Time-Varying Autoregressive (TVAR) Models for Multiple Radar Observations

Yuri I. Abramovich, *Senior Member, IEEE*, Nicholas K. Spencer, and Michael D. E. Turley

***Abstract*—We consider the adaptive radar problem where the properties of the (nonstationary) clutter signals can be estimated using multiple observations of radar returns from a number of sufficiently homogeneous range/azimuth resolution cells. We derive a method for approximating an arbitrary Hermitian covariance matrix by a time-varying autoregressive model of order m, $\mathrm{TVAR}(m)$, that is based on the Dym–Gohberg band-matrix extension technique which gives the unique $\mathrm{TVAR}(m)$ model for any nondegenerate covariance matrix. We demonstrate that the Dym–Gohberg transformation of the sample covariance matrix gives the maximum-likelihood (ML) estimate of the $\mathrm{TVAR}(m)$ covariance matrix. We introduce an example of $\mathrm{TVAR}(m)$ clutter modeling for high-frequency over-the-horizon radar that demonstrates its practical importance.**

***Index Terms*—Adaptive processing, autoregressive models, nonstationary clutter, nonstationary interference, radar observations, time-varying.**

I. Introduction

ANTENNA-ARRAYED radars with coherent repetitive waveforms, such as pulse trains or continuous-wave linear frequency modulation (FM) [1], generate a well-known "data cube" of radar returns (time sequences) from a number of range and azimuth resolution cells [2]. Even in conventional (single antenna beam) search radars, radar returns from adjacent range cells give multiple clutter observations (radar signal backscattered by terrain or sea surface) that are typically treated as independent observations of the same clutter process. Traditionally, clutter returns have been modeled as observations of a *stationary* process, then, after Burg's famous work [3], maximum-entropy spectral estimation methods (MEMs) were developed for clutter modeling and adaptive estimation [4], [5]. While the ergodicity of a stationary model theoretically allows spectral estimation with a conventional single observation (one range data train collected over the coherent integration time [CIT]), for a long time researchers have been using adjacent range cells for identifying the stationary AR clutter model [5]. According to Haykin *et al.* [5], because of the discontinuity in time between the pertinent data segments, "the segments cannot be simply combined to produce one long data record to be used as input for the usual MEM algorithm," which made *spatial averaging* that "involves using data segments from adjacent resolution cells" [5] an important means to overcome the discontinuity problem. Note that the combined spatial and temporal averaging suggested in [5] requires both stationarity (in time) and homogeneity (in space) for the clutter returns. Modern radar applications cannot guarantee both these assumptions. For example, it is well known that for airborne radar STAP, different terrain segments give quite different clutter returns, and the homogeneity assumptions can only be applied for a few neighboring range cells [6].

In this paper, we consider applications to high-frequency (HF) over-the-horizon (OTH) radar (sky-wave) systems [1], where backscattered surface-clutter signals are ionospherically propagated. Due to a number of different physical phenomena (e.g., traveling ionospheric disturbances), the two-way propagation path varies during the CIT. Such phase-path variations are the same for neighboring terrain/sea range resolution cells, but cause a significant Doppler frequency modulation over the CIT. Hence the homogeneity assumption can still be applied for clutter model identification over a reasonably small number of adjacent resolution cells, but the stationarity assumption is definitely invalid.

From a statistical viewpoint, this observation model comprises T independent identically distributed (i.i.d.) N-variate "training" vectors $\boldsymbol{x}_j (j = 1, \ldots, T)$ that we assume have a complex circular Gaussian distribution $\boldsymbol{x}_j \sim \mathcal{CN}(0, R_0)$, where R_0 is a positive-definite (p.d.) Hermitian covariance matrix. (As usual, we use boldface lowercase symbols for vectors, and uppercase symbols for matrices.) Note that this observation model is typical for adaptive antenna applications, where $\boldsymbol{x}_j$ is the "snapshot" of data collected simultaneously across N sensors at time j. It is important to emphasize that within this multiple i.i.d. observation model, the ML estimate $\hat{R}$ of the true covariance matrix R_0 (which exists for $T \geq N$) *does not* require any underlying model of R_0. The general (unstructured) ML covariance matrix estimate [7] is simply

$$\hat{R} = \frac{1}{T} \sum_{j=1}^{T} \boldsymbol{x}_j \boldsymbol{x}_j^H, \quad T \geq N \tag{1}$$

which is a consistent estimate with complex Wishart distribution $\mathcal{CW}(T \geq N, N, R_0)$. The efficiency of this ML estimate in adaptive antenna or adaptive detection algorithms has been comprehensively investigated [8]–[10]. Unfortunately, the number of "sufficiently homogeneous" adjacent range cells that can be used for such spatial averaging does not exceed the dimension of the problem N in many radar applications. For example, OTH radars often have CITs of $N = 128$ or 256 repetition periods,

Manuscript received June 17, 2005; revised May 10, 2006. Some results of this paper were presented at the Asilomar Conference, Pacific Grove, CA, October 2005. The associate editor coordinating the review of this paper and approving it for publication was Dr. Ta-Hsin Li.

Y. I. Abramovich and M. D. E. Turley are with the Intelligence, Surveillance and Reconnaissance Division (ISRD), Defence Science and Technology Organisation (DSTO), Adelaide, Australia (e-mail: Yuri.Abramovich@dsto.defence.gov.au; Mike.Turley@dsto.defence.gov.au).

N. K. Spencer is with the Adelaide Research & Innovation Pty. Ltd. (ARI), Adelaide, Australia (e-mail: Nick.Spencer@adelaide.edu.au).

Digital Object Identifier 10.1109/TSP.2006.888064

with only 40–60 available range cells (say), and the number of resolution cells sufficiently homogeneous in range and azimuth can be even smaller. Under these circumstances, one approach is to adopt a parametric model for the nonstationary process, with a relatively small number of free parameters. Indeed, for fewer parameters than in an arbitrary Hermitian covariance matrix (which has N^2 real-valued degrees of freedom), we expect the required sample size T for accurate adaptive model identification to be less than for general ML estimation (1).

When dealing with a nonstationary N-variate random process that is specified by a (non-Toeplitz) Hermitian covariance matrix, it seems logical to consider the time-varying AR process of order m, $\mathrm{TVAR}(m)$, to be our parametric model. Since the stationary $\mathrm{AR}(m)$ model has been proven in [11] to be an appropriate description of stationary clutter, ionospheric phase-path variations and the associated FM will generalize the $\mathrm{AR}(m)$ model into a $\mathrm{TVAR}(m)$ one. In any case, the number of free parameters that describe an N-variate covariance matrix $R_N^{(m)}$ of a $\mathrm{TVAR}(m)$ process is significantly smaller than in the general case, and, therefore, even for a non-$\mathrm{TVAR}(m)$ process, $\mathrm{TVAR}(m)$ approximations should be considered for adaptive signal detection (once the losses associated with such an approximation have been analyzed).

