resembling a target of interest contaminate the training data. This results in an incorrect threshold setting due to an erroneous estimate of the interference covariance matrix. Furthermore, the presence of outliers in the training data causes target cancellation resulting in degraded output signal to interference ratio and perforce degraded detection performance. A common signal processing method in this context is to excise outliers from the training data and use the resulting outlier free training data for covariance matrix estimation. Several algorithms for outlier removal have been proposed in recent times [13]–[15], [17], [27], [31] in a variety of dense target environments. For the purpose of this section, and in practice, the columns of the data matrix $\mathbf{X}$ are no longer independent, identically distributed and free from outlier contamination. The problem therefore is to classify the columns of $\mathbf{X}$ into groups sharing the same covariance matrix and thereby detect the presence of outliers, which have a deleterious impact on STAP performance, when used in covariance estimation. When outliers are in the form of strong clutter discretes, the generalized inner product (GIP) method [13], [31] and references therein gives a method for outlier removal as summarized below:

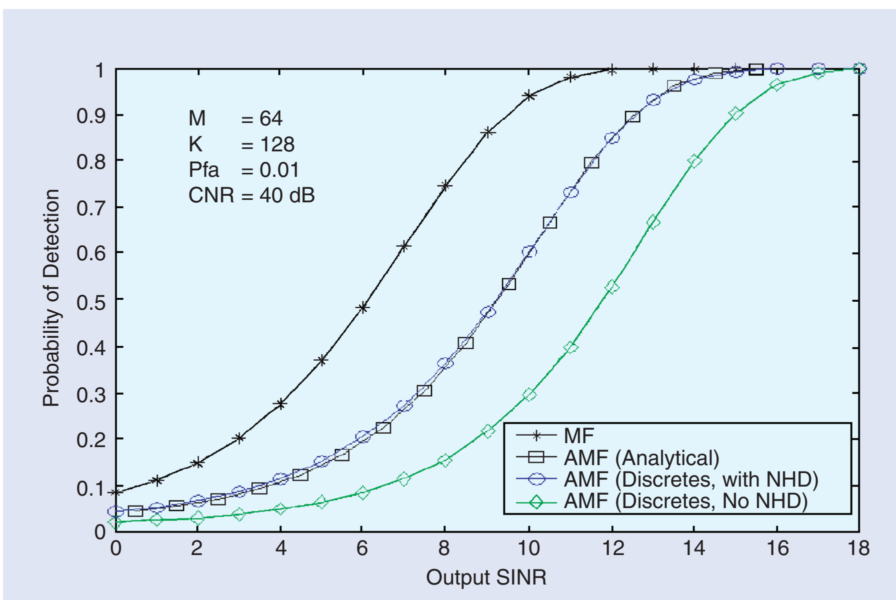
1) First an initial estimate of the covariance matrix using an extended training data set is formed as $\tilde{\mathbf{R}} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^H/L$ where $\tilde{\mathbf{X}}$ is a data matrix with L columns, where $L \gg 2JN$. For example, L may be equal to all available ranges R.

2) Each column of $\tilde{\mathbf{X}}$ is used in a sliding window process to form a statistic $p_i = \mathbf{x}_i^H\tilde{\mathbf{R}}^{-1}\mathbf{x}_i$ for the available range of i. Note that $\tilde{\mathbf{R}}$ for each $\mathbf{x}_i$ is formed by excluding that column and a one column on either side of $\mathbf{x}_i$ (to allow for guard cells).

3) If the columns $\tilde{\mathbf{X}}$ shared the same covariance matrix, empirical realizations of p_i will conform to an F-distribution [13], whose theoretical mean μ is readily calculated.

4) The absolute value $|p_i - \mu|$ are calculated and sorted in increasing order and $K \simeq 2(\mathrm{DOF})$ columns of $\tilde{\mathbf{X}}$ corresponding to p_i showing the least deviation from μ are retained for covariance matrix estimation. The remaining columns are discarded.

Approximately 3–5 dB of performance improvement in the AMF performance in heterogeneous clutter scenarios was demonstrated in [13] using simulated and measured data. However, such an approach relies on full dimension STAP



[FIG2] Impact of nonhomogeneous data on detection performance.

processing and therefore is not suited for conditions of limited sample support. An alternate reduced dimension extension of this procedure known as the innovations power sort was developed in [27], wherein a multichannel linear predictor approximation to $\tilde{\mathbf{R}}$ is employed along the lines of the multichannel AR model described in the PAMF. This form of the estimator has been found to be extremely valuable in conditions of small sample support. The procedure for outlier removal therein is very similar to the GIP approach described previously. Significant performance improvement over competing methods was demonstrated using measured radar data in [27]. When outliers resembling a target of interest contaminate the training data, it becomes imperative to use the steering vector in calculating the test statistic for use in outlier identification and removal. Motivated by this and the need to operate in conditions of limited sample support, the authors in [15] develop an eigen-based method, which relies on the simple principle that the output of a MF peaks when data containing a desired target is passed through the filter. This fundamental idea is used in an iterative manner in [15] to identify the outliers in training data. An extension of this method is pursued in [17] from a KB perspective to significantly reduce the sample support for covariance estimation, while obtaining near clairvoyant STAP detection performance. Other approaches include the use of the adaptive process as a NHD [36].

Theoretical approaches to the problem of nonhomogeneity include use of SIRPs. In other instances, there could be range varying clutter power properties due to environmental and system considerations. In this instance, the clutter statistics depart from the Gaussian behavior, which leads to unacceptably large false alarm rates. This calls into question a suitable model for these impulsive (heavy-tailed) clutter scenarios. There is no unique model for representing the joint probability density function (PDF) of a set of M correlated non-Gaussian random variables. However, a popular model for non-Gaussian radar clutter is the SIRP [14]. Every SIRP is equivalent to the product of a complex Gaussian process and a nonnegative random variable, whose PDF is defined to be the first order characteristic PDF of the SIRP. Consequently, every SIRP is uniquely determined by the specification of a mean vector, a covariance matrix and a characteristic first-order PDF. As a result, the sample covariance matrix is no longer the maximum likelihood estimate for the SIRP covariance matrix. Furthermore, the covariance matrix estimate cannot be calculated in closed form. Instead the maximum likelihood (ML) estimate is a weighted sample covariance matrix, which could be calculated iteratively using the expectation maximization algorithm. Key issues in this context include the convergence properties of the algorithm and the associated numerics. Having obtained the ML estimate of the covariance matrix (which is usually within a multiplicative constant of the covariance matrix of the Gaussian component of the SIRP), a scale invariant test statistic, such as the ACE, is called for. Using the statistics of the ACE test, a formal goodness-of-fit test is developed in [14] to detect and remove outliers. Performance of the approach is presented in [14] using simulated and measured data. The method outperforms all competing candidate algorithms. The extension of this method for sample support starved scenarios is the focus of ongoing research.

