Article

Fast Variational Bayesian Inference for Space-Time Adaptive Processing

Xinying Zhang, Tong Wang * and Degen Wang

National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China;
zxying@stu.xidian.edu.cn (X.Z.); wangdg@stu.xidian.edu.cn (D.W.)
* Correspondence: twang@mail.xidian.edu.cn

Abstract: Space-time adaptive processing (STAP) approaches based on sparse Bayesian learning (SBL) have attracted much attention for the benefit of reducing the training samples requirement and accurately recovering sparse signals. However, it has the problem of a heavy computational burden and slow convergence speed. To improve the convergence speed, the variational Bayesian inference (VBI) is introduced to STAP in this paper. Moreover, to improve computing efficiency, a fast iterative algorithm is derived. By constructing a new atoms selection rule, the dimension of the matrix inverse problem can be substantially reduced. Experiments conducted on the simulated data and measured data verify that the proposed algorithm has excellent clutter suppression and target detection performance.

Keywords: space-time adaptive processing; variational Bayesian inference; clutter suppress; sparse recovery

1. Introduction

Space-time adaptive processing (STAP) is widely studied and applied in airborne phased array radar for clutter suppression and target detection [1–5]. The performance of STAP is related to the clutter plus noise covariance matrix (CNCM), which is estimated by the training samples adjacent to the cell under test (CUT). More than two times the system degrees of freedom (DOFs) of independent and identically distributed (IID) training samples are required to estimate the CNCM [6,7]. However, it is difficult to obtain enough training samples in a practical environment.

Various algorithms have been proposed to improve the STAP performance with a small number of training samples. Reduced-rank (RR) approaches make use of the structural characteristics of the covariance matrix to reduce the requirement for samples, such as principal components (PC), the multistage winner filter (MWF), and cross-spectral metric (CSM) [8–11]. Reduced-dimension (RD) approaches reduce the requirement for samples by selecting space-time or angle-Doppler channels, such as joint-domain localization (JDL), the extend factor approach (EFA), and auxiliary channel processor (ACP) [12–15]. However, the performance of the RR or RD approaches is affected due to complex practical environments. Knowledge-aided (KA) [16–18] approaches use prior knowledge to improve clutter suppression performance. Nevertheless, it is hard to obtain precise prior knowledge. Additionally, some covariance estimation algorithms in limited samples have been proposed, such as the fast maximum likelihood (ML) estimation [19], rank-constrained ML estimation [20], knowledge-aided ML estimation [21], and distribution-free covariance estimation [22]. These approaches are sensitive to model mismatch.

Motivated by the compressed sensing (CS) theory, sparse recovery STAP (SR-STAP) methods have attracted much attention over the last few years [23–26]. Using the norm minimization method and the sparsity of the clutter spectrum, SR-STAP approaches can improve the clutter suppression performance with few samples or even a single sample.

Currently, several SR methods have been proposed. In [27], a convex optimization algorithm named the focal underdetermined system solver (FOCUSS) was proposed, which uses the ℓ_p norm penalty to relax the cost function. This kind of algorithm requires a suitable regularization parameter. However, the parameter chosen empirically is not robust to environmental change. A greedy pursuit approach, an orthogonal matching pursuit, is presented in [28]. It performs well when the signal sparsity is known. The difficulty in obtaining signal sparsity in practice limits the application of this type of approach. For its self-regularizing nature, sparse Bayesian learning (SBL) has received much attention and has been applied to STAP [29–31]. However, the problems of heavy computational burden and slow convergence speed prevent its practical application.

Variational Bayesian inference (VBI) can be seen as a Bayesian approximate method that deals with the SR problem by maximizing a lower bound of likelihood function [32–34]. Compared with SBL, VBI replaces the point estimation of parameters with approximate posterior distribution estimates, which gives it a fast convergence speed. The VBI has been widely used in the fields of engineering and signal processing, some of which include model degradation analysis [35], statistical properties estimation [36], and Orthogonal Time Frequency Space (OTFS) modulation [37]. In this paper, the VBI algorithm with multiple measurement vectors (MMV) is introduced in STAP (M-VBI-STAP). Because of the matrix inverse operation, the computational burden is still heavy. To reduce the computational complexity, a fast STAP algorithm based on VBI is proposed. The following are the main contributions of this paper:

(1) The frame of VBI with multiple measurement vectors is generalized to the STAP application. This allows for an increase in convergence speed.
(2) In order to improve computational efficiency, a fast STAP algorithm based on VBI (M-FVBI-STAP) is proposed. A new atoms selection rule is constructed, which is able to substantially reduce the dimension of the matrix inverse problem.
(3) Numerous experiment results on simulated data and measured data illustrate that the proposed algorithm can achieve better clutter suppression and target detection performance.

The remainder of this paper is outlined as follows. The signal model and Bayesian model are introduced in Section 2. In Section 3, the VBI algorithm is extended to STAP and proposes a fast STAP method based on VBI. Computational complexity analysis is detailed in Section 4. Section 5 shows the experimental results. The conclusions are presented in Section 6.

Notation: Matrices, vectors, and scalar quantities are denoted by boldface uppercase, boldface lowercase, and lightface letters, respectively. The symbols $\mathbb{C}$ and I represent the complex filed and the identity matrix. For a matrix X, $(X)^{-1}$, $(X)^T$, and $(X)^H$ denote the inverse, transpose and conjugate transpose of X, respectively. $\| \bullet \|_2$, $\| \bullet \|_{2,0}$, and $\| \bullet \|_F$ denote the ℓ_2, $\ell_{2,0}$ and Frobenius norms, respectively. $diag(x)$ denotes a diagonal matrix whose diagonal element is x. $\otimes$ stands for the kronecker product. $\langle \bullet \rangle$ denotes the expectation.

2. Signal Model and Bayesian Model

2.1. Signal Model

The system under consideration is a side-looking airborne-phased array radar with a uniform linear array (ULA), which consists of N elements spaced at half a wavelength. The radar transmits M pulses at a constant pulse repetition frequency (PRF) in a coherent processing interval (CPI). Ignoring the range ambiguity, the received space-time signal is modeled as

$$\begin{aligned} x &= x_c + x_t + n \\ &= \sum_{i=1}^{N_c} \alpha_i s(f_{d,i}, f_{s,i}) + \alpha_t a_t + n \end{aligned} \tag{1}$$

where x_c is the clutter component; N_c denotes the number of clutter patches for each range cell; x_t stands for the target component; n denotes the complex white Gaussian noise; $f_{d,i}$ and $f_{s,i}$ are the normalized Doppler frequency and normalized spatial frequency; α_t and α_i are the complex amplitude of the target and the i-th clutter patch, respectively; a_t represents the space-time steering vector of the target and $s(f_{d,i}, f_{s,i})$ is the space-time steering vector of the clutter with the form

$$s(f_{d,i}, f_{s,i}) = s_t(f_{d,i}) \otimes s_s(f_{s,i}) \tag{2}$$

where $s_t(f_{d,i})$ and $s_s(f_{s,i})$ are the temporal steering vector and spatial steering vector

$$s_t(f_{d,i}) = [1, \exp(j2\pi f_{d,i}), \cdots, \exp(j2\pi(M-1)f_{d,i})]^T \tag{3}$$

$$s_s(f_{s,i}) = [1, \exp(j2\pi f_{s,i}), \cdots, \exp(j2\pi(N-1)f_{s,i})]^T \tag{4}$$

The optimum weight of the STAP filter is obtained from the problem,

$$\begin{cases} \min_{w} w^H R_{c+n} w \\ s.t. \ w^H a_t = 1 \end{cases}. \tag{5}$$

The optimum weight of the STAP filter can be acquired by

$$w_{opt} = \frac{R_{c+n}^{-1} a_t}{a_t^H R_{c+n}^{-1} a_t}. \tag{6}$$

This assumes that the clutter and noise components are mutually uncorrelated. Then, the clutter plus noise covariance matrix (CNCM) can be written as

$$\begin{aligned} R_{c+n} &= R_c + R_n \\ &= \langle xx^H \rangle + \sigma^2 I \end{aligned} \tag{7}$$

where $R_c \in \mathbb{C}^{MN \times MN}$ and $R_n \in \mathbb{C}^{MN \times MN}$ represent the CCM and noise covariance matrix and σ^2 is the power of the noise.

For the SR-STAP algorithms, the angle-Doppler plane is uniformly discretized into $K = N_s N_d$ grids. Here, $N_s = \rho_s N$ and $N_d = \rho_d M$ are the number of angle bins, and Doppler bins, respectively. The received signal $X \in \mathbb{C}^{NM \times L}$ from L IID training samples can be written as

$$X = \Phi A + N \tag{8}$$

where $\Phi = [s_1, s_2 \cdots, s_K] \in \mathbb{C}^{MN \times K}$ is the dictionary matrix, $A = [\alpha_1, \alpha_2, \cdots, \alpha_L] \in \mathbb{C}^{K \times L}$ denotes the sparse coefficient matrix, and $N = [n_1, n_2, \cdots, n_L] \in \mathbb{C}^{MN \times L}$ represents the noise matrix.