Therefore, from a practical viewpoint, there are two major issues that need to be clarified. The first is not a statistical one; given some Hermitian covariance matrix R_0, we wish to find its $\mathrm{TVAR}(m)$ approximation and associated losses. In radar applications, we are typically interested in the signal-to-noise ratio (SNR) degradation (deterministic loss factor ρ and/or stochastic loss factor γ) due to adopting some approximate covariance matrix, such as $R_N^{(m)}$ instead of the true covariance matrix R_0 in the optimum filter, which gives us the deterministic loss factor

$$\rho = \frac{\left[\boldsymbol{s}^H \left[R_N^{(m)}\right]^{-1} \boldsymbol{s}\right]^2}{\boldsymbol{s}^H \left[R_N^{(m)}\right]^{-1} R_0 \left[R_N^{(m)}\right]^{-1} \boldsymbol{s}\boldsymbol{s}^H R_0^{-1}\boldsymbol{s}} \tag{2}$$

where $\boldsymbol{s}$ is the N-variate ("steering") vector of expected signal, and the optimum (Wiener) filter that extracts this signal from the interference described by R_0 is

$$\boldsymbol{w}_{opt} = R_0^{-1}\boldsymbol{s}. \tag{3}$$

The output SNR for an arbitrary filter $\boldsymbol{w}$ is

$$q(\boldsymbol{w}) = \frac{|\boldsymbol{w}^H\boldsymbol{s}|^2}{\boldsymbol{w}^H R_0 \boldsymbol{w}} \tag{4}$$

and for the model-based filter $\boldsymbol{w}_{TVAR} = [R_N^{(m)}]^{-1}\boldsymbol{s}$ we obtain the deterministic loss factor (2). Here the suitability of any model, such as $R_N^{(m)}$, depends not only on R_0, but also on the signal vector $\boldsymbol{s}$.

The second issue that needs to be clarified is a statistical one. It deals with (ML) covariance matrix estimation within the set of $\mathrm{TVAR}(m)$ matrices, given the $T(T < N)$ i.i.d. training samples

$$X \equiv [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T] \sim \mathcal{CN}_T(0, R_0), \quad T < N. \tag{5}$$

If such an estimate $R_N^{(m)}$ is available, then its quality in the adaptive filter (antenna) context is measured by the statistical properties of the random SNR loss factor with respect to the clairvoyant optimum filter (3) [8]

$$\gamma = \frac{\left[\boldsymbol{s}^H \left[\hat{R}_N^{(m)}\right]^{-1} \boldsymbol{s}\right]^2}{\boldsymbol{s}^H \left[\hat{R}_N^{(m)}\right]^{-1} R_0 \left[\hat{R}_N^{(m)}\right]^{-1} \boldsymbol{s}\boldsymbol{s}^H R_0^{-1}\boldsymbol{s}}. \tag{6}$$

When the true covariance matrix R_0 is indeed a $\mathrm{TVAR}(m)$ matrix, i.e.

$$R_0 = R_N^{(m)} \tag{7}$$

then the loss factor γ accounts for only the losses associated with the finite sample support T, since for any consistent estimate $\hat{R}_N^{(m)}$

$$\lim_{T\to\infty} \hat{R}_N^{(m)} = R_N^{(m)} \tag{8}$$

and so

$$\lim_{T\to\infty} \gamma \xrightarrow{p} 1. \tag{9}$$

Whereas the ML covariance matrix estimate is traditionally used as a substitute for the unknown exact covariance matrix for both adaptive filters [8] and adaptive detectors [9], [10], it is a fact that (from their performance viewpoint) ML estimates have not been theoretically justified, especially for small i.i.d. sample support $(T \approx N)$.

In adaptive antenna applications, it is typical to deal with "small-rank" covariance matrices, where R_0 can be written in the form

$$R_0 = \sigma_0^2 I_N + \sum_{i=1}^{n} \sigma_i^2 \boldsymbol{u}_i \boldsymbol{u}_i^H, \;\; n < N, \;\; \boldsymbol{u}_i^H \boldsymbol{u}_k = \delta_{ik}, \;\; \sigma_i^2 \gg \sigma_0^2 \tag{10}$$

where δ is the Kronecker symbol. Since the early 1980s, it has been known that "diagonal loading" of the sample matrix

$$\hat{R}_L = \alpha I_N + \hat{R}, \quad \sigma_0^2 < \alpha < \min_i \sigma_i^2 \tag{11}$$

gives a significant improvement in the "convergence rate," which is the T-dependent SNR loss factor (6). Moreover, in [12] it was proven that the probability density function (pdf) for γ (for $\mathcal{E}\{\hat{R}\} = R_0$) does not depend on the loading factor α for small-rank scenarios (10). In 1996, the optimal diagonal loading was suggested by Ledoit according to [13] to be a "shrinkage estimator" for covariance matrices, which is equivalent to finding the optimal linear shrinkage of eigenvalues. Of course, AR models for radar clutter [5], [11] are not necessarily small-rank (10), but a significant reduction in the number of free parameters describing the $\mathrm{TVAR}(m)$ covariance matrix is just another way of "shrinking toward structure" that stabilizes an unstructured covariance matrix estimate (1) [13]. Therefore, whereas the suitability of the ML criterion for covariance matrix estimation in adaptive radar processing is an important separate problem, in this paper we will consider ML estimation within the restricted class of nonstationary $\mathrm{TVAR}(m)$ models.

It is important to emphasize the essential difference between our problem with its multiple i.i.d. observations and most published studies on nonstationary processes which deal with a single sequence [14], [15]. It is obvious that any meaningful statistical analysis of a single nonstationary sequence is only possible if some additional restrictions or assumptions on the nature of the nonstationarity are imposed. In this way, the notion of a "locally stationary" process has been introduced in [14], with numerous following papers that explore this intuitively appealing idea of "slow nonstationarity" (e.g., [15]). In [15], a model with multiple observations of a *time-continuous* nonstationary process was considered, and a best-basis search algorithm proposed to convert the *covariance operator* estimation problem into the problem of estimating "a band or near diagonal matrix, although this condition is not required in the best basis search." In our discussion of multiple observations of a nonstationary *time series*, our restriction on the admissible covariance model to be within the class of $\mathrm{TVAR}(m)$ models already confines admissible (inverse) covariance matrices to a very sparse structure *without any additional assumptions* being made about the type of time variations. Of course, additional *valid* assumptions may improve the adaptive antenna properties if properly exploited. Yet in this paper we consider the generic $\mathrm{TVAR}(m)$ model for our nonstationary time series with T i.i.d. training samples (observations) of the same nonstationary process available for adaptive estimation of its covariance matrix. In fact, $\mathrm{TVAR}(m)$ models have been under quite intensive investigation for different applications, including acoustic (speech) processing [16]. In most of these applications, multiple (independent) observations of the same nonstationary process are not available, and so special assumptions regarding the nature of the $\mathrm{TVAR}(m)$ model are imposed.

In the first approach, the time-varying parameters are estimated using the dynamic model

$$a[n] = a[n-1] + \Delta a \tag{12}$$

where $a[n]$ are the parameters of the TVAR model, and n is discrete time $(n = 1, \ldots, N)$. In this case, the parameters $a[n]$ are updated depending on the utilized adaptive algorithm, such as steepest descent or the recursive least-squares [17]. In the second approach, the TVAR parameters are explicitly (*a priori*) defined as a linear combination of weighted time-dependent functions:

$$a[n] = \sum_k a_k f_k(n) \tag{13}$$

where a_k are the (estimated) weights, and $f_k(n)$ are some predefined time functions.