D^3 METHODS

Purely statistical algorithms, such as JDL and MWF, cannot suppress a discrete interference source within the primary range cell. For example, a large target within the test range cell but at a different angle and/or Doppler appears as a false alarm, through the sidelobes of the adapted beam pattern, at the look angle-Doppler domain. The secondary data cells do not carry information about the discrete nonhomogeneity, and hence a statistical algorithm cannot suppress discrete (uncorrelated) interference within the range cell under test. This issue of adaptive processing within nonhomogeneous cells has led to the investigation of a new class of algorithms—nonstatistical, or D^3, algorithms [23], [24]. D^3 algorithms use data from the primary range cell only and so bypass the problem of the required homogeneous secondary data support.

The basis of D^3 processing is that, as shown in (6), given the look angle and Doppler, the steering vector determines the phase shift of the target signal from one antenna element/transmitted pulse to the next. The look angle and Doppler determine z_s, the phase shift of the target signal from one antenna element to the next and z_t, the phase shift from one pulse to the next. If $x_j(n)$ represents the total signal at the j th element and n th pulse, terms such as $x_j(n) - z_s^{-1} x_{(j+1)}(n)$ and $x_j(n) - z_t^{-1} x_j(n+1)$ should therefore contain only interference and noise terms. The D^3 approach minimizes the power in these terms while maximizing processing gain in the look direction constant. For example, to determine a set of spatial weights, define the $N \times (J-1)$ interference and noise matrix $\mathbf{A}$ [see (17) at the bottom of the page] and the optimal weights $\mathbf{w}_s$ are the solution to the following optimization problem

$$\mathbf{w}_s^{\mathrm{opt}} = \arg \max_{\mathbf{w}_s, \mathbf{w}_s^H \mathbf{w}_s = 1} \left[\left| \mathbf{w}_s^H \mathbf{e}_{s,0:J-2}(\theta) \right|^2 - \kappa_s \mathbf{w}_s^H \mathbf{A}^T \mathbf{A}^* \mathbf{w}_s \right], \quad (18)$$

where the ^{T} and * represent the transpose and conjugation operators, respectively. This formulation is chosen to remain consistent with the notion that the conjugates of the weights multiply the data. The vector $\mathbf{e}_{s,0:J-2}(\theta)$ represents the first $J-1$ entries

$$\mathbf{A} = \begin{bmatrix} x_0(0) - z_s^{-1} x_1(0) & \cdots & x_{(J-2)}(0) - z_s^{-1} x_{(J-1)}(0) \\ x_0(1) - z_s^{-1} x_1(1) & \cdots & x_{(J-2)}(1) - z_s^{-1} x_{(J-1)}(1) \\ \vdots & \vdots & \vdots \\ x_0(N-1) - z_s^{-1} x_1(N-1) & \cdots & x_{(J-2)}(N-1) - z_s^{-1} x_{(J-1)}(N-1) \end{bmatrix} \quad (17)$$

of the length-J patial steering vector. The use of only $J-1$ weights represents the DOF lost due to the subtraction operation in $x_j(n) - z_s^{-1}x_{(j+1)}(n)$.

The first term in (18) represents the gain of the weight vector in the direction of the look angle, while the second term represents the residual interference power after the data is filtered by the same weights. Hence, the optimal D^3 weights maximize the difference between the gain of the antenna at the look Doppler and the residual interference power. The term κ_s is chosen as a tradeoff between gain and interference cancellation. Using the method of Lagrange multipliers, it can be shown that the desired weight vector is the eigenvector corresponding to the maximum eigenvalue of the $(J-1) \times (J-1)$ matrix $\mathbf{a}_{0:J-2}\mathbf{a}_{0:J-2}^H - \mathbf{A}^T\mathbf{A}^*$. A temporal weight vector $\mathbf{w}_t$ can be found analogously, and overall weight vector is

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_t \\ 0 \end{bmatrix} \otimes \begin{bmatrix} \mathbf{w}_s \\ 0 \end{bmatrix}, \tag{19}$$

where $\otimes$ represents the Kronecker product and the zeros appended represent the loss of one DOF in space and time. The steps in implementing the D^3 processor are the following:

1) Choose the emphasis parameter κ and form matrix $\mathbf{A}$ using (17) and data from within the range cell of interest only.
2) Find the eigenvector corresponding to its largest eigenvalue of $\mathbf{a}_{0:J-2}\mathbf{a}_{0:J-2}^H - \mathbf{A}^T\mathbf{A}^*$. This is $\mathbf{w}_s$.
3) Repeat steps 1 and 2 to obtain a temporal weight vector and then the overall weights $\mathbf{w}$ using (19).

Note that the adaptive weight vector in (18) is obtained using data from the primary range cell only, without estimation of a covariance matrix. This property gives D^3 processing its greatest advantage and its greatest disadvantage. The lack of an estimation of correlation allows use of D^3 processing in severely nonhomogeneous situations. In theory, it could be used by itself, however, the nonhomogeneous range cells have two components of interference—the discrete and the homogeneous components. By their very nature, D^3 algorithms are effective against discrete interference, but they are not as effective against the homogeneous component of the interference. This is because they ignore all statistical information.

HYBRID APPROACH

We present here a hybrid technique, a two-stage process based on the D^3 and JDL algorithms, that combines the benefits of D^3 and statistical processing [24]. Consider the framework of any STAP algorithm. The algorithm processes received data to obtain a complex weight vector for each range bin and each look angle/Doppler. The weight vector multiplies the primary data vector to yield a complex number. The process of obtaining a real scalar from this number for threshold comparison is part of the postprocessing and not inherent to the algorithm itself. The adaptive process effectively estimates the signal component in the look angle/Doppler, i.e., it is a 2-D adaptive spectral estimate. The adaptive weights can therefore be viewed in a role similar to that of the nonadaptive steering vectors in JDL processing, used to transform the space-time data to the angle-Doppler domain.