The angle-Doppler profile can be estimated by solving the following problem:

$$A = \arg \min_{A} \|A\|_{2,0} \quad s.t. \|X - \Phi A\|_F^2 \leq \varepsilon \tag{9}$$

where ε is the noise error tolerance. Thus, the CNCM can be estimated by

$$\tilde{R}_{c+n} = \frac{1}{L} \sum_{l=1}^{L} \Phi \, diag(\alpha_l^2) \Phi^H + \sigma^2 I_{MN}. \tag{10}$$

2.2. Bayesian Model

From a Bayesian perspective, noise is assumed to be a zero-mean complex of Gaussian distribution. Then, the complex Gaussian likelihood for the $\mathbf{X}$ can be modeled as

$$p\left(\mathbf{X}\middle|\mathbf{A},\sigma^2\right) = \prod_{l=1}^{L}\mathcal{CN}\left(\mathbf{x}_l\middle|\mathbf{\Phi}\boldsymbol{\alpha}_l,\delta^{-1}\mathbf{I}_{MN}\right) \tag{11}$$

where $\delta = \sigma^{-2}$ is the noise precision. According to the conjugate prior principle [38], the δ follows a Gamma distribution,

$$p(\delta) = Gamma(\delta|c,d) \tag{12}$$

where $Gamma(\delta|c,d) = \frac{d^c}{\gamma(c)}\delta^{c-1}\exp(-d\delta)$, $\gamma(c)$ is the Gamma function and c and d represent the shape parameter and scale parameter.

To model the sparse coefficient, a complex Gaussian prior is employed first for each column of $\mathbf{A}$

$$p(\boldsymbol{\alpha}_l|\boldsymbol{\lambda}) = \prod_{k=1}^{K}\mathcal{CN}\left(a_{k,l}\middle|0,\lambda_k^{-1}\right) \tag{13}$$

where $\boldsymbol{\lambda} = [\lambda_1,\cdots,\lambda_k,\cdots,\lambda_K]^T$. A Gamma prior is assigned on the hyper-parameter λ_k

$$p(\lambda_k) = Gamma(\lambda_k|a,b) \tag{14}$$

In a word, the Bayesian model of the received signal has been established. Figure 1 shows the corresponding graphical model. Each node represents a conditional probability density.
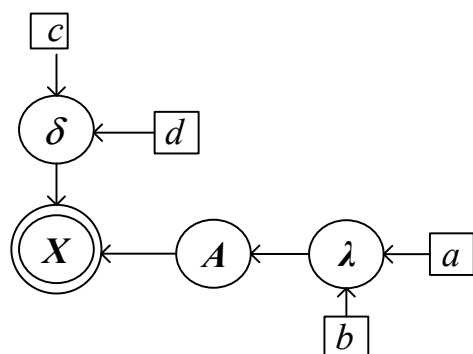

Figure 1. Bayesian graphical model.

3. Proposed Algorithm

In this section, a VBI-based STAP algorithm is derived for the MMV case. Then, a fast iterative structure is proposed to improve the computational efficiency of M-VBI-STAP.

3.1. VBI Algorithm

The set of hiding variables for the Bayesian model denotes $\Omega = [\mathbf{A},\boldsymbol{\lambda},\delta]$. The marginal likelihood function $p(\mathbf{X})$ needs to be calculated to obtain the closed form of the posterior distribution $p(\Omega|\mathbf{X})$. Since $p(\mathbf{X})$ is intractable, it can be decomposed as

$$\ln p(\mathbf{X}) = \mathcal{L}(q) + \mathcal{KL}(q|p) \tag{15}$$

with

$$\mathcal{L}(q) = \int q(\Omega)\ln\left(\frac{p(\mathbf{X},\Omega)}{q(\Omega)}\right)d\Omega \tag{16}$$

and

$$\mathcal{KL}(q|p) = -\int q(\Omega)\ln\left(\frac{p(\Omega|\mathbf{X})}{q(\Omega)}\right)d\Omega \tag{17}$$

where $q(\Omega)$ denotes a variational distribution, $\mathcal{L}(q)$ is the lower bound of $\ln p(X)$, and $\mathcal{KL}(q|p)$ represents the Kullback–Leibler divergence between the $p(\Omega|X)$ and $q(\Omega)$ regularization parameters. The VBI algorithm approximates the posterior distribution using the factorized variable distribution. Assuming the variables are independent of each other, the factorized variable distribution can be expressed as

$$q(\Omega) = q(X)q(\boldsymbol{\lambda})q(\delta). \tag{18}$$

Maximizing $\mathcal{L}(q)$, the optimal factorized variable distribution is

$$\ln q^*(\Omega) = \langle \ln p(X,\Omega)\rangle_{q(\Omega\backslash\Omega_i)} + \text{const} \tag{19}$$

where $\Omega\backslash\Omega_i$ means removing Ω_i from Ω, and

$$p(X,\Omega) = p(X|A,\boldsymbol{\lambda},\delta)p(A|\boldsymbol{\lambda})p(\boldsymbol{\lambda})p(\delta) \tag{20}$$

Combining Equations (19) and (20), the following gives the iterative update procedure of the variational parameters:

(1)   update A

$$
\begin{aligned}
\ln q^*(A) &= \langle \ln p(X|A,\boldsymbol{\lambda},\delta) + \ln p(A|\boldsymbol{\lambda})\rangle_{q(\Omega\backslash X)} + \text{const} \\
&= -\delta\sum_{l=1}^{L}\|x_l - \boldsymbol{\Phi}\boldsymbol{\alpha}_l\|_2^2 + \sum_{l=1}^{L}\boldsymbol{\alpha}_l^H diag(\boldsymbol{\lambda})\boldsymbol{\alpha}_l + \text{const}
\end{aligned} \tag{21}
$$

It can be found that $q^*(A)$ obeys the Gaussian distribution

$$q^*(A) = \prod_{l=1}^{L}\mathcal{CN}(\boldsymbol{\alpha}_l|\boldsymbol{\mu}_l,\boldsymbol{\Sigma}) \tag{22}$$

where

$$\boldsymbol{\mu}_l = \delta\boldsymbol{\Sigma}\boldsymbol{\Phi}^H x_l \tag{23}$$

$$\boldsymbol{\Sigma} = diag(\boldsymbol{\lambda}) - diag(\boldsymbol{\lambda})\boldsymbol{\Phi}^H\left(\delta\boldsymbol{I}_{MN} + \boldsymbol{\Phi}diag(\boldsymbol{\lambda})\boldsymbol{\Phi}^H\right)^{-1}\boldsymbol{\Phi}diag(\boldsymbol{\lambda}) \tag{24}$$

(2)   Update $\boldsymbol{\lambda}$

$$
\begin{aligned}
\ln q^*(\boldsymbol{\lambda}) &= \langle \ln p(A|\boldsymbol{\lambda}) + \ln p(\boldsymbol{\lambda})\rangle_{q(\Omega\backslash\lambda)} + \text{const} \\
&= \sum_{k=1}^{K}\left[(a+L-1)\ln\lambda_k - \left(b + \sum_{l=1}^{L}\left\langle a_{k,l}a_{k,l}^H\right\rangle\right)\lambda_k\right]
\end{aligned} \tag{25}
$$

From Equation (25), $q^*(\boldsymbol{\lambda})$ follows Gamma distribution and λ_k can be obtained by

$$\lambda_k = \frac{a+L}{b + \sum_{l=1}^{L}\left\langle a_{k,l}a_{k,l}^H\right\rangle} \tag{26}$$

where $\left\langle a_{k,l}a_{k,l}^H\right\rangle = \mu_{k,l}^2 + \Sigma_{k,k}$, $\mu_{k,l}$ is the k-th element of $\boldsymbol{\mu}_l$ and $\Sigma_{k,k}$ denotes the k-th main diagonal element of $\boldsymbol{\Sigma}$.

(3)   Update δ:

$$
\begin{aligned}
\ln q^*(\delta) &= \langle \ln\ln p(X|A,\boldsymbol{\lambda},\delta) + \ln p(\delta)\rangle_{q(\Omega\backslash\delta)} + \text{const} \\
&= (MNL + c - 1)\ln\delta - \left(\left\langle\sum_{l=1}^{L}\|x_l - \boldsymbol{\Phi}\boldsymbol{\alpha}_l\|_2^2\right\rangle + d\right)\delta
\end{aligned} \tag{27}
$$

Then, the result of δ is

$$\delta = \frac{MNL + c}{\|\boldsymbol{X} - \boldsymbol{\Phi\mu}\|_F^2 + trace(\boldsymbol{\Sigma\Phi}^H\boldsymbol{\Phi}) + d} \tag{28}$$

When the maximum number of iterations or the predefined stop condition is reached, the iterative procedure is terminated. Finally, the CNCM can be estimated by

$$\tilde{\boldsymbol{R}}_{c+n} = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\|\mu_{k,l}\|^2\boldsymbol{s}_k\boldsymbol{s}_k^H + \sigma^2\boldsymbol{I}_{MN}. \tag{29}$$

For clarification, the steps of the M-VBI-STAP algorithm are described in Algorithm 1.

Algorithm 1: Pseudocode of the M-VBI.