We can see that, in addition to the $\mathrm{TVAR}(m)$ model restriction, both methods enforce quite restrictive additional assumptions, and neither can directly generate a $\mathrm{TVAR}(m)$ model approximation $R_N^{(m)}$ of the given covariance matrix R_N for some nonstationary process. For multiple observations we do not need to impose any of these restrictions. In dealing with our first nonstatistical issue (2), we have to specify the necessary and sufficient conditions for an arbitrary Hermitian matrix to have a $\mathrm{TVAR}(m)$ approximation of a certain kind. For the second statistical issue (6), (7), we have to find the ML estimate of a $\mathrm{TVAR}(m)$ covariance matrix $\hat{R}_N^{(m)}$ given T multiple i.i.d. observations.

In this paper we address mainly the second "statistical" issue, driven by the problem of adaptive target detection masked by nonstationary interference (clutter), with a very limited sample size $T \ll N$. Our method is derived mainly by providing a "signal processing" interpretation to the analytical results on the "band method" for positive matrix and operator extensions. This method was presented in 1981 by Dym and Gohberg [18], and was further extended in [19]–[21].

This paper is organized as follows. In Section II, we compare the well-known properties of the stationary autoregressive model, $\mathrm{AR}(m)$, with those of the time-varying autoregressive model, $\mathrm{TVAR}(m)$. We show that these models share an important covariance matrix property, namely that the inverse of the Toeplitz covariance matrix in the stationary case and the inverse of the Hermitian covariance matrix in the nonstationary case are both band (also called banded) matrices with bandwidth $(2m+1)$. On the other hand, an important difference between the models is that the stability condition that is necessary for the stationary $\mathrm{AR}(m)$ model does not exist in the time-dependent case. We also reintroduce the Dym–Gohberg results on band matrix extensions (Theorems 1–3), which, given the $(2m+1)$-wide band of some Hermitian matrix, allow us to calculate its unique extension that has zeros outside this band in its inverse, under certain specified conditions. We demonstrate that, when applied to covariance matrices, this method can be considered a generalization of the famous Burg maximum-entropy extension of the $(2m+1)$-wide band of some Toeplitz covariance matrix.

Section III proves an important new result (Theorem 4): that for a given set of independent Gaussian observations (training samples), the ML covariance matrix estimate of a $\mathrm{TVAR}(m)$ model is specified by exactly the same equations as the Dym–Gohberg transformation of the sample covariance matrix. We demonstrate that the only necessary and sufficient condition for the ML estimate of a $\mathrm{TVAR}(m)$ covariance matrix to exist is that the number of training samples T exceeds the order m of the model $(T \geq m+1)$. We also derive results arising from Theorem 3 that deal with the properties of likelihood-ratio (LR) tests which determine whether or not a $\mathrm{TVAR}(m)$ covariance matrix is the covariance matrix of a given set of independent training samples (Theorems 5 and 6).

Section IV describes an example of our $\mathrm{TVAR}(m)$ modeling for HF OTH radar applications. We introduce some sea-clutter data collected by a surface-wave radar, whereby the radar and backscattered signals propagate in a surface (Norton) wave mode over the highly saline ocean surface [22], with no ionospheric propagation. Hence this clutter is true sea clutter, unlike that from a sky-wave radar, where returns are contaminated by ionospheric propagation. We demonstrate that this data can be accurately modeled by a stationary AR model with the relatively high order $m = 23$, and then use this model to represent ionospherically propagated sea clutter that is affected by phase-path variations during the collection period. We show that conventional Doppler processing fails to resolve a hypothetical target that has a similar Doppler frequency to the clutter, whereas adaptive filtering based on the $\mathrm{TVAR}(m)$ model can successfully detect such slow-moving targets.

Our summary and conclusions are presented in Section V.

II. Properties of $\mathrm{TVAR}(m)$ and $\mathrm{AR}(m)$ Models Relevant to the Matrix Extension Problem

Consider the observation of T i.i.d. N-variate complex Gaussian vectors

$$\boldsymbol{x}_j \equiv \left[x_1^{(j)}, \ldots, x_N^{(j)}\right]^T \sim \mathcal{CN}(0, R_0), \quad j = 1, \ldots, T \tag{14}$$

where

$$R_0 = \mathcal{E}\left\{\boldsymbol{x}_j \boldsymbol{x}_j^H\right\} = \mathcal{E}\left\{x_i^{(j)} x_k^{(j)*}\right\} \equiv \{r_{ik}\}_{i,k=1,\ldots,N} \tag{15}$$

and where $\mathcal{E}\{\cdot\}$ is the expectation operator. Let us compare the properties of a stationary $\mathrm{AR}(m)$ discrete-time process with those of a time-varying (nonstationary) $\mathrm{TVAR}(m)$ process. The vector $\boldsymbol{x} \equiv [x_1, \ldots, x_N]^T$ is an N-variate sample of a stationary $\mathrm{AR}(m)$ process if its elements satisfy [23]

$$x_t = -\sum_{k=1}^{m} a_k^* x_{t-k} + \eta_t \quad \text{for} \quad t = m+1, \ldots, N \tag{16}$$

where

$$\mathcal{E}\left\{\eta_p \eta_q^*\right\} = \sigma_0^2 \delta_{pq}, \quad a_0 = 1. \tag{17}$$

For any *stationary* discrete-time process, its N-variate covariance matrix is a non-negative definite Toeplitz matrix, since

$$\mathcal{E}\left\{x_i x_k^*\right\} = \mathcal{E}\left\{x_{i+p} x_{k+p}^*\right\} = t_{i-k} \tag{18}$$

where $\mathcal{T}_N \equiv \mathrm{Toep}[t_{i-k}]_{i,k=1,\ldots,N}$. For the $\mathrm{AR}(m)$ process (16), multiplying by x_{t-i}^* and taking expectations then leads to

$$\sum_{k=0}^{m} a_k^* t_{k-i} = \begin{cases} \sigma_0^2 & \text{for } i = 0 \\ 0 & \text{for } i > 0. \end{cases} \tag{19}$$

Since $t_{k-i} = t_{i-k}^*$, this equation can be presented in matrix form

$$\mathcal{T}_{m+1} \boldsymbol{a}_{m+1} = \sigma_0^2 \boldsymbol{e}_{m+1} \tag{20}$$

where

$$\mathcal{T}_{m+1} \equiv \mathrm{Toep}[t_{i-k}]_{i,k=0,\ldots,m}, \quad \boldsymbol{a}_{m+1} \equiv [1, a_1, \ldots, a_m]^T, \\ \boldsymbol{e}_{m+1} \equiv [1, 0, \ldots, 0]^T. \tag{21}$$

This equation is known as the m^{th} order Yule-Walker equation [23]. Since an $\mathrm{AR}(m)$ process can also be viewed as an $\mathrm{AR}(N-1)$ process with $N > m$ and $a_k = 0$ for $k > m$, it can be rewritten as [23]

$$\begin{bmatrix} t_0 & t_1^* & \cdots & t_{N-1}^* \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1^* \\ t_{N-1} & \cdots & t_1 & t_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_m \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_0^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{22}$$

This equation means that the vector $\sigma_0^2[1, a_1, \ldots, a_m, 0, \ldots, 0]^T$ is the first column of the inverse matrix, i.e.,

$$\sigma_0^{-2} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_m \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \left[\mathcal{T}_N^{(m)}\right]^{-1} \boldsymbol{e}_N. \tag{23}$$

The Gohberg–Semencul formula [24] can now be used to restore the entire inverse matrix $[\mathcal{T}_N^{(m)}]^{-1}$ using just its first column (or row). Specifically, the (non-Toeplitz) inverse matrix is

$$\left[\mathcal{T}_N^{(m)}\right]^{-1} = \sigma_0^{-2}(BB^H - CC^H) \tag{24}$$

where B is the N-variate lower-triangular Toeplitz matrix whose elements are

$$B_{ik} \equiv a_{i-k} \quad \text{for} \quad i, k = 1, \ldots, N, \quad i \geq k \tag{25}$$

with the definitions $a_0 = 1$ and $a_i = 0$ for $i > m$, and where C is the N-variate lower-triangular Toeplitz matrix

$$C_{ik} \equiv a_{N-i-k}^* \quad \text{for} \quad i, k = 1, \ldots, N, \quad i \geq k \tag{26}$$

with $a_N = 0$.