The JDL processing algorithm begins with a transformation of the data from the space-time domain to the angle-Doppler domain. Statistical adaptive processing within a LPR in the angle-Doppler domain follows. The hybrid approach uses the D^3 weights, replacing the nonadaptive steering vectors used earlier. By choosing the set of look angles and Dopplers to form the LPR, the D^3 weights perform a function analogous to the nonadaptive transform. The D^3 algorithm is used repeatedly with the η_a look angles and the η_d look Doppler frequencies to form the LPR using the same primary data. This implies that there is a main look direction for the overall hybrid STAP process but a set of auxiliary look directions for use with the D^3 algorithm.

[THE JDL ALGORITHM MAPS THE RECEIVED DATA TO THE ANGLE-DOPPLER DOMAIN AND LOCALIZES THE TARGET AND INTERFERENCE TO A FEW ANGLE AND DOPPLER BINS.]

The steps in implementing the hybrid adaptive processor are as follows:

1) Choose the size of the LPR (η_a and η_d), the number of secondary data vectors that will be used to estimate the covariance matrix (usually of the order of $2\eta_a\eta_d$–$4\eta_a\eta_d$) and the number of guard cells (usually 2–4).
2) Choose a set of η_a angles centered around (and including) the look angle.
3) For each range bin and Doppler bin of interest, choose a set of η_d Doppler bins centered around (and including) the look Doppler.
4) Using only the primary data, use the D^3 algorithm repeatedly ($\eta_a\eta_d$ times) with each combination of the chosen angles and Dopplers as the look direction. (Note that this implies that there is a main look angle/Doppler for the overall STAP process but a set of auxiliary look directions for use with the D^3 algorithm.) These $\eta_a\eta_d$ weight vectors form the transformation matrix $\mathbf{T}$ as in (11).
5) JDL processing continues as in (12) and (13).

KNOWLEDGE-AIDED APPROACHES

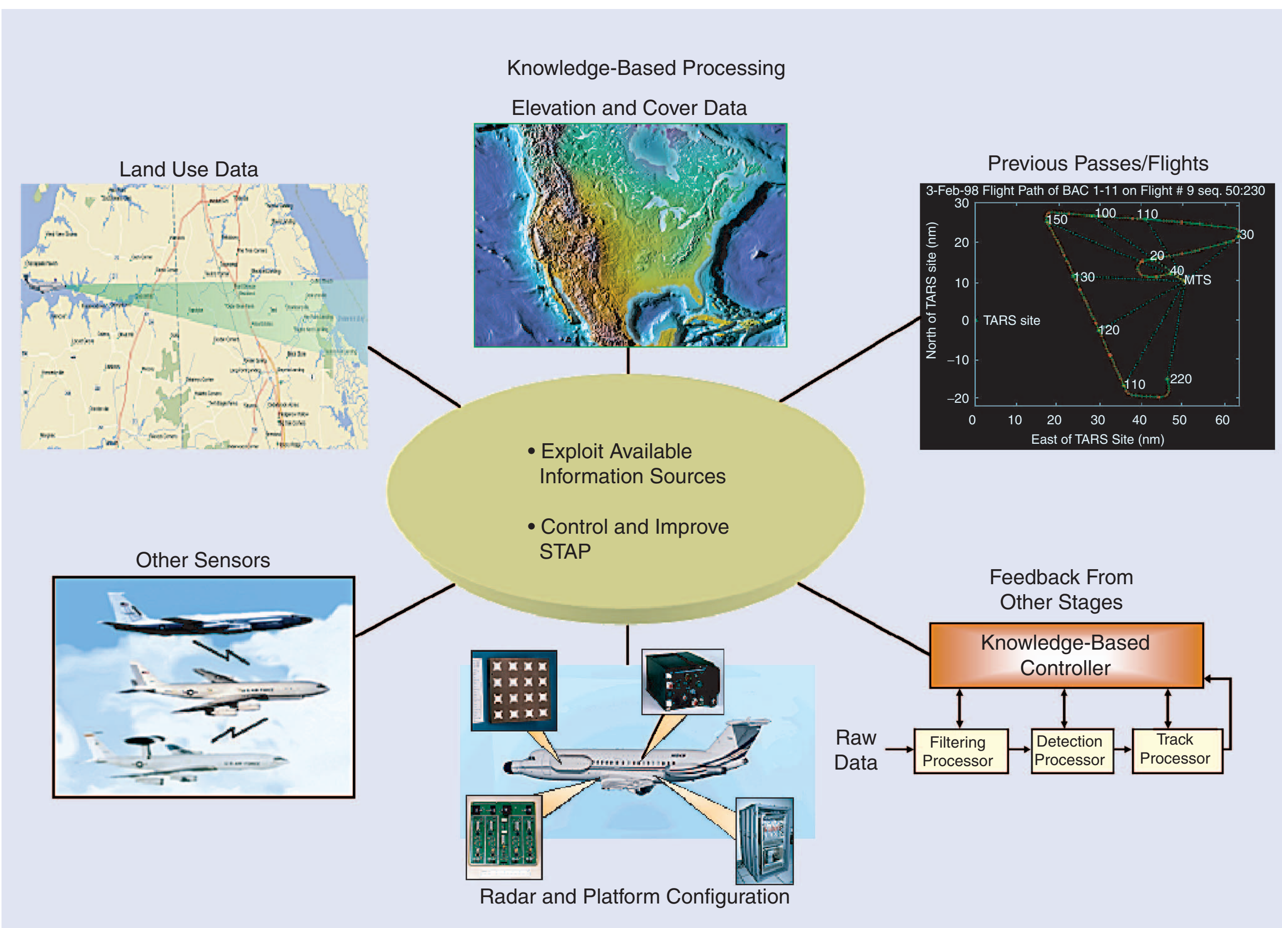
The previous sections have addressed the three fundamental issues associated with practical adaptive processing for airborne radar: computation load, required sample support, and nonhomogeneity detection (including adaptive processing within heterogeneous ranges). Clearly, for each issue, there exists an embarrassment of riches—this article has detailed only a few key schemes addressing each issue. An equally important issue that arises is therefore a scheme to pick within all these potential approaches. One should start with the fundamental notion that there is no one-best approach—different algorithms have their own advantages and disadvantages. This introduces the need for knowledge-aided approaches wherein a database informs the choice of algorithm, sample support both in terms of quantity and choice of range bins, the threshold level that sets the probability of false alarm, potentially even radar parameters such as frequency of transmission, PRF, and transmitted waveform.

Figure 3 illustrates the potential knowledge sources that could be exploited—it includes land-use and coverage data, information from earlier passes over the same terrain, radar parameters, and feedback from other stages in the detection and tracking process. Clearly, this requires a massively complex series of decisions to be made in real time. The figure serves more to illustrate the long-term goal of knowledge-aided processing.