---

step 1: Input: data X and dictionary $\boldsymbol{\Phi}$;
step 2: Initialize: $a = b = c = d = 10^{-6}$, $\boldsymbol{\lambda}^{(0)} = \boldsymbol{1}$;
step 3: While if it does not converge:
　　1. Calculate $\boldsymbol{\mu}_l$, and $\boldsymbol{\Sigma}$ by Equations (23) and (24);
　　2. Calculate $\boldsymbol{\lambda}$ by Equation (26);
　　3. Calculate δ by Equation (28);
　　End
step 4: Obtain CNCM $\tilde{\boldsymbol{R}}_{c+n}$ by Equation (29);
step 5: Output: the STAP weight.

3.2. Variational Fast Solution

In a previous subsection, the updated criteria for the variational parameters were derived. However, the real-time performance of the algorithm is also crucial due to the matrix inverse operation in Equation (24). In addition, the overcomplete dictionaries used in the traditional SR-STAP algorithms also affect the computational efficiency. Here, a fast algorithm to reduce the computational complexity is introduced by analyzing the behavior of the distribution $q^*(\boldsymbol{\lambda})$. The high dimensional matrix inverse operation and a full dictionary are not needed in the fast algorithm.

The $\mu_{k,l}^2 + \Sigma_{k,k}$ in Equation (26) can be rewritten as $\boldsymbol{e}_k^H(\boldsymbol{\mu}_l\boldsymbol{\mu}_l^H + \boldsymbol{\Sigma})\boldsymbol{e}_k$, where $\boldsymbol{e}_k$ is the k-th column of the unit matrix with the dimension K. By noting $\boldsymbol{\mu}_l\boldsymbol{\mu}_l^H = \delta^2\boldsymbol{\Sigma\Phi}^H\boldsymbol{x}_l\boldsymbol{x}_l^H\boldsymbol{\Phi\Sigma}^H$ and $diag(\boldsymbol{\lambda}) = \sum_{k=1}^{K}\lambda_k\boldsymbol{e}_k\boldsymbol{e}_k^H$, Equation (24) is equivalent to

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \left(\delta\boldsymbol{\Phi}^H\boldsymbol{\Phi} + \sum_{j\neq k}\lambda_k\boldsymbol{e}_k\boldsymbol{e}_k^H + \lambda_j\boldsymbol{e}_j\boldsymbol{e}_j^H\right)^{-1} \\
&= \overline{\boldsymbol{\Sigma}}_j - \frac{\overline{\boldsymbol{\Sigma}}_j\boldsymbol{e}_j\boldsymbol{e}_j^H\overline{\boldsymbol{\Sigma}}_j}{\lambda_j^{-1} + \boldsymbol{e}_j^H\overline{\boldsymbol{\Sigma}}_j\boldsymbol{e}_j}
\end{aligned} \tag{30}
$$

where the Woodbury Matrix Identity is used in the second expression, and $\overline{\boldsymbol{\Sigma}}_j = \left(\delta\boldsymbol{\Phi}^H\boldsymbol{\Phi} + \sum_{j\neq k}\lambda_k\boldsymbol{e}_k\boldsymbol{e}_k^H\right)^{-1}$.

Defining $\varsigma_j = \boldsymbol{e}_j^H\overline{\boldsymbol{\Sigma}}_j\boldsymbol{e}_j$ and $\rho_{j,l} = \delta\boldsymbol{e}_j^H\overline{\boldsymbol{\Sigma}}_j\boldsymbol{\Phi}^H\boldsymbol{x}_l$, the modified version of Equation (26) can be obtained

$$\lambda_j^{(t+1)} = \frac{(a+L)\left(1 + \lambda_j^{(t)}\varsigma_j\right)}{b + \sum_{l=1}^{L}\left(\rho_{j,l}^2 + \varsigma_j + \lambda_j^{(t)}\varsigma_j^2\right)} \tag{31}$$

where t is the number of iterations. For brevity, when setting $a = b = 0$, Equation (31) can be seen as nonlinear maps $\lambda_j^{(t+1)} = \Gamma\left(\lambda_j^{(t)}\right)$. Calculating the stationary points of this map [39], the optimal λ_j^* is

$$\lambda_j^* = \begin{cases} L\left[\sum\limits_{l=1}^{L}\left(\rho_{j,l}^2 - \varsigma_j\right)\right]^{-1}, & \sum\limits_{l=1}^{L}\left(\rho_{j,l}^2 - \varsigma_j\right) > 0 \\ \infty, & \sum\limits_{l=1}^{L}\left(\rho_{j,l}^2 - \varsigma_j\right) \leq 0 \end{cases} \tag{32}$$

This means that:

(1)　If s_j is excluded from the model and $\sum\limits_{l=1}^{L}\left(\rho_{j,l}^2 - \varsigma_j\right) > 0$, add s_j to the model;

(2)　If s_j is in the model and $\sum\limits_{l=1}^{L}\left(\rho_{j,l}^2 - \varsigma_j\right) > 0$, re-estimate λ_j;

(3)　If s_j is in the model and $\sum\limits_{l=1}^{L}\left(\rho_{j,l}^2 - \varsigma_j\right) \leq 0$, remove s_j from the model.

To efficiently compute ς_j and $\rho_{j,l}$, the dictionary and sparse coefficient are set as $\mathbf{\Psi}^{(0)} \in \varnothing$, $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$. Thus, the $\overline{\mathbf{\Sigma}}_j$ can be written as

$$\overline{\mathbf{\Sigma}}_j = \begin{pmatrix} \delta\mathbf{\Psi}^H\mathbf{\Psi} + diag(\boldsymbol{\alpha}) & \delta\mathbf{\Psi}^H s_j \\ \delta s_j^H\mathbf{\Psi} & \delta s_j^H s_j \end{pmatrix}^{-1} \tag{33}$$

Using the block matrix inversion, Equation (35) is equivalent to

$$\overline{\mathbf{\Sigma}}_j = \begin{pmatrix} \mathbf{\Sigma} + \gamma_j\delta^2\mathbf{\Sigma}\mathbf{\Psi}^H s_j s_j^H\mathbf{\Psi}\mathbf{\Sigma} & -\gamma_j\delta\mathbf{\Sigma}\mathbf{\Psi}^H s_j \\ -\gamma_j\delta s_j^H\mathbf{\Psi}\mathbf{\Sigma} & \gamma_j \end{pmatrix} \tag{34}$$

where $\gamma_j = \left(\delta s_j^H s_j - \delta^2 s_j^H\mathbf{\Psi}\mathbf{\Sigma}\mathbf{\Psi}^H s_j\right)^{-1}$, $\mathbf{\Sigma} = \left(\delta\mathbf{\Psi}^H\mathbf{\Psi} + diag(\boldsymbol{\alpha})\right)^{-1}$ is the variance and the mean is $\boldsymbol{\mu}_l = \delta\mathbf{\Sigma}\mathbf{\Psi}^H x_l$.

Then, ς_j and $\rho_{j,l}$ can be updated as

$$\varsigma_j = \left(\delta s_j^H s_j - \delta^2 s_j^H\mathbf{\Psi}\mathbf{\Sigma}\mathbf{\Psi}^H s_j\right)^{-1} \tag{35}$$

$$\rho_{j,l} = \delta\varsigma_j s_j^H x_l - \delta^2\varsigma_j s_j^H\mathbf{\Psi}\mathbf{\Sigma}\mathbf{\Psi}^H x_l \tag{36}$$

The computation complexity of sequential selecting all dictionary atoms is still large. Inspired by the marginal likelihood maximization algorithm [40], a function $f(\lambda_j)$ can be constructed which has a unique maximum value for λ_j. Combining Equations (25) and (31), we have

$$f(\lambda_j) = L\ln\lambda_j - \sum\limits_{l=1}^{L}\left(\rho_{j,l}^2 - \varsigma_j\right)\lambda_k \tag{37}$$

In the $(t+1)$ iteration, we maximize the incremental $f(\lambda_j)$ to select a candidate atom. The serial number z is obtained by

$$z = \underset{j,1\leq j\leq N_s N_d}{\arg\max}\left[f\left(\left(\lambda_j^*\right)^{(t+1)}\right) - f\left(\left(\lambda_j^*\right)^{(t)}\right)\right] \tag{38}$$

Utilizing Equation (38), the computational complexity decreases because the number of the selected atoms is smaller than the size of dictionary. In summary, an iterative process to update parameter λ_j can be obtained. The update formulas are detailed in Appendix A. The pseudocode of the proposed M-F-VBI algorithm is described in Algorithm 2.

Algorithm 2: Pseudocode of the M-F-VBI.

step 1: Input: data X and dictionary $\boldsymbol{\Phi}$;
step 2: Initialize: $\boldsymbol{\lambda}^{(0)} = \boldsymbol{0}$, $\boldsymbol{\Psi}^{(0)} = \varnothing$, and $\boldsymbol{E}^0 = \boldsymbol{I}$;
step 3: While, if it does not converge:
    1. Calculate all λ_j^*, $\forall j$ by (31) and choose the atomic index by (38);

    2. If s_j is in the model and $\sum_{l=1}^{L} \left(\rho_{j,l}^2 - \varsigma_j \right) \le 0$, use (44)–(47) to update parameters; If s_j is in

the model and $\sum_{l=1}^{L} \left(\rho_{j,l}^2 - \varsigma_j \right) > 0$, use (48)–(51) to update parameters; If s_j is excluded from the

model and $\sum_{l=1}^{L} \left(\rho_{j,l}^2 - \varsigma_j \right) > 0$, use (53)–(56) to update parameters.
    3. Update δ by (28).
    End
step 4: Obtain CNCM $\tilde{\boldsymbol{R}}_{c+n}$ by (29);
step 5: Output: the STAP weight.