Incidentally, this unique restoration of the entire inverse matrix using just the vector (23) demonstrates that this inverse matrix is a $(2m+1)$-wide band matrix, *i.e.,* all its elements outside the band ($|i-i| > m$) are zero. This is a consequence of the Gohberg–Semencul formula (24) where both triangular matrices B and C are in fact $(m+1)$-wide band matrices for the $\mathrm{AR}(m)$ model.

Meanwhile, it is well-known [23] that the positive-definiteness of the general Toeplitz covariance matrix $\mathcal{T}_{p+1} \equiv \mathrm{Toep}[t_0, \ldots, t_p]$ is equivalent to the condition that the polynomial

$$w(z) \equiv \sum_{k=0}^{p} d_k z^k \neq 0 \quad \text{for} \quad |z| \leq 1 \tag{27}$$

i.e., does not have any zeros inside the unit disk. Here

$$\boldsymbol{d}_{p+1} \equiv [d_0 = 1, d_1, \ldots, d_p] = \frac{\mathcal{T}_{p+1}^{-1} \boldsymbol{e}_{p+1}}{\boldsymbol{e}_{p+1}^T \mathcal{T}_{p+1}^{-1} \boldsymbol{e}_{p+1}}. \tag{28}$$

For an $\mathrm{AR}(m)$ model with $a_i = 0$ for $i > m$, (27) is the same as the condition for a stable process, meaning that the covariance lag t_N approaches zero as $N \to \infty$.

It is an important fact that the N-variate Toeplitz matrix $\mathcal{T}_N^{(m)}$, whose band inverse is (23), is a solution to the famous Carathéodory moment problem [25].

> *For a given finite sequence $\{t_k\}_{k=0,\ldots,m}$ of covariance lags of some stationary random process with discrete time, find the set of all possible functions $D(z)$ of the*

Carathéodory class $\mathcal{D}$ (corresponding to the spectral functions $\sigma(\beta)$), such that

$$D(z) = \frac{d_0}{2} + \sum_{k=1}^{\infty} d_k z^k, \quad d_k = \int_{-\pi}^{\pi} \exp(-ik\beta) d\sigma(\beta) \quad (29)$$

and

$$d_k = t_k \quad \text{for} \quad k = 0, \ldots, m. \quad (30)$$

In terms of Toeplitz covariance matrix extension, the Carathéodory problem is to specify for $N > m$ all possible extensions (completions) to the central $(2m+1)$-wide band of this matrix, which is specified by the entries $t_0, \ldots, t_m$. The Carathéodory problem has a solution if and only if the $(m+1)$-variate Toeplitz matrix $\mathcal{T}_{m+1} \equiv \mathrm{Toep}[t_0, \ldots, t_m]$ is non-negative-definite. The solution is unique if and only if $\mathcal{T}_{m+1}$ is not invertible. Moreover, if $\mathcal{T}_{m+1} > 0$, then the problem has infinitely many solutions, and all of them are described by the Artemenko–Geronimus formula [25], [26]. Amongst all these solutions, there is a single one that, under the constraints of (30), maximizes the function

$$H\left(\mathcal{T}_N^{(m)}\right) = \log\left|\mathcal{T}_N^{(m)}\right| \quad (31)$$

which is proportional to the entropy of the distribution $\mathcal{CN}(0, \mathcal{T})$. For the given set of parameters $\{\sigma_0^2, a_1, \ldots, a_m\}$, this solution is specified by (20)–(26), and is known by the signal processing community as Burg's maximum-entropy (ME) solution. Therefore, the $\mathrm{AR}(m)$ model is the ME approximation of a stationary process with the given initial $(m+1)$ covariance lags (moments) $\{t_0, \ldots, t_m\}$. Equivalently, the ME matrix extension of the band matrix $\mathrm{Toep}[t_0, \ldots, t_m]$ is uniquely specified by the parameters $\{\sigma_0^2, a_1, \ldots, a_m\}$.

Finally, let us introduce another important representation of the p.d. Toeplitz matrix $\mathcal{T}_N^{(m)}$. For an arbitrary $(p+1)$-variate Hermitian Toeplitz covariance matrix $\mathcal{T}_{p+1} > 0$, the Yule-Walker equation can be written as [23]

$$\mathcal{T}_{p+1} \boldsymbol{a}_{p+1} = \sigma_p^2 \boldsymbol{e}_{p+1} \quad \text{for} \quad \sigma_p^2 > 0 \quad (32)$$

where

$$\boldsymbol{a}_{p+1} \equiv [a_{p0} = 1, a_{p1}, \ldots, a_{pp}]^T. \quad (33)$$

Then the predictor coefficients provide a decomposition of $\mathcal{T}_{p+1}$ as a product of lower triangular, diagonal, and upper triangular matrices (see [23, Theorem 2.13])

$$\mathcal{T}_{p+1} = A_{p+1}^{-1} \Sigma_{p+1} A_{p+1}^{-H} \quad (34)$$

where $\Sigma_{p+1} \equiv \mathrm{diag}[\sigma_0^2, \ldots, \sigma_p^2]$ and

$$A_{p+1} = \begin{bmatrix} a_{00} & 0 & \ldots & 0 \\ a_{11} & a_{10} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ a_{pp} & a_{p,p-1} & \ldots & a_{p0} \end{bmatrix}, \quad a_{i0} = 1 \quad \text{for} \quad i = 0, \ldots, p. \quad (35)$$

Note that for the $\mathrm{AR}(m)$ model with $m < p$, all solutions for (32) are specified by $(m+1)$ nonzero parameters, which means that the covariance matrix $\mathcal{T}_{p+1}^{(m)}$ is uniquely specified by the *all different* $m+1 \geq p$-variate central matrices of the matrix $\mathcal{T}_{p+1}^{(m)}$.

We have introduced these well-known properties of the $\mathrm{AR}(m)$ model to compare them with the $\mathrm{TVAR}(m)$ properties that are described below.