KB STAP WILL REQUIRE SEVERAL ORDERS OF MAGNITUDE GAINS IN AVAILABLE COMPUTATION CAPABILITIES.

A PRELIMINARY KB PROCESSOR

This section implements a very preliminary KB processor [37]. KB processing best matches the adaptive processing algorithm to the interference scenario. The STAP technique is chosen using knowledge gained by processing the received data. In the KB processor of Figure 4, each range cell is classified into one of only two types: homogeneous or nonhomogeneous, with different algorithms used for each type of cell. This classification is made using the NHD based on whether the JDL detection statistic crosses a chosen threshold. Within the range cells deemed nonhomogeneous, the interference is assumed to have discrete and homogeneous components, and the hybrid algorithm is used for target detection. We use the JDL processor. This choice of statistical processing allows for the use of the JDL algorithm in all three components of the KB processor. The only difference between processing in the homogeneous cells and in the nonhomogeneous cells is the choice of transformation matrix. Within the homogeneous cells, the transformation matrix is the nonadaptive transform of (11). Within the nonhomogeneous range cells, the transformation matrix is given by the D^3 weights. In both cases, the secondary data used to estimate the angle-Doppler covariance matrix are chosen from range cells deemed homogeneous.



[FIG3] **Sources informing a KB processor.**

The steps in implementing the simple KB processor are as follows:

1) For each Doppler bin of interest, repeat the following steps:
2) For all range bins, identify homogeneous and nonhomogeneous cells using the JDL-NHD.
3) For each range cell of interest, if it is homogeneous, apply the JDL algorithm, but now using other homogeneous cells as sample support.
4) If it is nonhomogeneous, apply the hybrid algorithm, using other homogeneous cells as sample support.

Another KB processor is the fast maximum likelihood reiterative self-censoring adaptive power residue concurrent block processing two weight vector adaptive cosine estimator (FRACTA) [17], which employs a priori information pertaining to the clutter covariance matrix. The FRACTA method demonstrates near clairvoyant detection performance while employing 30% of the sample support needed in reduced rank STAP. For reduced rank STAP, the RMB rule requires $K = 2r$ (where r is the clutter rank; typically $r \ll M$) training data snapshots to obtain performance within 3 dB of the optimum. Performance analysis of the FRACTA algorithm is carried out using data from the KASSPER program. Due to constraints of space, the interested reader is referred to [17] for further details. Finally, the LRNAMF developed in [15] is another example of knowledge-aided adaptive processing, where a priori information about the clutter rank gained from system parameters such as platform speed, pulse repetition interval, array element spacing, number of antenna array elements, and number of pulses processed in a coherent processing interval is used to significantly reduce the training data support for covariance matrix estimation. Performance of the LRNAMF is benchmarked using data from the KASSPER program.

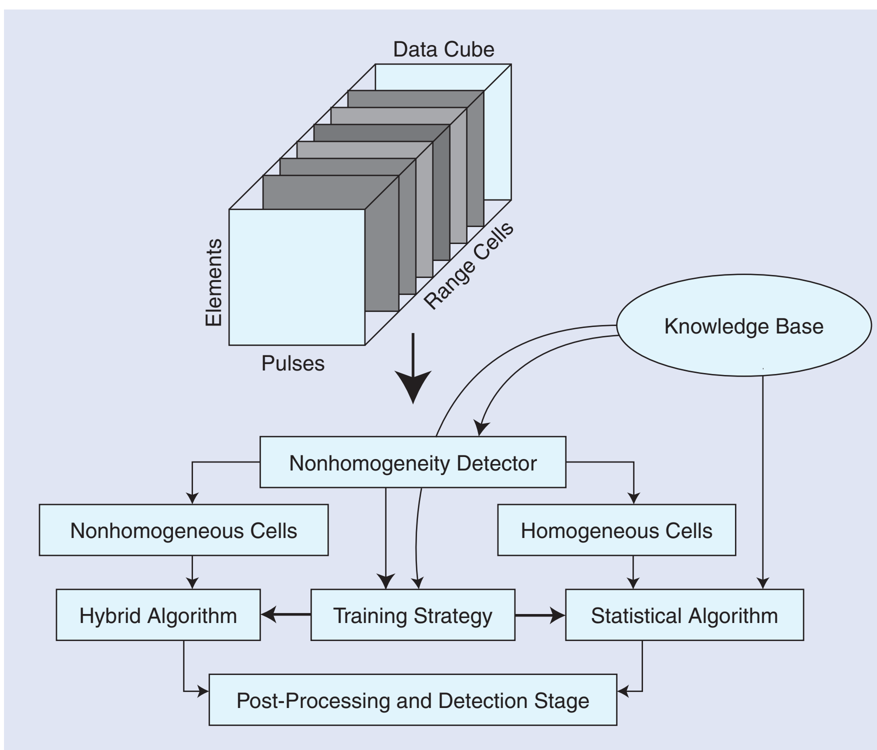
NUMERICAL EXAMPLE

The motivation for the KB processor is practical implementation of STAP in airborne radars for ground moving target indicators (GMTIs). With this in mind, we present here a result of using the KB formulation of Figure 4 using measured data from the multichannel airborne radar measurements (MCARM) program [38]. The example chosen here uses the data from Acquisition 575 on Flight 5. Included with the data is information regarding the position, aspect, and velocity of the airborne platform and the mainbeam transmit direction. This information is used to correlate target detections with ground features.

While recording this acquisition, the radar platform was at latitude-longitude coordinates of (39.379°, −75.972°), placing the aircraft close to Chesapeake Haven, Maryland, USA. The plane was flying mainly south with velocity 223.78 mi/h and east with velocity 26.48 mi/h. The aircraft location and the transmit mainbeam are shown in Figure 5. The mainbeam is close to broadside. Note that the mainbeam illuminates terrain of various types, including several major highways. Each data cube comprises 22 elements ($J = 22$), 128 pulses ($N = 128$) at a PRF of 1984 Hz, and 630 range bins sampled at 0.8 μs (corresponding to 0.075 mi). The array is a 2×11 rectangular array. The array operates at a center frequency of 1.24 GHz.

[KB PROCESSING BEST MATCHES THE ADAPTIVE PROCESSING ALGORITHM TO THE INTERFERENCE SCENARIO.]