4. Computational Complexity Analysis

The computational complexity of the proposed algorithm is investigated via the number of complex multiplications, and the lower-order terms of the multiplication are omitted. In the proposed M-FVBI-STAP algorithm, the computational complexity is related to the dimension of $\boldsymbol{\Psi} \in \mathbb{C}^{MN \times P}$. P is the number of selected atoms that is assumed to be twice the rank of the clutter. The most computational part of M-VBI-STAP is Equation (24). Table 1 shows the computational complexity of different approaches. T_{SBL}, T_{FOCUSS}, T_{VBI} and T_{FVBI} represent the number of iterations of M-FOCUSS-STAP, M-SBL-STAP [24], M-VBI-STAP and M-FVBI-STAP. The results are plotted in Figure 2. As the number of system DOFs increases, it is observed that the computational complexity of the M-VBI-STAP is lower than M-SBL-STAP and M-FOCUSS-STAP. Compared to M-VBI-STAP, M-FVBI-STAP can improve computational efficiency.

Table 1. Computational complexity.

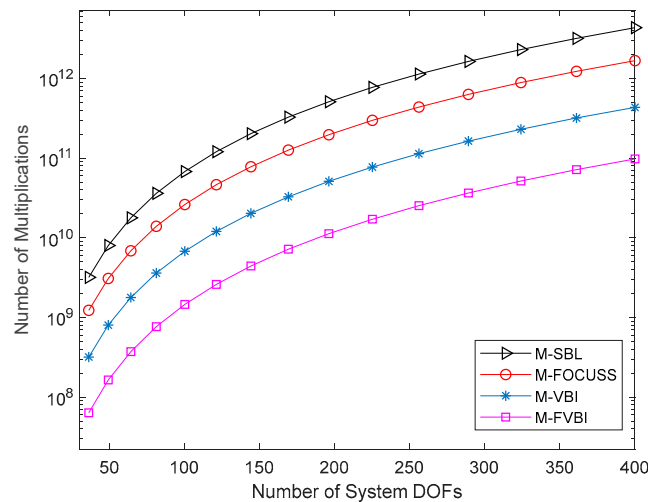| Algorithm	Computational Load
M-SBL	$o\left(\left((MN)^3 + 2K(MN)^2 + K^2 MN + KMNL \right) T_{SBL} \right)$
M-FOCUSS	$o\left(\left((MN)^3 + 2(MN)^2 K + MNLK \right) T_{FOCUSS} \right)$
M-VBI	$o\left(\left((MN)^3 + 2K(MN)^2 + K^2 MN + KMNL \right) T_{VBI} \right)$
M-FVBI	$o\left((MNK + MNL + 4MNKP + MNLP) T_{FVBI} \right)$

Figure 2. Computational complexity versus the number of system DOFs.

5. Numerical Experiments

5.1. Simulated Data

In this section, numerical experiments are conducted to illustrate the STAP and target detection performance of the proposed algorithm on simulated data. The main parameters of the simulations are listed in Table 2 and $\rho_s = \rho_d = 5$. For comparison, M-SBL-STAP and M-FOCUSS-STAP are presented. In addition, the optimal STAP with known CNCM and diagonal loading SMI-STAP (LSMI-STAP) are also included. The results are averaged over 100 Monte Carlo trials, and the number of training samples used in all experiments is ten. To measure the performance of clutter suppression, the improvement factor (IF) and signal-to-interference plus noise ratio (SINR) loss are used, which are denoted as

$$IF = \frac{\left|w^H a_t\right|}{w^H R_{c+n} w} \frac{tr(R_{c+n})}{a_t^H a_t} \tag{39}$$

$$L_{SINR} = \frac{\sigma^2}{MN} \frac{\left|w^H a_t\right|^2}{w^H R_{c+n} w} \tag{40}$$

Table 2. Parameters of radar system.

Parameter	Value	Unit
Number of elements	8	/
Number of pulses	8	/
Wavelength	0.3	m
Bandwidth	2.5	MHz
Height of platform	150	m/s
Velocity of platform	9000	m
Pulse repetition frequency (PRF)	2000	Hz
Clutter to noise ratio (CNR)	40	dB

The STAP performance in terms of Capon spectra is first investigated. Figure 3a plots the optimal spectrum with known CNCM. From Figure 3b–f, it can be noted that the estimation results of the proposed M-FVBI, M-VBI algorithm, and M-SBL-STAP are close to the optimal spectrum. In addition, the spectra of M-FOCUSS-STAP have an extension in the main-lobe region. It can also be seen that the spectrum is inaccurate when estimated by LSMI-STAP due to the lack of training samples. Figure 4 plots the Capon spectra with gain–phase (GP) error, which are set as 0.05 and 2°. Compared to Figure 3a, it can be seen that the results of the proposed M-FVBI, M-VBI algorithm, and M-SBL-STAP are close to the optimal spectrum.
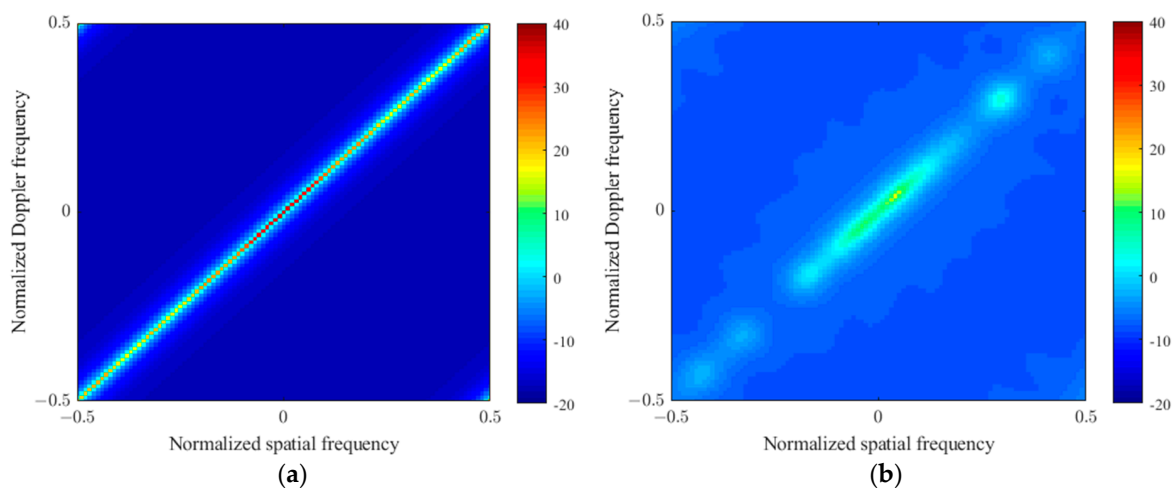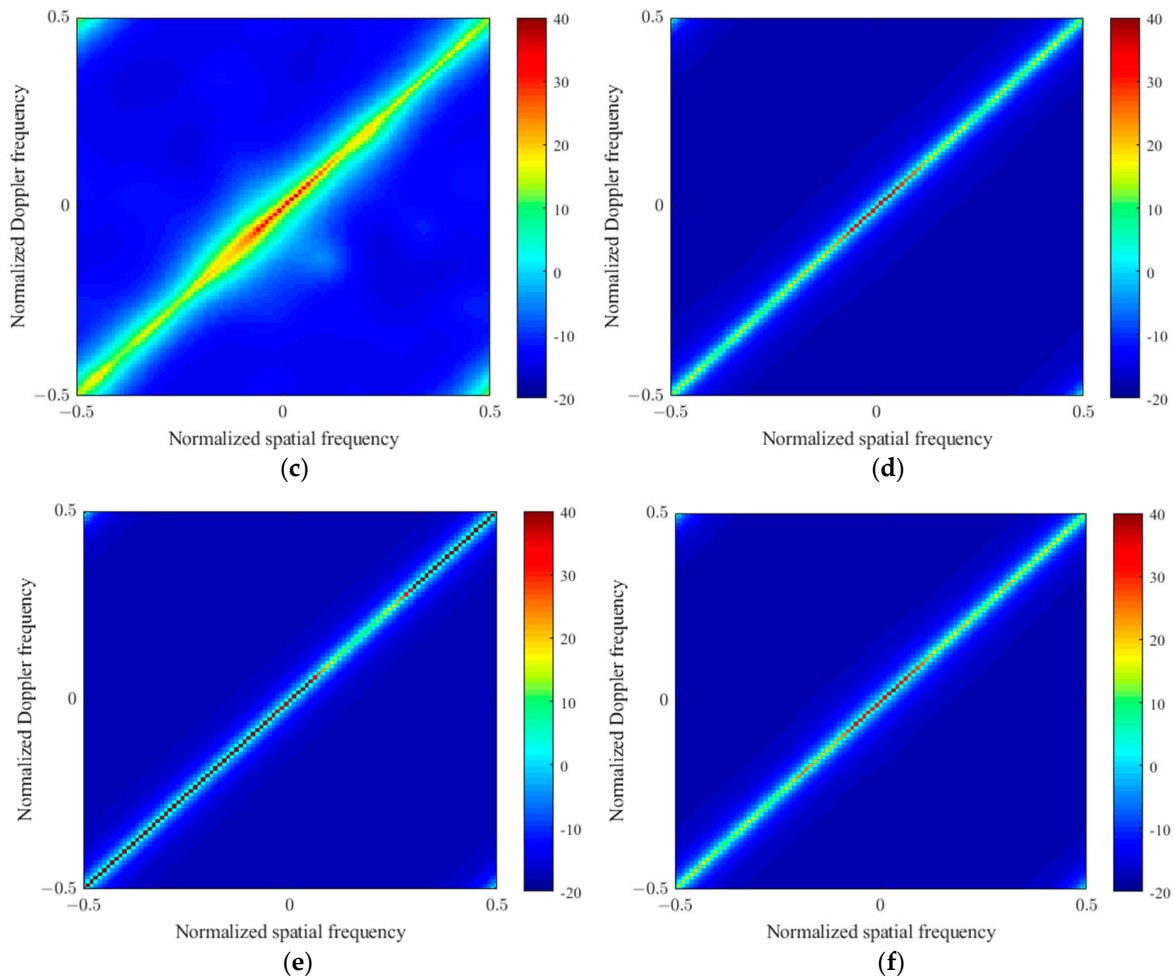