The $\mathrm{TVAR}(m)$ model for the N-variate vector $\boldsymbol{x} \equiv [x_1, \ldots, x_N]^T$ is defined similarly to (16), but with time-dependent AR coefficients $a_t(k)$:

$$x_t = -\sum_{k=1}^{m} a_t^*(k) x_{t-k} + \eta_t \quad \text{for} \quad t = m+1, \ldots, N \quad (36)$$

where

$$\mathcal{E}\{\eta_t \eta_s^*\} = \sigma_t^2 \delta_{ts}. \quad (37)$$

Again, multiplying by x_i^* and taking expectations gives us

$$\sum_{k=0}^{m} r_{i,t-k} a_t(k) = \sigma_t^2 \delta_{ti} \quad \text{for} \quad t = m+1, \ldots, N, \; i = 1, \ldots, N, \; a_t(0) = 1. \quad (38)$$

From this equation, we can directly derive the set of $(m+1)$ linear equations:

$$\begin{bmatrix} r_{tt} & r_{t,t-1} & \ldots & r_{t,t-m} \\ r_{t-1,t} & r_{t-1,t-1} & \ldots & r_{t-1,t-m} \\ \vdots & \vdots & & \vdots \\ r_{t-m,t} & r_{t-m,t-1} & \ldots & r_{t-m,t-m} \end{bmatrix} \begin{bmatrix} 1 \\ a_t(1) \\ \vdots \\ a_t(m) \end{bmatrix} = \sigma_t^2 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (39)$$

Since $a_t(0) = 1$, we may rewrite this as

$$J R_t^{(m)} J \boldsymbol{a}_t = \sigma_t^2 \boldsymbol{e}_{m+1} \quad (40)$$

where

$$R_t^{(m)} \equiv \begin{bmatrix} r_{t-m,t-m} & r_{t-m,t-m+1} & \ldots & r_{t-m,t} \\ r_{t-m+1,t-m} & r_{t-m+1,t-m+1} & \ldots & r_{t-m+1,t} \\ \vdots & \vdots & & \vdots \\ r_{t,t-m} & r_{t,t-m+1} & \ldots & r_{tt} \end{bmatrix} \quad (41)$$

with $t = m+1, \ldots, N$, and where J is the exchange matrix:

$$J \equiv \begin{bmatrix} 0 & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & 0 \end{bmatrix}. \quad (42)$$

If we let $\tilde{\boldsymbol{a}}_t \equiv J \boldsymbol{a}_t$ and $\tilde{\boldsymbol{e}}_{m+1} \equiv J \boldsymbol{e}_{m+1}$, then we get

$$R_t^{(m)} \tilde{\boldsymbol{a}}_t = \sigma_t^2 \tilde{\boldsymbol{e}}_{m+1}. \quad (43)$$

Note that $R_t^{(m)}$ is the $(m+1)$-variate central (diagonal) block of the covariance matrix $R_N^{(m)} = \mathcal{E}\{\boldsymbol{x}_i \boldsymbol{x}_i^H\}$. We can treat (43) as the $\mathrm{TVAR}(m)$ version of the m^{th} order Yule-Walker equation (20) which specifies all nonzero elements of the $\mathrm{TVAR}(m)$ model $a_t(k)$, $k = 1, \ldots, m;\; t = m+1, \ldots, N$.

Equation (39) can now be used to prove that the inverse of $R_N^{(m)}$ is a p.d. $(2m+1)$-wide band matrix. Let us introduce the N-variate $(m+1)$-wide band upper-triangular matrix A_N, similarly to [23, Theorem 2.13] (for the stationary case): [see (44) at the bottom of the page]. The first m-variate block is built from the vectors $\tilde{\boldsymbol{a}}_t$ that are calculated from (43) by

$$\tilde{\boldsymbol{a}}_t = \frac{\left[R_t^{(m)}\right]^{-1}\tilde{\boldsymbol{e}}_t}{\tilde{\boldsymbol{e}}_t^T\left[R_t^{(m)}\right]^{-1}\tilde{\boldsymbol{e}}_t} \quad \text{for} \quad t=1,\dots,m \tag{45}$$

(where σ is removed by the normalization). The remaining $(N-m)$ columns are formed from the $\tilde{\boldsymbol{a}}_t$ for $t=m+1,\dots,N$, augmented by zeros. Let us now investigate the product $R_N^{(m)}A_N$. According to (39) and (45), it is

$$R_N^{(m)}A_N = \left[\begin{array}{cccc|cccc} d_1 & 0 & \dots & 0 & 0 & \dots & \dots & 0 \\ z & d_2 & \ddots & \vdots & \vdots & & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & & \vdots \\ z & & \ddots & d_m & 0 & & & \vdots \\ z & \dots & \dots & z & \sigma_{m+1}^2 & \ddots & & \vdots \\ \vdots & & & \vdots & z & \ddots & \ddots & \vdots \\ \vdots & & & \vdots & \vdots & \ddots & \ddots & 0 \\ z & \dots & \dots & z & z & \dots & z & \sigma_N^2 \end{array}\right] \equiv ZD_N \tag{46}$$

where

$$d_t \equiv \frac{1}{\tilde{\boldsymbol{e}}_t^T\left[R_t^{(m)}\right]^{-1}\tilde{\boldsymbol{e}}_t} \quad \text{for} \quad t=1,\dots,m,$$
$$D_N \equiv \operatorname{diag}\left[d_1,\dots,d_m,\sigma_{m+1}^2,\dots,\sigma_N^2\right] \tag{47}$$

and where z denotes a (so-far) unspecified complex number, and Z is a lower-triangular matrix with ones on the diagonal and unspecified elements below it. Since $R_N^{(m)}$ is a Hermitian matrix, we have

$$R_N^{(m)} = ZD_NA_N^{-1} = A_N^{-H}D_NZ^H \tag{48}$$

so $Z^H = A_N^{-1}$, and finally

$$R_N^{(m)} = A_N^{-H}D_NA_N^{-1}, \quad \left[R_N^{(m)}\right]^{-1} = A_ND_N^{-1}A_N^H. \tag{49}$$

Since A_N is an $(m+1)$-wide band upper-triangular matrix, $[R_N^{(m)}]^{-1}$ is a $(2m+1)$-wide band Hermitian matrix. In fact, (49) is the well-known Cholesky decomposition of the p.d. matrix $R_N^{(m)}$, and its positive-definiteness is guaranteed by $D_N > 0$ regardless of A_N. Therefore no special conditions on the admissible values of $a_t(k)(k=1,\dots,m;\ t=m+1,\dots,N)$ are imposed for the generic $\mathrm{TVAR}(m)$ model. In particular, the p.d. requirement on the Toeplitz covariance matrix $\mathcal{T} \equiv \mathrm{Toep}[t_0,\dots,t_m]$ for the stationary $\mathrm{AR}(m)$ model (27) does not exist for the generic $\mathrm{TVAR}(m)$ model. In fact, since there exists a unique Cholesky decomposition for any p.d. Hermitian matrix, we have demonstrated that the single necessary and sufficient condition for a p.d. Hermitian matrix $R_N^{(m)}$ to be a covariance matrix of a generic $\mathrm{TVAR}(m)$ process is the condition

$$\left\{\left[R_N^{(m)}\right]^{-1}\right\}_{ik} = 0 \quad \text{for} \quad |i-k| > m \tag{50}$$

i.e., the elements of its inverse are zero outside the band. The fact that this condition does not have an associated stability condition analogous to (27) has a straightforward physical interpretation. A p.d. Toeplitz covariance matrix can be expanded to an arbitrary dimension, which means that the underlying AR model has to generate statistically the same signals as $T \to \infty$ (the stability condition). On the contrary, a nonstationary $\mathrm{TVAR}(m)$ process is observed over the particular time interval $t=1,\dots,N$, and *none* of the properties of this (generic) $\mathrm{TVAR}(m)$ process can be accurately extrapolated beyond this interval. Therefore, the "stability condition" is nonsensical for a generic TVAR process and is not required, but estimation of its properties (specifically the $\mathrm{TVAR}(m)$ coefficients $a_t(k)$) is possible *only* if multiple i.i.d. observations are available. For adaptive radar clutter estimation, this opportunity exists and is frequently exploited.