To illustrate the effects of nonhomogeneities in secondary training data, we inject two targets at closely spaced range bins. These artificial targets are in addition to the ground targets of opportunity on the roadways illuminated by the array. The artificial targets are injected in range bins 290 and 295. In this acquisition, the zero range is referenced to range bin 74, and so these injected targets are at ranges of 16.2 mi and 16.575 mi respectively. The parameters of the injected targets are given in Table 1. These values are chosen to ensure that the targets cannot be distinguished using nonadaptive, MF processing. Note that the two targets are at the same look angle and Doppler frequency



[FIG4] **A preliminary KB process.**

and the second target is 20 dB stronger than the first.

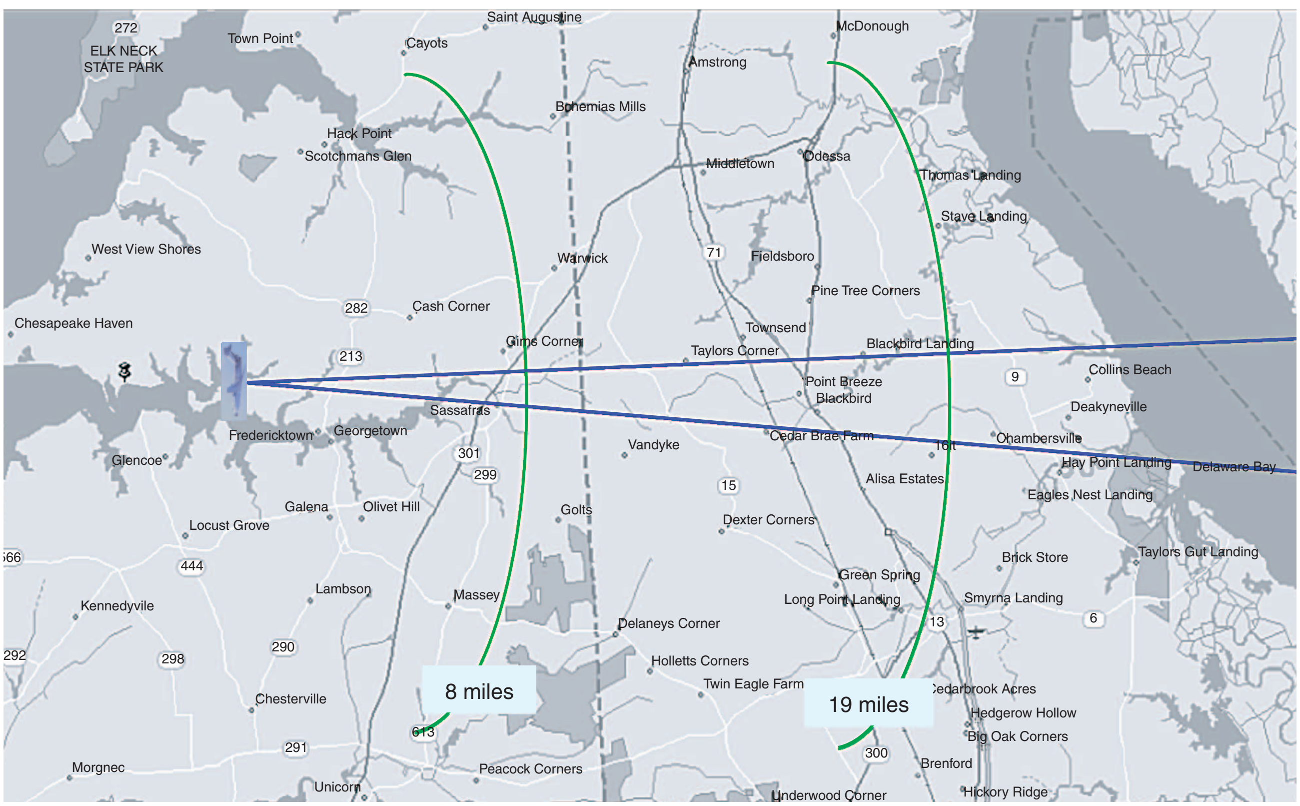
This example is based on the JDL algorithm in all stages. The NHD uses the JDL-NHD discussed earlier, while the statistical algorithm is the JDL algorithm using homogeneous range cells for sample support. The hybrid algorithm, as discussed earlier, is the JDL algorithm with an adaptive D^3 transform to the angle-Doppler domain. All stages use three angle bins and three Doppler bins (a 3×3 LPR). Thirty-six secondary data vectors are used to estimate the 9×9 angle-Doppler LPR covariance matrix. Two guard cells are used on either side of the primary data vector. Based on these numbers, without a NHD stage, range bin 295 would be used as a secondary data vector for detection within range bin 290, violating the homogeneity assumption of statistical STAP algorithms. The example compares the original JDL algorithm of [20] and the KB STAP algorithm of Figure 4.

AN ADAPTIVE ARRAY OF SPATIALLY DISTRIBUTED SENSORS OVERCOMES THE DIRECTIVITY AND RESOLUTION LIMITATIONS OF A SINGLE SENSOR.

Figure 6 plots the results of the original JDL algorithm without attempting to compensate for array effects or nonhomogeneities. The plot is of the modified sample matrix inversion (MSMI) statistic as a function of range and Doppler. The red spots correspond to higher statistics, i.e., the red tend to correspond to target detections. The figure shows that targets are detected in almost all range and Doppler bins, including at extremely high velocities. If using the original JDL algorithm with measured data, therefore, one must deal with several false alarms. Also, while the second injected target is clearly visible, the first target is not detected at all. This inability to detect the target is because the second target is present in the secondary data while attempting to detect the first target at range bin 290. The presence of a target-like nonhomogeneity in the secondary data makes detection of a weak target practically impossible.

The KB processor, illustrated in Figure 4, matches the processing to the interference in that it uses JDL processing in the homogeneous range cells and hybrid processing in the nonhomogeneous cells. Figure 7 plots the AMF statistic obtained by using the KB processor. The improved discrimination, as compared to Figures 6, between a few target signals and residual interference is clear. The first target is now clearly visible. This is possible because the NHD treats the second injected target as a nonhomogeneity, and it is eliminated from the secondary data while processing the range cell corresponding to the first, weaker, injected target. The KB processor can, therefore, detect weak targets buried in nonhomogeneous interference.