(a)

(b)

Figure 3. *Cont.*

Figure 3. Capon spectra of different approaches. (**a**) Capon spectrum estimated by known CNCM; (**b**) Capon spectrum estimated by LSMI-STAP; (**c**) Capon spectrum estimated by M-FOCUSS-STAP; (**d**) Capon spectrum estimated by M-SBL-STAP; (**e**) Capon spectrum estimated by M-VBI-STAP; (**f**) Capon spectrum estimated by the M-FVBI-STAP.

Then, the IF curves of different approaches against the normalized Doppler frequency in the main beam direction are presented in Figure 5. It is shown in Figure 5 that M-FVBI-STAP, M-VBI-STAP, and M-SBL-STAP have a similar performance, which is near the optimal performance. In the main-lobe area, the curves of the M-FOCUSS-STAP are broader than others.

For the next experiment, the average SINR loss curves versus the number of training samples are plotted in Figure 6, which were calculated as the mean SINR loss curves for $f_d = (-0.5, -0.1) \cup (0.1, 0.5)$. It can be seen from Figure 6 that all the approaches achieved a relatively stable performance when the number of training samples was more than three, and the stable performance of M-FOCUSS-STAP was worse than others. The M-VBI-STAP and M-FVBI-STAP had the best performance.

For a clearer illustration, the target detection performance of the proposed algorithm was accessed. According to [41], the probability of detection (PD) versus the probability of false alarm (PFA) is defined as

$$PD = PFA^{\frac{1}{1+\xi}} \tag{41}$$

where ξ is the output signal-to-clutter-plus-noise ratio (SCNR). We set the target located in the main beam direction.
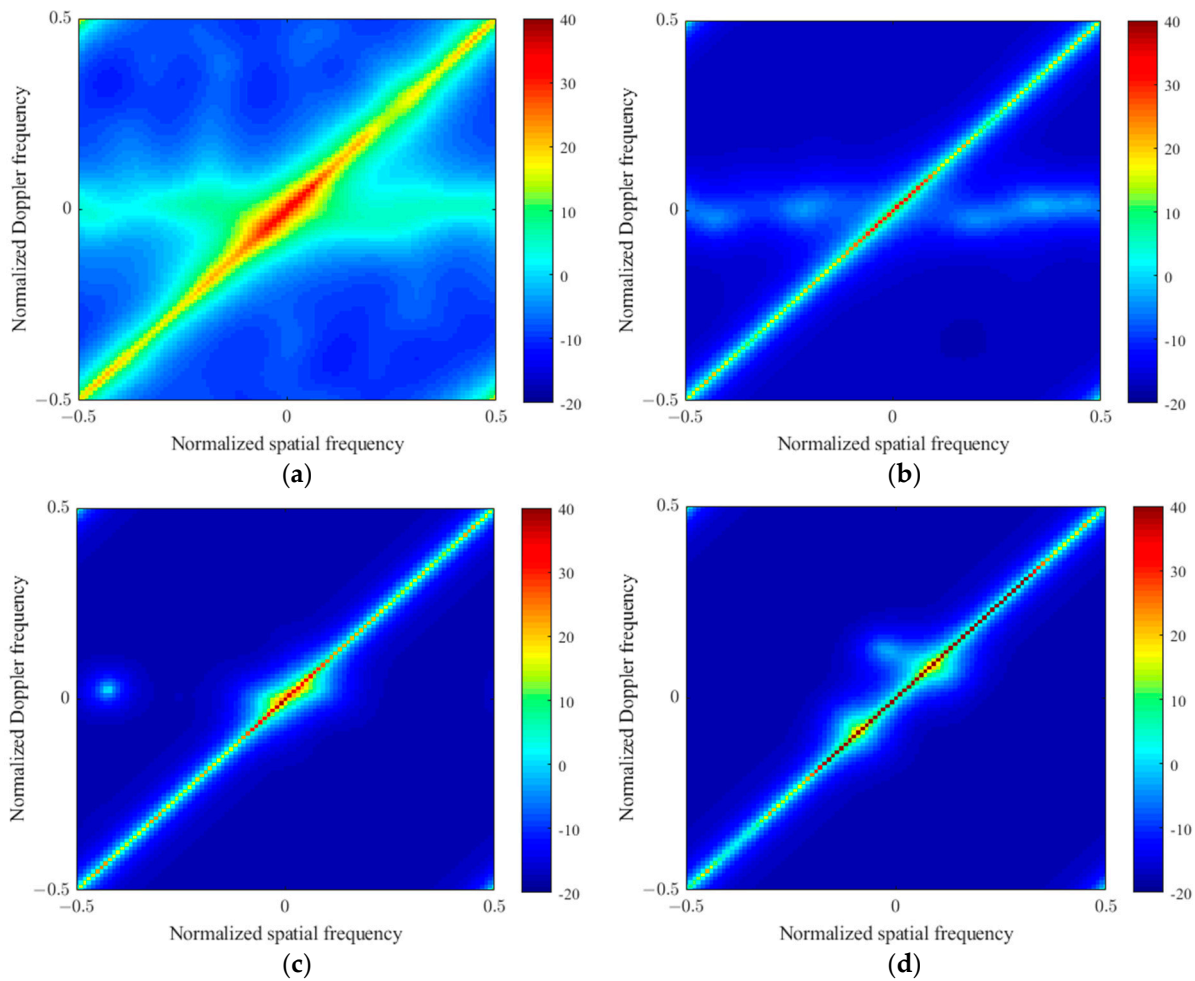
Figure 4. Capon spectra of different approaches with GP error. (**a**) Capon spectrum estimated by M-FOCUSS-STAP; (**b**) Capon spectrum estimated by M-SBL-STAP; (**c**) Capon spectrum estimated by M-VBI-STAP; (**d**) Capon spectrum estimated by the M-FVBI-STAP.
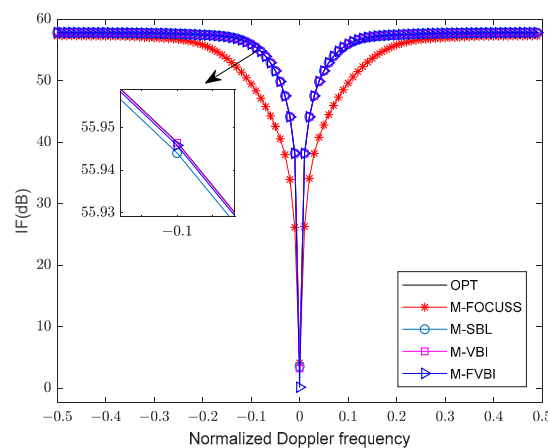

Figure 5. IF against the normalized Doppler frequency.