Now that the properties of a $\mathrm{TVAR}(m)$ covariance matrix $R_N^{(m)}$ have been specified, we may formulate the $\mathrm{TVAR}(m)$ approximation and $\mathrm{TVAR}(m)$ covariance matrix extension problems.

$$A_N = \left[\begin{array}{cccc|cccc} 1 & a_2(1) & \dots & a_m(m-1) & a_{m+1}(m) & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots & \vdots & a_{m+2}(m) & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_m(1) & \vdots & \vdots & \ddots & 0 \\ 0 & \dots & & 1 & a_{m+1}(1) & \vdots & & a_N(m) \\ 0 & \dots & \dots & 0 & 1 & a_{m+2}(1) & & \vdots \\ \vdots & & & \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & & & \vdots & \vdots & \ddots & \ddots & a_N(1) \\ 0 & \dots & \dots & 0 & 0 & \dots & 0 & 1 \end{array}\right]. \tag{44}$$

Definition 1: Suppose that a given N-variate Hermitian matrix $R_N \equiv \{r_{ik}\}$ is positive-definite, then the Dym–Gohberg $TVAR(m)$ approximation of R_N is the p.d. Hermitian matrix $R_N^{(m)}$ such that

$$\begin{cases} \left\{R_N^{(m)}\right\}_{ik} = r_{ik} & \text{for } |i-k| \le m \\ \left\{\left[R_N^{(m)}\right]^{-1}\right\}_{ik} = 0 & \text{for } |i-k| > m. \end{cases} \quad (51)$$

The solution to the problem of finding $R_N^{(m)}$ given R_N is provided by the following theorem that was first proven by H. Dym and I. Gohberg [18].

Theorem 1 [18]: Given an N-variate Hermitian matrix $R_N \equiv \{r_{ik}\}_{i,k=1,\ldots,N}$, suppose that

$$\begin{bmatrix} r_{ii} & \cdots & r_{i,i+m} \\ \vdots & \ddots & \vdots \\ r_{i+m,i} & \cdots & r_{i+m,i+m} \end{bmatrix} > 0$$
$$\text{for} \quad i = 1, \ldots, N-m; \quad |i-k| \le m. \quad (52)$$

For $q = 1, \ldots, N$, let

$$\begin{bmatrix} y_{qq} \\ \vdots \\ y_{L(q),q} \end{bmatrix} = \begin{bmatrix} r_{qq} & \cdots & r_{q,L(q)} \\ \vdots & \ddots & \vdots \\ r_{L(q),q} & \cdots & r_{L(q),L(q)} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (53)$$

and

$$\begin{bmatrix} z_{\Gamma(q),q} \\ \vdots \\ z_{qq} \end{bmatrix} = \begin{bmatrix} r_{\Gamma(q),\Gamma(q)} & \cdots & r_{\Gamma(q),q} \\ \vdots & \ddots & \vdots \\ r_{q,\Gamma(q)} & \cdots & r_{qq} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (54)$$

where $L(q) \equiv \min\{N, q+m\}$ and $\Gamma(q) \equiv \max\{1, q-m\}$. Furthermore, let the N-variate triangular matrices U and V be defined as

$$V_{ik} \equiv \begin{cases} y_{ik} y_{kk}^{-\frac{1}{2}} & \text{for} \quad k \le i \le L(k) \\ 0 & \text{otherwise} \end{cases} \quad (55)$$

$$U_{ik} \equiv \begin{cases} z_{ik} z_{kk}^{-\frac{1}{2}} & \text{for} \quad \Gamma(k) \le i \le k \\ 0 & \text{otherwise} \end{cases}. \quad (56)$$

Then the N-variate matrix given by

$$R_N^{(m)} \equiv (U^H)^{-1} U^{-1} = (V^H)^{-1} V^{-1} \quad (57)$$

is the unique p.d. Hermitian matrix that satisfies (51).

See [18] for the proof.

In fact, the Dym–Gohberg theorem now comes as no surprise, after we have introduced our "$\mathrm{TVAR}(m)$ Yule-Walker" equation (40), augmented by (45), and proven the decomposition (49).

Clearly, Theorem 1 gives a specific solution to the "nonstationary" Carathéodory problem, whereby we are given R_N "inside the band", *i.e.*, $\{r_{ik}\}_{i,k=1,\ldots,N}$, $|i-k| \le m$, rather than the stationary covariance lags $\{t_{i-k}\}_{i,k=0,\ldots,m}$. All possible solutions of this extension problem are described by the formula introduced in [18], [21] that could be treated as the nonstationary (time-varying) generalization of the Artemenko–Geronimus formula [26]. Obviously, (57) and (34) coincide for the stationary case (*i.e.*, $\mathrm{TVAR}(m) = \mathrm{AR}(m)$).

Recall that $\mathcal{T}_N^{(m)}$ in (23) is the specific ME (Burg) solution to the Carathéodory problem. The following theorem can be considered a $\mathrm{TVAR}(m)$ generalization of this result.

Theorem 2 [20], [21]: Given an N-variate Hermitian matrix $R_N \equiv \{r_{ik}\}_{i,k=1,\ldots,N}$ that satisfies (52), the band extension $R_N^{(m)}$ (53)–(57) is the unique positive extension of the band $\{r_{ik}\}_{|i-k|\le m}$ for which the determinant is maximal.

The next theorem specifies the necessary and sufficient conditions for a matrix R_N to have a positive $R_N^{(m)}$ approximation.

Theorem 3 [21]: Given an N-variate Hermitian matrix $R_N \equiv \{r_{ik}\}_{i,k=1,\ldots,N}$, the necessary and sufficient condition for a p.d. extension of the band $\{r_{ik}\}_{|i-k|\le m}$ to exist is that all the submatrices are p.d., i.e.,

$$\begin{bmatrix} r_{ii} & \cdots & r_{i,i+m} \\ \vdots & \ddots & \vdots \\ r_{i+m,i} & \cdots & r_{i+m,i+m} \end{bmatrix} > 0 \quad \text{for} \quad i = 1, \ldots, N-m. \quad (58)$$

While condition (58) is clearly similar to $\mathcal{T}_{m+1} \ge 0$ for the Carathéodory problem, the main distinction is that in the nonstationary (non-Toeplitz) case, no requirements on the zeros of the $\mathrm{TVAR}(m)$ polynomials

$$w_t(z) = \sum_{k=0}^{m} a_t(k) z^k, \quad a_t(0) = 1 \quad (59)$$

are made. Instead, we have the above p.d. condition on all $(N-m)$ $(m+1)$-variate different block matrices in R_N.