The final step in determining the presence or absence of a target is to apply a threshold to the MSMI statistic of Figures 6 and 7 to yield target declarations. Here, a target is declared at all



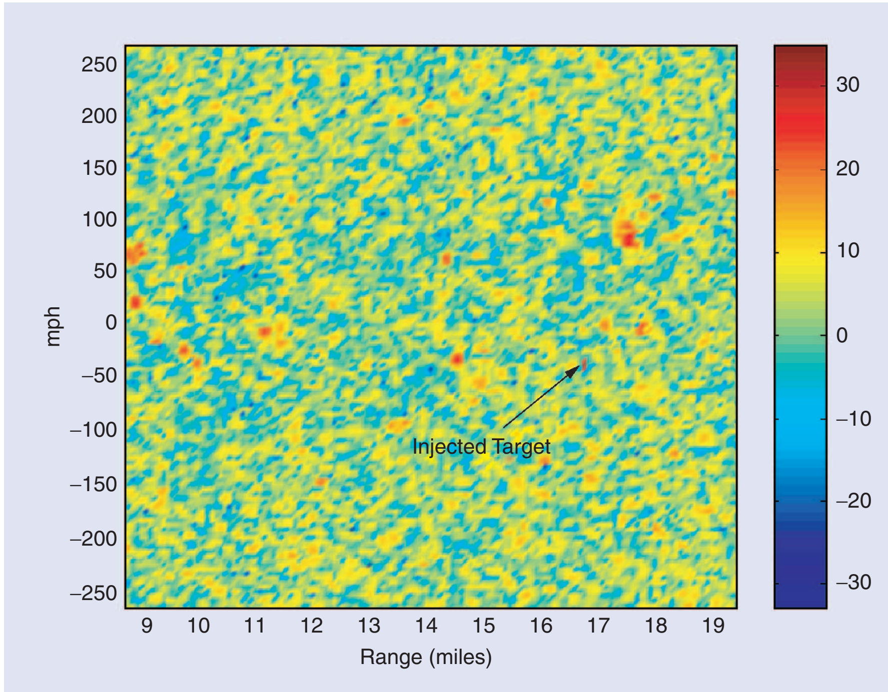
[FIG5] **Location and transmit direction of the MCARM airplane during acquisition 575.**

points with an estimated MVDR statistic of greater than 40. Figures 8 and 9 plot the declared target locations as a function of Doppler and range. These locations are correlated with the map of Figure 5. In Figure 8, note the extremely high number of false alarms. Also, as in Figure 6, the weak injected target is not detected. On the other hand, nearly all the target declarations by the KB processor in Figure 9 correlate directly with major highways in Maryland and Delaware illuminated by the radar mainbeam. Routes 290 and 301 in Maryland are closely spaced at a range of 9.0 and 9.8 mi. Accounting for the platform motion, the ground speed of the target(s) is approximately 50 mi/h.

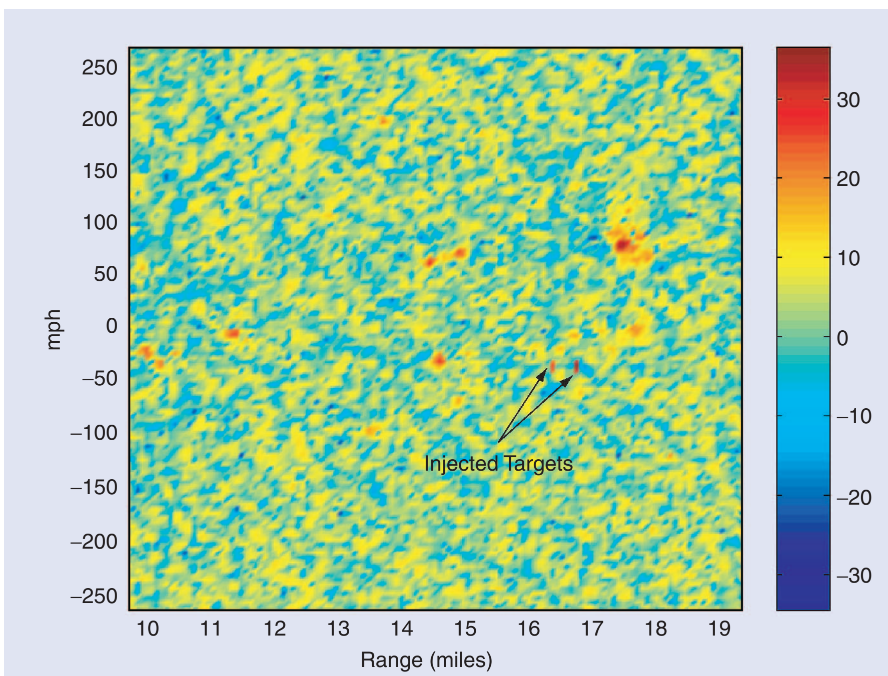
The target detections at the far range shown in the plot are between 19.4 and 20.4 mi. The range to Route 9 varies between 19.1 and 21.1 mi within the transmit mainbeam. These far range detections therefore correspond to Route 9. The targets detected at these ranges are present in both Figures 8 and 9.

[TABLE 1] PARAMETERS DEFINING THE INJECTED TARGETS.

	TARGET 1	TARGET 2
AMPL	1×10^{-4}	1×10^{-3}
ANGLE BIN	1°	1°
DOPPLER	$-9 \equiv -139.5$ Hz	$-9 \equiv -139.5$ Hz
RANGE BIN	$290 \equiv 16.2$ Mi	$295 \equiv 16.575$ Mi



[FIG6] JDL processing ignoring array effects and nonhomogeneities.

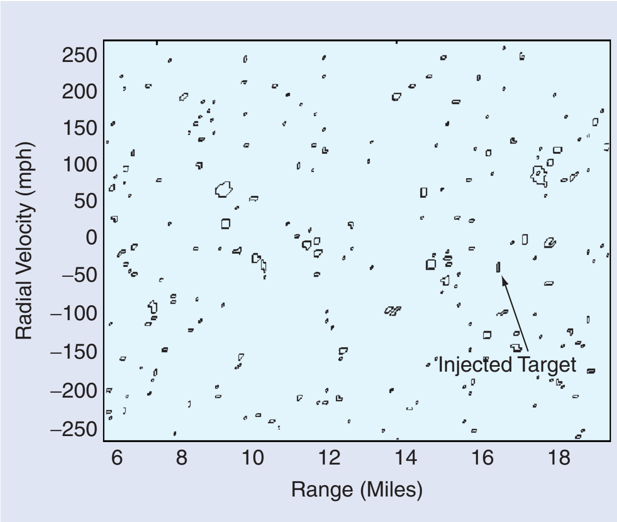


[FIG7] KB processor matching the STAP algorithm to the interference scenario.