As shown in Figure 7, the receiver operating characteristics (ROC) (i.e., PD versus PFA) are plotted with the target signal-to-noise ratio (SNR), 0 dB and −4 dB, respectively. The normalized Doppler frequency is $f_d = 0.1$. It can be seen that the detection performances of the M-SBL-STAP, M-VBI-STAP, and M-FVBI-STAP are close to optimal. Another notable result is that the detection performances increased with the growth of the SNR value. The

ROC results with $f_d = 0.3$ are shown in Figure 8. The detection performance increased compared to $f_d = 0.1$. In both cases, M-VBI-STAP and M-FVBI-STAP are considered optimal.
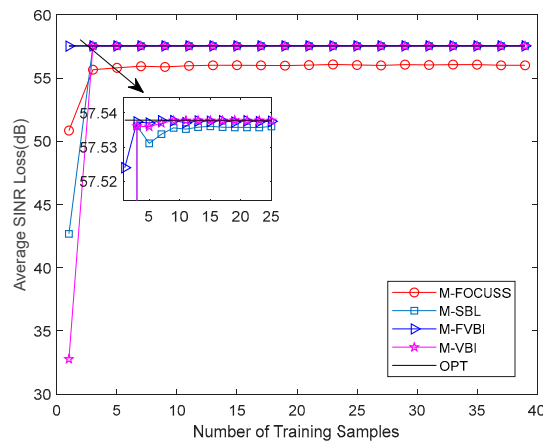

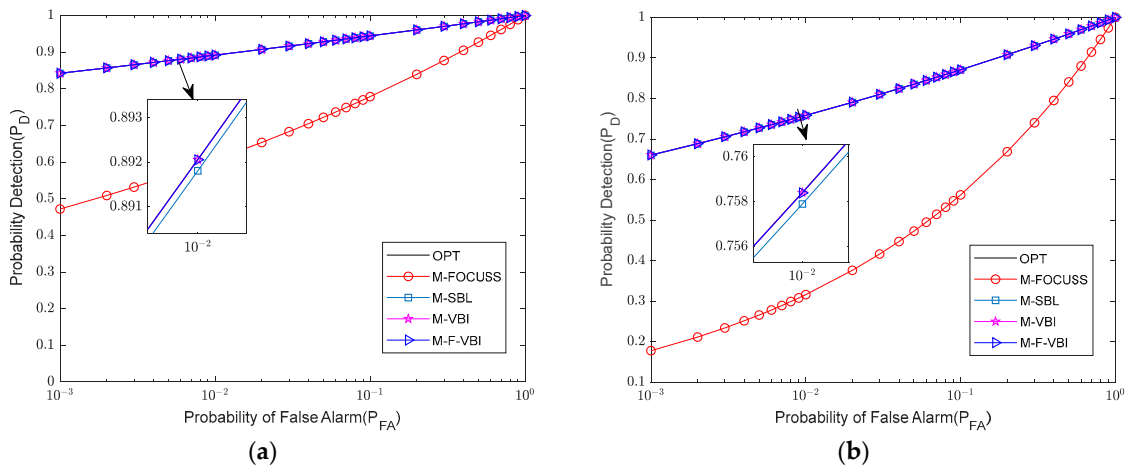**Figure 6.** Average IF against the number of training samples.

Figure 7. Receiver operating characteristics with $f_d = 0.1$. (**a**) SNR = 0 dB; (**b**) SNR = −4 dB.
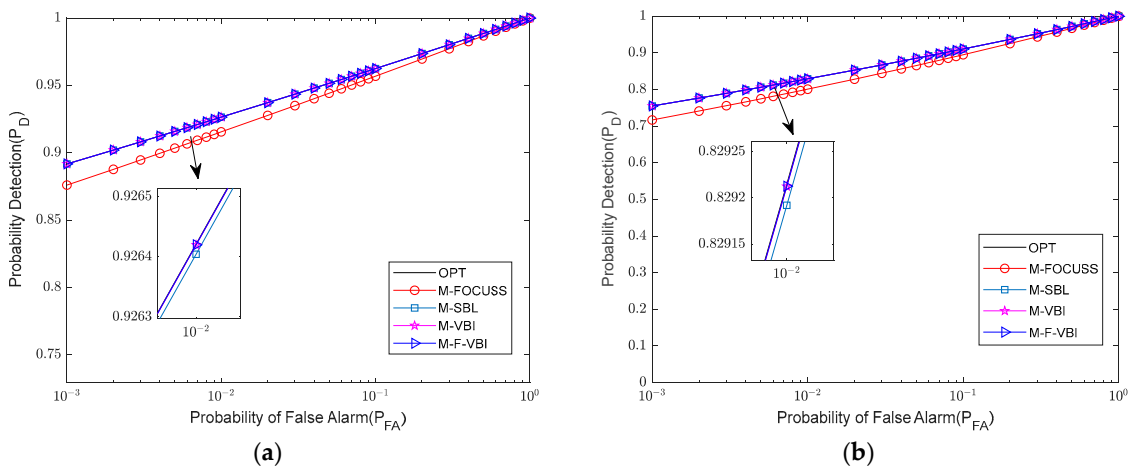

Figure 8. Receiver operating characteristics with $f_d = 0.3$. (**a**) SNR = 0 dB; (**b**) SNR = −4 dB.

Finally, the average running time for different cases is summarized in Table 3. The experiments were conducted by MATLAB 2017b on the computer equipped with Intel(R) Xeon(R) E5-2620 CPU @ 2.10Hz 2.10Hz and 64G RAM. It was found that the average

running time of M-VBI-STAP was less than M-SBL-STAP. M-FVBI-STAP significantly accelerated the average running time.

Table 3. The average running time (second).

Algorithms	Running Time
M-SBL	34.88 s
M-FOCUSS	11.16 s
M-VBI	3.70 s
M-FVBI	0.05 s

5.2. Measured Data

In this subsection, the performance of the proposed algorithm is accessed with publicly available data from Multi-Channel Airborne Radar Measurement (MCARM) data [42]. The airborne radar of the MCARM can be viewed as side looking. In CPI data, the number of coherent pluses, array elements, and range cells are 128, 22, and 630, respectively. For 12 pluses, data from the first space channels are selected for experimental analysis. There is a target added in the 320th range cell. Since the first 150 range cells cannot be used, the clutter power spectrum shown in Figure 9 was estimated using ten training samples selected from ye 151st range cell to the 630th range cell. It can be seen that all the approaches can recover the clutter ridge. The spectra of the M-FOCUSS-STAP have an extension in the main-lobe region. In addition, M-FVBI-STAP estimated CNCM by selecting the atoms. Thus, the clutter ridge of the M-FVBI-STAP is concentrated on a diagonal line.
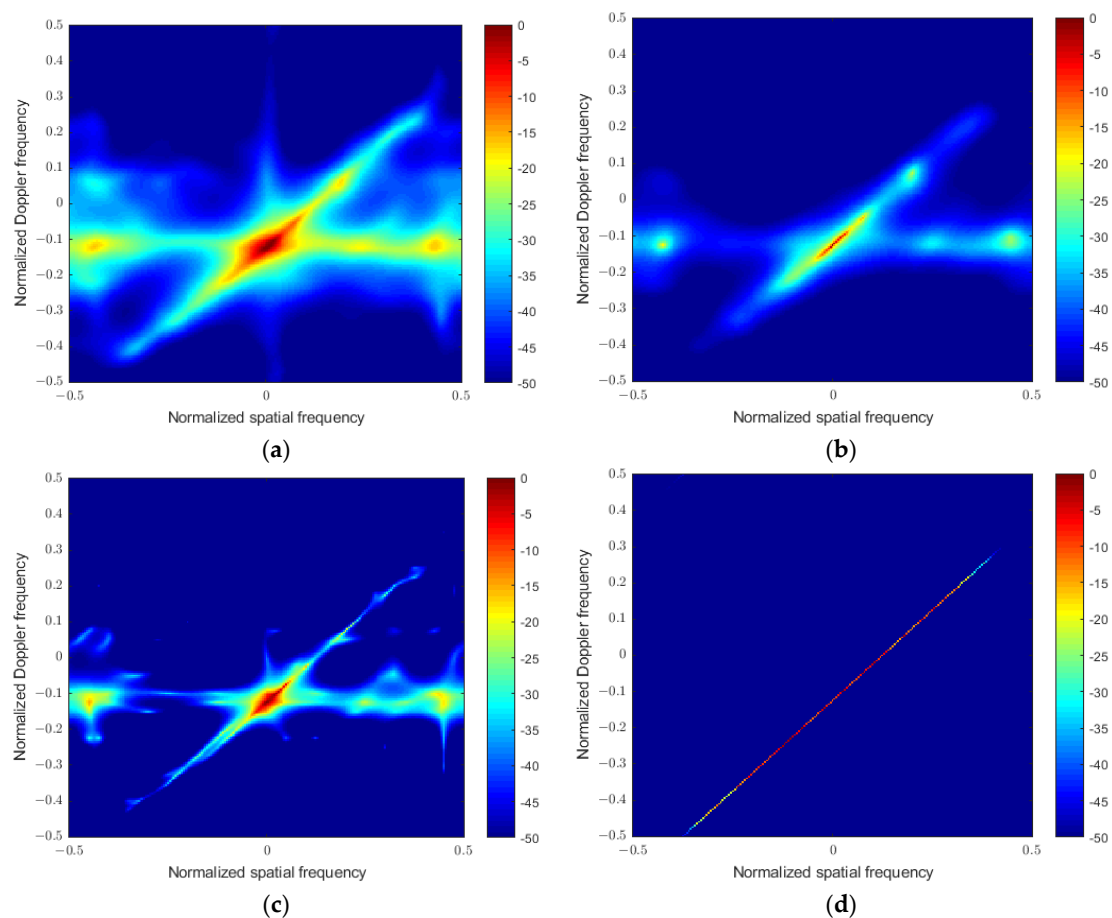

Figure 9. Estimated clutter power spectra of different approaches. (**a**) Capon spectrum estimated by M-FOCUSS-STAP; (**b**) Capon spectrum estimated by M-SBL-STAP; (**c**) Capon spectrum estimated by M-VBI-STAP; (**d**) Capon spectrum estimated by the M-FVBI-STAP.