We define the $\mathrm{TVAR}(m)$ time-frequency function (time-frequency distribution) as [17]

$$F_t^{(m)}(w_t) = \frac{\sigma_t^2}{\left|\sum_{k=0}^{m} a_t(k) \exp(-i w_t k)\right|^2},$$
$$a_t(0) = 1, \quad \text{for} \quad t = m+1, \ldots, N \quad (60)$$

which serves as an obvious generalization for the stationary $\mathrm{AR}(m)$ process spectrum [23]

$$F^{(m)}(w) = \frac{\sigma_0^2}{\left|\sum_{k=0}^{m} a(k) \exp(-iwk)\right|^2}, \quad a(0) = 1. \quad (61)$$

Again, the time-frequency function $F_t^{(m)}(w_t)$ is specified for arbitrary $a_t(k)$, while (27) is required for $F^{(m)}(w)$ to be an $\mathrm{AR}(m)$ spectrum (see [27] for details).

Of course, the $\mathrm{AR}(m)$ "Burg spectrum" (61) is indeed the spectrum of a stationary process that may be estimated from a single (sufficiently long) observation. In this regard, (60) is not a *spectrum* as such, and may be calculated only for particular values of $a_t(k)$ estimated via multiple observations.

Finally, we mention that our use of the Dym–Gohberg approximation with the properties (51) is just one possible $\mathrm{TVAR}(m)$ approximation of a p.d. Hermitian matrix R_0 as far as the "nonstatistical criterion" (2) is concerned. For example, it is quite possible that small perturbations in the elements of the band $\{r_{ik}\}_{|i-k|\le m}$ would improve the SNR loss factor (2) for a given R_0 and $\boldsymbol{s}$, and similarly for the "statistical" loss factor (3). Yet, for ML covariance matrix estimation within the class of $\mathrm{TVAR}(m)$ matrices, the optimality of the Dym–Gohberg approximation is proven here.

III. ML $\mathrm{TVAR}(m)$ Model Identification

According to our multiple observation model (14), we consider T i.i.d. N-variate complex Gaussian vectors $X \equiv [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]$, so that the likelihood function (LF) can be introduced in the usual way [23]:

$$\mathcal{L}(R|X) = \frac{\exp\left[-\mathrm{tr}(R^{-1}XX^H)\right]}{\pi^T \det^T R}. \tag{62}$$

We need to find the maximum of the LF over the class of structured p.d. Hermitian matrices with

$$\{R^{-1}\}_{ik} = 0 \quad \text{for} \quad |i-k| > m \tag{63}$$

which according to (50) is the only necessary and sufficient condition for the matrix R to be a $\mathrm{TVAR}(m)$ matrix $R_N^{(m)}$. Let

$$C \equiv R^{-1}, \quad C_{ik} = 0 \quad \text{for} \quad |i-k| > m \tag{64}$$

then

$$\log \mathcal{L}(C|X) = T \log\det C - T\mathrm{tr} C\hat{R} \tag{65}$$

where

$$\hat{R} \equiv \frac{1}{T} XX^H \tag{66}$$

is the sample (direct data) covariance matrix. For $C_{ik} \neq 0$ (*i.e.*, $|i-k| \leq m$), we now need to solve the ML equation

$$\frac{\partial \mathcal{L}(C|X)}{\partial C_{ik}} = 0 \quad \text{for} \quad |i-k| \leq m. \tag{67}$$

Since only the C_{ik} for $|i-k| \leq m$ are subject to optimization, the ML equation is then

$$\begin{cases} \frac{\partial \log \mathcal{L}(C|X)}{\partial C_{ik}} = 0 & \text{for } |i-k| \leq m \\ C_{ik} = 0 & \text{for } |i-k| > m. \end{cases} \tag{68}$$

Since [28]

$$\frac{\partial \log\det C}{\partial C_{ki}} = \{C^{-1}\}_{ik}, \quad \frac{\partial \mathrm{tr}[C\hat{R}]}{\partial C_{ki}} = \hat{R}_{ik} \equiv \hat{r}_{ik} \tag{69}$$

we get the following ML equation:

$$\begin{cases} \left\{\hat{C}_{ML}^{-1}\right\}_{ik} = \hat{R}_{ik} & \text{for } |i-k| \leq m \\ \{\hat{C}_{ML}\}_{ik} = 0 & \text{for } |i-k| > m. \end{cases} \tag{70}$$

(We have introduced the notation $\hat{R}_{ik}$ here to emphasize its similarity to $\{\hat{C}^{-1}\}_{ik}$.) On the other hand, for $T \geq m+1$, all central $(m+1)$-variate blocks of the sample covariance matrix $\hat{R}$ are positive-definite (with probability one), i.e.,

$$\begin{bmatrix} \hat{r}_{ii} & \ldots & \hat{r}_{i,i+m} \\ \vdots & & \vdots \\ \hat{r}_{i+m,i} & \ldots & \hat{r}_{i+m,i+m} \end{bmatrix} > 0 \quad \text{for} \quad i = 1, \ldots, N-m. \tag{71}$$

Since this means that the condition for Theorem 3 is satisfied for $\hat{R}$, the unique Dym–Gohberg transformation $DG^{(m)}(\hat{R}) = \hat{R}_N^{(m)}$ such that

$$\begin{cases} \left\{\hat{R}_N^{(m)}\right\}_{ik} = \hat{R}_{ik} & \text{for } |i-k| \leq m \\ \left\{\left[\hat{R}_N^{(m)}\right]^{-1}\right\}_{ik} = 0 & \text{for } |i-k| > m \end{cases} \tag{72}$$

satisfies the ML condition (70), *i.e.*, $\hat{C}_{ML}^{-1} = \hat{R}_N^{(m)}$. In fact, we have just proven the following theorem.

Theorem 4: Let $X \sim \mathcal{CN}_T(0, R_N > 0)$ be T i.i.d. random vectors, then the ML estimate of the $TVAR(m)$ covariance matrix $\hat{R}_N^{(m)}$ is given by (72), provided that $T \geq m+1$.

We see that, unlike the stationary $\mathrm{AR}(m)$ model with its Toeplitz covariance matrix (with a band inverse), the Dym–Gohberg band-extension method gives a closed-form solution to the problem of finding the ML Hermitian covariance matrix estimate of a $\mathrm{TVAR}(m)$ model, analogously to Anderson's famous solution of the ML estimate of an arbitrary Hermitian matrix [7]. This result has a straightforward physical interpretation: since no special conditions are imposed on the set of admissible AR lags $a_t(k)$, they are selected in the way that retains the ML covariance lag estimates within the band $\{\hat{r}_{ik}\}_{|i-k|\leq m}$.