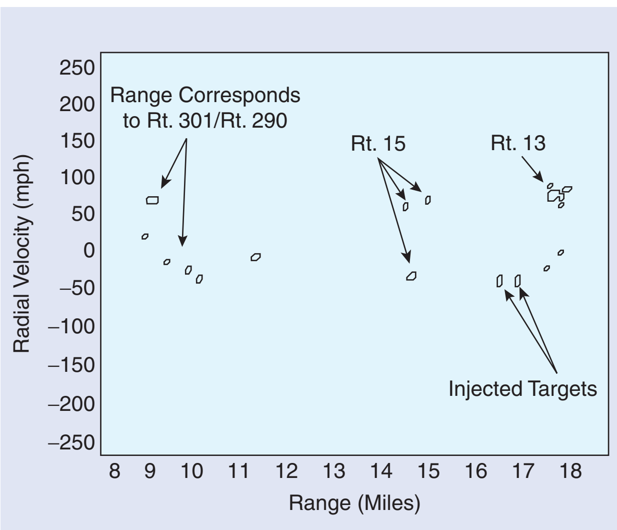
A LONG-TERM VIEW

Several years of research has shown that KB approaches are essential for a practical implementation of STAP in airborne radars. The twin issues of data nonhomogeneity and adequate data support necessitate real-time analysis of the received data and the choice of an appropriate adaptive algorithm (with its associated parameters). As shown in Figure 3, there are several knowledge sources that make the decision process more effective. Furthermore, as the references show, recent research has developed the many pieces of the overall knowledge-aided STAP puzzle.

The fundamental question that is yet to be addressed is feasibility. As must be clear, STAP by itself is a computationally complex process. Receiving, basebanding, and processing multichannel signals in real time places an enormous burden on available digital signal processing technologies. As it stands, KB STAP will require several orders of magnitude gains in available computation capabilities. Furthermore, as is



[FIG8] **Target declarations using JDL ignoring array effects and nonhomogeneities.**



[FIG9] **Target declarations using KB STAP.**

readily apparent, there is no single best solution, and every choice involves a tradeoff. The feasibility of implementation of KB STAP needs to be examined on a case-by-case basis. The computational requirements, storage, access, and communications overhead in addition to system considerations such as size, weight, power, and cost will dictate the implementation of KB STAP for each application of interest. We predict that looking forward, it will be these factors that limit what will or will not be implemented.

CONCLUSIONS

This review has attempted to provide the reader an intuitive and theoretical basis of STAP. The focus has been on the importance of STAP, the fundamental issues that have guided research in this area. Two central problems arise in the application of STAP—the issue of computation load and the homogeneity of the sample support needed to train the adaptive filter. There have been several algorithms to address either of these issues, the key concepts of which have been presented here. However, most researchers would agree that there is no one best algorithm, and the only practical approach is to use a KB scheme that best matches the signal processing to the interference scenario at hand. This matching could be in the choice of adaptive algorithm including its parameters, the scheme used to distinguish nonhomogeneities and the training data used.

We presented an example of using a preliminary KB processor on measured data. The example illustrates the immense potential of KB approaches in detecting weak targets and reducing false alarm rates. However, it must be emphasized that the algorithm used is extremely simple—in fact, the example emphasizes the vast amount of work remaining, such as that undertaken in the KASSPER program.

ACKNOWLEDGMENTS

Muralidhar Rangaswamy's work on this article was supported by the Air Force Office of Scientific Research under project 2304IN. The authors would like to acknowledge Mr. Genello from AFRL and Dr. Guerci from DARPA for their continued support of this research. The authors wish to thank Dr. J. Guerci for permission to use, in Figure 3, a picture from the DARPA/SPO KASSPER program.

AUTHORS

Michael C. Wicks is a member of the scientific and professional cadre of senior executives and the senior scientist for sensor signal processing in the U.S. Air Force. He specializes in the science and technology needed for superior U.S. air and space systems for reconnaissance, surveillance, precision engagement, and electronic warfare. He is a scientific and technical adviser specializing in radio frequency sensors, radar systems, and signal processing research as it applies to advanced ground, air, and space systems. He began his career with the Air Force in 1981 after receiving his B.S.E.E. from Rensselaer Polytechnic Institute, Troy, New York. In 1995 he earned his Ph.D. in electrical engineering from Syracuse University, New York. He holds 12 U.S. patents, and he has authored or coauthored two books, three book chapters, and more than 125 journal, conference, and technical papers.

Muralidhar Rangaswamy received the B.E. degree in electronics engineering from Bangalore University, India in 1985, the M.S. and Ph.D. degrees in electrical engineering from Syracuse University, New York, in 1992. He is currently a senior electronics engineer at the Air Force Research Laboratory (AFRL) SNHE, Hanscom Air Force Base, Massachusettes. His research interests include radar signal processing, spectrum estimation, modeling non-Gaussian interference phenomena, and statistical communication theory. He has authored more than 70 journal and conference papers in these areas in addition to one book and two book chapters. He is a co-inventor on two U.S. patents. He received the 2005 Nathanson Memorial Radar Engineer award from the IEEE Aerospace and Electronic

Systems Society and the Charles Ryan Memorial basic research award from the Sensors Directorate of AFRL as well as 20 AFRL scientific achievement awards.

Ravi Adve received his B.Tech. degree in electrical engineering from IIT, Bombay, in 1990 and his Ph.D. from Syracuse University, New York in 1996. Between 1997 and August 2000, he worked for Research Associates for Defense Conversion Inc. on contract with the Air Force Research Laboratory at Rome, New York. He joined the faculty at the University of Toronto in August 2000. His research interests include practical signal processing algorithms for smart antennas with applications in wireless communications and radar. He is currently focused on linear precoding in wireless communications, cooperative communications, and augumenting STAP with a waveform dimension.

Todd B. Hale is an assistant professor of electrical engineering and chief of the Electrical Engineering Division, Department of Electrical and Computer Engineering, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio. He obtained a B.S.E.E. degree from Milwaukee School of Engineering, Wisconsin, in 1993 and M.S.E.E. and Ph.D. degrees from the Air Force Institute of Technology in 1997 and 2002, respectively. His research interests include adaptive and statistical signal processing for radar, STAP, radar interference modeling, image processing, automatic target recognition, and electronic warfare. As an AFIT professor, he is responsible for the entire radar curriculum.

REFERENCES

[1] R.A. Monzingo and T.W. Miller, *Introduction to Adaptive Arrays*. New York: Wiley, 1980.

[2] R. Klemm, *Principles of Space-Time Adaptive Processing.* Bodmin, U.K.: IEE Press, 2002.

[3] J.R. Guerci, *Space-Time Adaptive Processing for Airborne Radar*. Norwood, MA: Artech House, 2003.