The STAP output power from the 280th range cell to the 360th range cell is depicted in Figure 10. The generalized inner product (GIP) method [43] was used to select ten range cells of the twenty range cells around the CUT as the training samples. Additionally, four range cells on each side of the CUT were chosen as the guard cells. It can be observed that all algorithms clearly detected the target. To further illustrate the detection performance, the output signal to clutter plus noise ratio (SCNR) was calculated at the target range cell. The result corresponding to the algorithms is listed in Table 4. The proposed algorithm clearly had the best detection performance of all approaches.
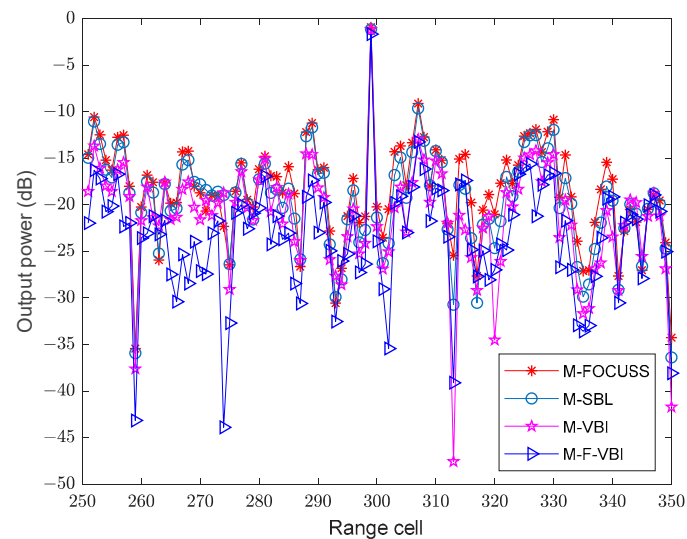

Figure 10. STAP output power versus range cells.

Table 4. The output SCNR (dB).

Algorithms	Results
M-SBL	16.43 dB
M-FOCUSS	16.89 dB
M-VBI	18.86 dB
M-FVBI	19.58 dB

6. Conclusions

In this paper, the framework of VBI into the STAP algorithm has been developed, which can improve the convergence speed compared to SBL. Due to the heavy computational burden caused by the matrix inverse operation, a fast iterative algorithm was derived. The numerical results of the simulated data and measured data demonstrate that the proposed algorithm has superior STAP and target detection performance and can effectively reduce computational complexity. VBI still belongs to SR algorithms, which require dictionary construction. When the grid mismatch exists, the performance of VBI decreases. The direction for future research is mitigating the grid mismatch effect on SR-STAP.

Author Contributions: Conceptualization, T.W.; methodology, X.Z.; validation, X.Z. and D.W.; writing—original draft preparation, X.Z.; writing—review and editing, D.W.; visualization, X.Z. and D.W.; supervision, T.W.; project administration, T.W. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Appendix A

(1)   Update formulas for removing the atoms.

We define the $\overline{\Sigma}_j$ as

$$\overline{\Sigma}_j = \begin{pmatrix} \delta\mathbf{\Psi}_{\bar{j}}^H\mathbf{\Psi}_{\bar{j}} + diag(\alpha_{\bar{j}}) & \delta\mathbf{\Psi}_{\bar{j}}^H s_j \\ \delta s_j^H \mathbf{\Psi}_{\bar{j}} & \delta s_j^H s_j \end{pmatrix}^{-1} \tag{A1}$$

where $\mathbf{\Psi}_{\bar{j}}$ and $\alpha_{\bar{j}}$ are the dictionary and sparse coefficient by removing s_j and λ_j. Using the block matrix inversion, Equation (A5) is equivalent to

$$\overline{\Sigma}_j = \begin{pmatrix} \tilde{\Sigma} + \gamma_j\delta^2\tilde{\Sigma}\mathbf{\Psi}_{\bar{j}}^H s_j s_j^H \mathbf{\Psi}_{\bar{j}}\tilde{\Sigma} & -\gamma_j\delta\tilde{\Sigma}\mathbf{\Psi}_{\bar{j}}^H s_j \\ -\gamma_j\delta s_j^H \mathbf{\Psi}_{\bar{j}}\tilde{\Sigma} & \gamma_j \end{pmatrix} \tag{A2}$$

where $\gamma_j = \left(\delta s_j^H s_j - \delta^2 s_j^H \mathbf{\Psi}_{\bar{j}}\tilde{\Sigma}\mathbf{\Psi}_{\bar{j}}^H s_j \right)^{-1}$, $\tilde{\Sigma}$ represents the variance of the current model,

$$\tilde{\Sigma} = \left(\delta\mathbf{\Psi}_{\bar{j}}^H\mathbf{\Psi}_{\bar{j}} + diag\left(\alpha_{\bar{j}}\right) \right)^{-1}$$
$$= \left[\Sigma - \frac{\Sigma e_j e_j^H \Sigma}{e_j^H \Sigma e_j} \right]_{\bar{j}\bar{j}} \tag{A3}$$

and the mean of the model is

$$\tilde{\mu}_l = \left[\mu_l - \frac{\Sigma e_j e_j^H \mu_l}{e_j^H \Sigma e_j} \right]_{\bar{j}} \tag{A4}$$

Then, ς_j and $\rho_{j,l}$ can be updated as

$$\varsigma_j = \left(\delta s_j^H s_j - \delta^2 s_j^H \mathbf{\Psi}_{\bar{j}}\tilde{\Sigma}\mathbf{\Psi}_{\bar{j}}^H s_j \right)^{-1} \tag{A5}$$

$$\rho_{j,l} = \delta\varsigma_j s_j^H x_l - \delta^2\varsigma_j s_j^H \mathbf{\Psi}_{\bar{j}}\tilde{\Sigma}\mathbf{\Psi}_{\bar{j}}^H x_l \tag{A6}$$

(2)   Update the formulas of re-estimating atoms.

The variance of the current model is

$$\tilde{\Sigma} = \Sigma - \kappa\Sigma e_j e_j^H \Sigma \tag{A7}$$

where $\kappa = \frac{\lambda_j - e_j^H\alpha}{1 + \left(e_j^H\Sigma e_j\right)\left(\lambda_j - e_j^H\alpha\right)}$. The mean of the model is

$$\tilde{\mu}_l = \mu_l - \kappa e_j^H \Sigma e_j e_j^H \mu_l \tag{A8}$$

Then, ς_j and $\rho_{j,l}$ can be updated as

$$\varsigma_j = \left(\delta s_j^H s_j - \delta^2 s_j^H \mathbf{\Psi}\tilde{\Sigma}\mathbf{\Psi}^H s_j \right)^{-1} \tag{A9}$$

$$\rho_{j,l} = \delta\varsigma_j s_j^H x_l - \delta^2\varsigma_j s_j^H \mathbf{\Psi}\tilde{\Sigma}\mathbf{\Psi}^H x_l \tag{A10}$$

(3)   Update the formulas for adding atoms.

The variance of the current model can be defined as

$$\tilde{\mathbf{\Sigma}} = \begin{pmatrix} \delta \mathbf{\Psi}^H \mathbf{\Psi} + diag(\boldsymbol{\alpha}) & \delta \mathbf{\Psi}^H \boldsymbol{s}_j \\ \delta \boldsymbol{s}_j^H \mathbf{\Psi} & \delta \boldsymbol{s}_j^H \boldsymbol{s}_j \end{pmatrix}^{-1} \tag{A11}$$

where $\boldsymbol{s}_j \notin \mathbf{\Psi}$. Using block matrix inversion, Equation (A5) is equivalent to

$$\tilde{\mathbf{\Sigma}} = \begin{pmatrix} \mathbf{\Sigma} + \gamma_j \delta^2 \mathbf{\Sigma} \mathbf{\Psi}^H \boldsymbol{s}_j \boldsymbol{s}_j^H \mathbf{\Psi} \mathbf{\Sigma} & -\gamma_j \delta \mathbf{\Sigma} \mathbf{\Psi}^H \boldsymbol{s}_j \\ -\gamma_j \delta \boldsymbol{s}_j^H \mathbf{\Psi} \mathbf{\Sigma} & \gamma_j \end{pmatrix} \tag{A12}$$

where $\gamma_j = \left(\delta \boldsymbol{s}_j^H \boldsymbol{s}_j + \lambda_j - \delta^2 \boldsymbol{s}_j^H \mathbf{\Psi} \mathbf{\Sigma} \mathbf{\Psi}^H \boldsymbol{s}_j \right)^{-1}$. The mean of the model is

$$\tilde{\boldsymbol{\mu}}_l = \delta \tilde{\mathbf{\Sigma}} \mathbf{\Psi}_j^H \boldsymbol{x}_l \tag{A13}$$

where $\mathbf{\Psi}_j = \left[\mathbf{\Psi}, \boldsymbol{s}_j \right]$. Then, ς_j and $\rho_{j,l}$ can be updated as

$$\varsigma_j = \left(\delta \boldsymbol{s}_j^H \boldsymbol{s}_j + \lambda_j - \delta^2 \boldsymbol{s}_j^H \mathbf{\Psi}_j \tilde{\mathbf{\Sigma}} \mathbf{\Psi}_j^H \boldsymbol{s}_j \right)^{-1} \tag{A14}$$

$$\rho_{j,l} = \delta \varsigma_j \boldsymbol{s}_j^H \boldsymbol{x}_l - \delta^2 \varsigma_j \boldsymbol{s}_j^H \mathbf{\Psi}_j \tilde{\mathbf{\Sigma}} \mathbf{\Psi}_j^H \boldsymbol{x}_l \tag{A15}$$