[4] H.L. VanTrees, *Detection, Estimation and Modulation Theory Part-I*. New York: Wiley, 1968.

[5] J. Ward, "Space-time adaptive processing for airborne radar," MIT Lincoln Laboratory, Tech. Rep. 1015, Dec. 1994.

[6] F. Robey, D. Fuhrmann, E. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, 1992.

[7] B. Widrow, P.E. Mantey, L.J. Griffiths, and B.B. Goode, "Adaptive antenna systems," *Proc. IEEE*, vol. 55, no. 12, pp. 2143–2161, 1967.

[8] S. Applebaum, "Adaptive arrays," Syracuse Univ. Res. Corp., Tech. Rep. SPL TR 66-1, Aug. 1966.

[9] I. Reed, J. Mallett, and L. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 6, pp. 853–863, 1974.

[10] E. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 115–127, 1986.

[11] S. Kraut, L.L. Scharf, and L. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Processing*, vol. 49, no. 1, pp. 1–16, 2001.

[12] J. Bergin and P. Techau, "High-fidelity site-specific radar simulation: Kassper'02 workshop datacube," Defense Advanced Res. Projects Agency, Tech. Rep. ISL-SCRD-TR-02-105, 2002.

[13] M. Rangaswamy, J.H. Michels, and B. Himed, "Statistical analysis of the non-homogeneity detector for STAP applications," *Dig. Signal Process.*, vol. 14, no. 3, pp. 253– 267, 2004.

[14] M. Rangaswamy, "Statistical analysis of the non-homogeneity detector for non-Gaussian interference backgrounds," *IEEE Trans. Signal Processing*, vol. 53, no. 6, pp. 2101–2111, 2005.

[15] M. Rangaswamy, F. Lin, and K. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Signal Process.*, vol. 84, no. 9, pp. 1653–1665, 2004.

[16] D.J. Zywicki, W.L. Melvin, G.A. Showman, and J.R. Guerci, "Stap performance in site-specific clutter environments," in *Proc. 2003 IEEE Aerospace Conf.*, Apr. 2004, pp. 2005–2020.

[17] S. Blunt, K. Gerlach, and M. Rangaswamy, "The enhanced FRACTA algorithm with knowledge aided covariance estimation," in *Proc. Third IEEE Workshop on Sensor Array Multichannel Processing*, Barcelona, Spain, 2004.

[18] J. Bergin, C. Teixeira, P. Techau, and J. Guerci, "STAP with knowledge aided data pre-whitening," in *Proc. IEEE Radar Conf.,* Philadelphia, PA, 2004, pp. 289–294.

[19] W.L. Melvin, G.A. Showman, and J.R. Guerci, "A knowledge-aided GMTI detection architecture," in *Proc. 2004 IEEE Radar Conf.*, Apr. 2004, pp. 301–306.

[20] H. Wang and L. Cai, "A localized adaptive MTD processor," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, no. 3, pp. 532–539, 1991.

[21] J. Roman, M. Rangaswamy, D. Davis, Q. Zhang, B. Himed, and J. Michels, "Parametric adaptive matched filter for airborne radar applications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 677–692, 2000.

[22] J. Goldstein, L. Scharf, and I. Reed, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2943–2959, 1998.

[23] T.K. Sarkar and N. Sangruji, "An adaptive nulling system for a narrow-band signal with a look-direction constraint utilizing the conjugate gradient method," *IEEE Trans. Antennas Propagat.*, vol. 37, no. 7, pp. 940–944, July 1989.

[24] R.S. Adve, T.B. Hale, and M.C. Wicks, "Joint domain localized adaptive processing in homogeneous and non-homogeneous environments, Part II: Non-homogeneous environments," *Proc. Inst. Elect. Eng. (Radar Sonar Navig.)*, vol. 147, pp. 66–73, Apr. 2000.

[25] R.S. Adve, T.B. Hale, and M.C. Wicks, "Joint domain localized adaptive processing in homogeneous and non-homogeneous environments. Part I: Homogeneous environments," *Proc. Inst. Elect. Eng. (Radar Sonar Navig.)*, vol. 147, pp. 57–65, Apr. 2000.

[26] S. Marple, *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice Hall, 1987.

[27] J.H. Michels, B. Himed, and M. Rangaswamy, "Robust STAP detection in a dense signal airborne radar environment," *Signal Process.*, vol. 84, no. 9, pp. 1625–1636, 2004.

[28] J. Goldstein, I. Reed, and P. Zulch, "Multistage partially adaptive STAP CFAR detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 2, pp. 645–661, 1999.

[29] M. Zoltowski, "Conjugate gradient adaptive filtering with application to space time processing for wireless digital communications," in *Proc. Second IEEE Conf. Sensor Array and Multichannel Processing*, Key Bridge, VA, Aug. 2002.

[30] I. Kirsteins and D. Tufts, "Adaptive detection using a low rank approximation to a data matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 1, pp. 55–67, 1994.

[31] W.L. Melvin, M.C. Wicks, and P. Chen, "Nonhomogeneity detection method and apparatus for improved adaptive signal processing," U.S. Patent 5 706 013, 1999.

[32] Y.-L. Wang, J.-W. Chen, Z. Bao, and Y.-N. Peng, "Robust space-time adaptive processing for airborne radar in nonhomogeneous clutter environments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 1, pp. 70–81, 2003.

[33] J. Guerci and J. Bergin, "Principal components, covariance matrix tapers, and the subspace leakage problem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 1, pp. 152–162, 2002.

[34] K.R. Gerlach, "Outlier resistant adaptive matched filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 885–901, 2002.

[35] S. Vorobyov, A. Gershman, and Z. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Processing*, vol. 51, no. 2, pp. 313–324, 2003.

[36] R.S. Adve, T.B. Hale, and M.C. Wicks, "Transform domain localized processing using measured steering vectors and non-homogeneity detection," in *Proc. 1999 IEEE Nat. Radar Conf.*, Boston, MA, Apr. 1999, pp. 285–290.

[37] R.S. Adve, T.B. Hale, M.C. Wicks, and P.A. Antonik, "Ground moving target indication using knowledge based space-time adaptive processing," in *Proc. 2000 IEEE Int. Radar Conf.*, Washington, DC, May 2000, pp. 735–740.

[38] D. Sloper, D. Fenner, J. Arntz, and E. Fogle, "Multi-channel airborne radar measurement (MCARM), MCARM fight test," Westinghouse Electronic Systems, Contract F30602-92-C-0161, Apr. 1996. [Online] Available: http://sunrise.deepthought.rl.af.mil