References

1. Ward, J. *Space-Time Adaptive Processing for Airborne Radar*; Technical Report; MIT Lincoln Laboratory: Lexington, KY, USA, 1994.
2. Brennan, L.E.; Reed, L.S. Theory of adaptive radar. *IEEE Trans. Aerosp. Electron. Syst.* **1973**, *9*, 237–251. [CrossRef]
3. Smith, S.T. Adaptive radar. In *Wiley Encyclopedia of Electrical and Electronics Engineering*; Wiley: Hoboken, NJ, USA, 2001.
4. Melvin, W.L. Space-time adaptive processing and adaptive arrays: Special collection of papers. *IEEE Trans. Aerosp. Electron. Syst.* **2000**, *36*, 508509. [CrossRef]
5. Klemm, R. *Principles of Space-Time Adaptive Processing*; The Institution of Electrical Engineers: London, UK, 2006.
6. Reed, L.S.; Mallet, J.D.; Brennan, L.E. Rapid convergence rate in adaptive arrays. *IEEE Trans. Aerosp. Electron. Syst.* **1974**, *10*, 853–863. [CrossRef]
7. Pallotta, L.; Farina, A.; Smith, S.T.; Giunta, G. Phase-only space-time adaptive processing. *IEEE Access* **2021**, *9*, 147250–147263. [CrossRef]
8. Haimovich, A.M.; Bar-Ness, Y. An eigenanalysis interference canceller. *IEEE Trans. Signal Process.* **1991**, *39*, 76–84. [CrossRef]
9. Hua, Y.; Nikpour, M.; Stoica, P. Optimal reduced rank estimation and filtering. *IEEE Trans. Signal Process.* **2001**, *49*, 457–469.
10. Goldstein, J.S.; Reed, I.S.; Scharf, L.L. A multistage representation of the Wiener filter based on orthogonal projections. *IEEE Trans. Inf. Theory* **1998**, *44*, 2943–2959. [CrossRef]
11. Haimovich, A. The eigencanceler: Adaptive radar by eigenanalysis methods. *IEEE Trans. Aerosp. Electron. Syst.* **1996**, *32*, 532–542. [CrossRef]
12. Wang, H.; Cai, L. On adaptive spatial-temporal processing for airborne surveillance radar systems. *IEEE Trans. Aerosp. Electron. Syst.* **1994**, *30*, 660–670. [CrossRef]
13. Klemm, R. Adaptive airborne MTI: An auxiliary channel approach. *IEEE Proc. F Commun. Radar Signal Process.* **1987**, *134*, 269–276. [CrossRef]
14. DiPietro, R.C. Extended factored space-time processing for airborne radar systems. *Signals Syst. Comput.* **1992**, *1*, 425–430.
15. Zhang, W.; He, Z.; Li, J.; Liu, H.; Sun, Y. A method for finding best channels in beam-space post-Doppler reduced-dimension STAP. *IEEE Trans. Aerosp. Electron. Syst.* **2014**, *50*, 254–264. [CrossRef]
16. Lin, X.; Blum, R.S. Robust STAP algorithms using prior knowledge for airborne radar applications. *Signal Process.* **1999**, *79*, 273–287. [CrossRef]
17. Guerci, J.R.; Baranoski, E.J. Knowledge-aided adaptive radar at DARPA: An overview. *IEEE Signal Process.* **2006**, *23*, 41–50. [CrossRef]
18. Gerlach, K.; Picciolo, M.L. Airborne/spacebased radar STAP using a structured covariance matrix. *IEEE Trans. Aerosp. Electron. Syst.* **2003**, *39*, 269–281. [CrossRef]
19. Steiner, M.; Gerlach, K. Fast converging adaptive processor or a structured covariance matrix. *IEEE Trans. Aerosp. Electron. Syst.* **2000**, *36*, 1115–1126. [CrossRef]

20. Kang, B.; Monga, V.; Rangaswamy, M. Rank-constrained maximum likelihood estimation of structured covariance matrices. *IEEE Trans. Aerosp. Electron. Syst.* **2014**, *50*, 501–515. [CrossRef]

21. De Maio, A.; De Nicola, S.; Landi, L.; Farina, A. Knowledge-aided covariance matrix estimation: A MAXDET approach. *IET Radar Sonar Navig.* **2009**, *3*, 341–356. [CrossRef]

22. Aubry, A.; De Maio, A.; Pallotta, L. A geometric approach to covariance matrix estimation and its applications to radar problems. *IEEE Trans. Signal Process.* **2017**, *66*, 907–922. [CrossRef]

23. Sun, K.; Zhang, H.; Li, G.; Meng, H.D.; Wang, X.Q. A novel STAP algorithm using sparse recovery technique. *IEEE Int. Geosci. Remote Sens. Symp.* **2009**, *1*, 3761–3764.

24. Sun, K.; Meng, H.; Wang, Y. Direct data domain STAP using sparse representation of clutter spectrum. *Signal Process.* **2011**, *91*, 2222–2236. [CrossRef]

25. Yang, Z.; Li, X.; Wang, H.; Jiang, W. On clutter sparsity analysis in space–time adaptive processing airborne radar. *IEEE Geosci. Remote. Sens. Lett.* **2013**, *10*, 1214–1218. [CrossRef]

26. Cotter, S.F.; Rao, B.D.; Engan, K.; Kreutz-Delgado, K. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans. Signal Process.* **2005**, *53*, 2477–2488. [CrossRef]

27. Tropp, J.; Gilbert, A. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **2007**, *53*, 4655–4666. [CrossRef]

28. Gorodnitsky, I.; Rao, B. Sparse signal reconstruction from limited data using FOCUSS a re-weighted minimum norm algorithm. *IEEE Trans. Signal Process.* **1997**, *45*, 600–616. [CrossRef]

29. Wu, Q.; Zhang, Y.; Amin, M.G.; Himed, B. Space-time adaptive processing and motion parameter estimation in multistatic passive radar using sparse Bayesian learning. *IEEE Trans. Geosci. Remote. Sens.* **2016**, *54*, 944–957. [CrossRef]

30. Duan, K.Q.; Wang, Z.T.; Xie, W.C.; Chen, H.; Wang, Y.L. Sparsity-based STAP algorithm with multiple measurement vectors via sparse Bayesian learning strategy for airborne radar. *IET Signal Process.* **2017**, *11*, 544–553. [CrossRef]

31. Wang, Z.T.; Xie, W.C.; Duan, K.Q. Clutter suppression algorithm base on fast converging sparse Bayesian learning for airborne radar. *Signal Process.* **2017**, *130*, 159–168. [CrossRef]

32. Tzikas, D.; Likas, A.; Galatsanos, N. The variational approximation for Bayesian inference. *IEEE Signal Process.* **2008**, *25*, 131–146. [CrossRef]

33. Guo, K.; Zhang, L.; Li, Y.; Zhou, T.; Yin, J. Variational Bayesian Inference for DOA Estimation under Impulsive Noise and Non-uniform Noise. *IEEE Trans. Aerosp. Electron. Syst.* **2023**. [CrossRef]

34. Shutin, D.; Buchgraber, T.; Kulkarni, S.; Poor, H. Fast variational sparse Bayesian learning with automatic relevance determination for superimposed signals. *IEEE Trans. Signal Process.* **2011**, *95*, 6257–6261. [CrossRef]

35. Zhou, S.; Xu, A.; Tang, Y.; Shen, L. Fast Bayesian inference of reparameterized gamma process with random effects. *IEEE Trans. Reliab.* **2023**, *in press*. [CrossRef]

36. Turlapaty, A.C. Variational Bayesian estimation of statistical properties of composite gamma log-normal distribution. *IEEE Trans. Signal Process.* **2020**, *68*, 6481–6492. [CrossRef]

37. Yuan, W.; Wei, Z.; Yuan, J.; Ng, D.W.K. A Simple Variational Bayes Detector for Orthogonal Time Frequency Space (OTFS) Modulation. *IEEE Trans. Veh. Technol.* **2020**, *69*, 7976–7980. [CrossRef]

38. Bishop, C. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.

39. Thomas, B. *Variational Sparse Bayesian Learning: Centralized and Distributed Processing*; Graz University of Technology: Graz, Austria, 2013.

40. Tipping, M.; Faul, A. Fast marginal likelihood maximization for sparse Bayesian models. In Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Key West, FL, USA, 3–6 January 2003; pp. 276–283.

41. Robey, F.C.; Fuhrmann, D.; Kelly, E.; Nitzberg, R. A CFAR adaptive matched filter detector. *IEEE Trans. Aerosp. Electron. Syst.* **1992**, *28*, 208–216. [CrossRef]

42. Wu, Y.; Wang, T.; Wu, J. A knowledge-aided space time adaptive processing based on road network data. *J. Electron. Sci. Technol.* **2015**, *37*, 613–618.

43. Melvin, W.L.; Wicks, M.C. Improving practical space-time adaptive radar. In Proceedings of the 1997 IEEE National Radar Conference, Syracuse, NY, USA, 13–15 May 2005; pp. 48–53.