Corollary 1: The maximum of the LF over the class of $TVAR(m)$ models is

$$\max_{C_{ik}=0, |i-k|<m} \mathcal{L}(C|X) = \pi^{-T} \exp(-NT) \left[\prod_{q=1}^{N} \hat{y}_{qq}\right]^T \tag{73}$$

where

$$\hat{y}_{qq} \equiv \boldsymbol{e}_{L(q)-q}^T \begin{bmatrix} \hat{r}_{qq} & \ldots & \hat{r}_{q,L(q)} \\ \vdots & & \vdots \\ \hat{r}_{L(q),q} & \ldots & \hat{r}_{L(q),L(q)} \end{bmatrix}^{-1} \boldsymbol{e}_{L(q)-q}. \tag{74}$$

Indeed, for the ML $\mathrm{TVAR}(m)$ covariance matrix estimate $\hat{R}_N^{(m)}$ (72)

$$\sum_{|i-k|<m} \hat{R}_{ik} \left\{\left[\hat{R}_N^{(m)}\right]^{-1}\right\}_{ik} = \mathrm{tr}\hat{R}\hat{C}_{ML} = \mathrm{tr}\hat{R}_N^{(m)}\hat{C}_{ML} = N \tag{75}$$

since $\hat{C}_{ML}^{-1} = \hat{R}_N^{(m)}$. According to (57)

$$\det \hat{C}_{ML} = \det \hat{U}\hat{U}^H = \prod_{q=1}^{N} \hat{y}_{qq} \tag{76}$$

since U is a triangular matrix whose determinant is equal to the product of its diagonal elements. Therefore

$$\begin{aligned} \mathcal{L}(\hat{C}_{ML}|X) &= \pi^{-T} \exp\left[-\mathrm{tr}(\hat{C}_{ML}\hat{R})\right] \det{}^T \hat{C}_{ML} \\ &= \pi^{-T} \exp(-NT) \left[\prod_{q=1}^{N} \hat{y}_{qq}\right]^T. \end{aligned} \tag{77}$$

This corollary is important for proving our next theorem, which deals with the following hypothesis test. Suppose that for a given set of i.i.d. complex Gaussian observations $\boldsymbol{x}_j$ $(j = 1, \ldots, T)$ we wish to test the hypothesis

$$H_0: \quad \mathcal{E}\left\{\boldsymbol{x}_j \boldsymbol{x}_j^H\right\} = R^{(m)} \tag{78}$$

where $R^{(m)}$ is a particular covariance matrix that belongs to the class of $\mathrm{TVAR}(m)$ models, against the alternative hypothesis

$$H_1: \quad \mathcal{E}\left\{\boldsymbol{x}_j\boldsymbol{x}_j^H\right\} \neq R^{(m)}. \tag{79}$$

The main difference between this testing problem and the classical one studied in, for example, [29], is that here the admissible class of covariance matrices is the set of $\mathrm{TVAR}(m)$ covariance matrices, rather than the entire class of Hermitian matrices $\mathcal{H}$.

In the classical testing problem

$$H_0: \quad \mathcal{E}\left\{\boldsymbol{x}_j\boldsymbol{x}_j^H\right\} = R, \tag{80}$$

the likelihood ratio (LR) is [29]

$$LR(R|X) = \frac{\mathcal{L}(R|X)}{\max\limits_{R\in\mathcal{H}}\mathcal{L}(R|X)}. \tag{81}$$

The ML over the set of p.d. Hermitian matrices is (for $T \geq N$)

$$\max_{R\in\mathcal{H}}\mathcal{L}(R|X) = \pi^{-T}\exp(-NT)\det{}^{-T}\hat{R} \tag{82}$$

since $R_{ML} = \hat{R}$, which leads to the LR [29]

$$LR(R|X) = \left[\frac{\exp(N)\det(R^{-1}\hat{R})}{\exp\mathrm{tr}(R^{-1}\hat{R})}\right]^T. \tag{83}$$

For the $\mathrm{TVAR}(m)$ model, according to Corollary 1:

$$\max_{R\in\mathcal{R}_{TVAR}^{(m)}}\mathcal{L}(R|X) = \pi^{-T}\exp(-NT)\left[\prod_{q=1}^{N}\hat{y}_{qq}\right]^T \tag{84}$$

and so

$$LR\left(R^{(m)}|X\right) = \left[\frac{\exp(N)\det\left(\left[R^{(m)}\right]^{-1}\hat{R}_N^{(m)}\right)}{\exp\mathrm{tr}\left(\left[R^{(m)}\right]^{-1}\hat{R}_N^{(m)}\right)}\right]^T \tag{85}$$

since $\mathrm{tr}([R^{(m)}]^{-1}\hat{R}) = \mathrm{tr}([R^{(m)}]^{-1}\hat{R}_N^{(m)})$, where $\hat{R}_N^{(m)}$ is the ML $\mathrm{TVAR}(m)$ covariance matrix estimate which, according to (72), has the same elements as the sample matrix $\hat{R}$ in the band $|i-k| \leq m$.

Comparison of the $\mathrm{TVAR}(m)$ LR (85) with the classical one (83) shows that, instead of the unstructured ML covariance matrix estimate $R_{ML} = \hat{R}$ (for $T \geq N$), we now use the $\mathrm{TVAR}(m)$ estimate $R_{ML} = \hat{R}_N^{(m)}$, and so instead of the sample support requirement

$$T \geq N \tag{86}$$

for the general test, we can now use only

$$T \geq m+1 \tag{87}$$

i.i.d. training samples. All this can be summarized as follows.

Theorem 5: For T given i.i.d. training vectors $\boldsymbol{x}_j$ ($j = 1,\ldots,T$) and a given $TVAR(m)$ covariance matrix model $R^{(m)}$, the LR test of size α for

$$H_0: \quad \mathcal{E}\left\{\hat{R} \equiv \frac{1}{T}\sum_{j=1}^{T}\boldsymbol{x}_j\boldsymbol{x}_j^H\right\} = R^{(m)} \tag{88}$$

rejects H_0 if $\Lambda_1 \leq c_\alpha$, where

$$\Lambda_1 \equiv \frac{\exp(N)\det\left(\left[R^{(m)}\right]^{-1}\hat{R}_N^{(m)}\right)}{\exp\mathrm{tr}\left(\left[R^{(m)}\right]^{-1}\hat{R}_N^{(m)}\right)} \leq 1 \tag{89}$$

where $\hat{R}_N^{(m)}$ is the ML estimate of the $TVAR(m)$ covariance matrix (72), and c_α is chosen so that the size of the test is α.

Let us introduce our final theorem.

Theorem 6: For any given $TVAR(m)$ covariance matrix model $R^{(m)}$ and i.i.d. training vectors $\boldsymbol{x}_j$ ($j = 1,\ldots,T$), the LR test of size α for the hypothesis

$$H_0: \quad \mathcal{E}\left\{\hat{R} \equiv \frac{1}{T}\sum_{j=1}^{T}\boldsymbol{x}_j\boldsymbol{x}_j^H\right\} = cR^{(m)}, \quad c > 0 \tag{90}$$

rejects H_0 if

$$\Lambda_2 \equiv \frac{\det\left(\left[R^{(m)}\right]^{-1}\hat{R}_N^{(m)}\right)}{\left[\frac{1}{N}\mathrm{tr}\left(\left[R^{(m)}\right]^{-1}\hat{R}_N^{(m)}\right)\right]^N} \leq d_\alpha \leq 1 \tag{91}$$

where d_α is chosen so that the size of the test is α.

This theorem is the $\mathrm{TVAR}(m)$-generalization of the well-known sphericity test [29] for an arbitrary Hermitian covariance matrix.

Theorems 5 and 6 verify the expected fact that the ML estimate can be treated as a sufficient statistic for any hypothesis test regarding some model that belongs to the admissible class used in the ML derivation. In our case, the admissible set is the class of *all* $\mathrm{TVAR}(m)$ models, and the ML estimate $\hat{R}_N^{(m)}$ clearly gives the ultimate value of unity for both LRs Λ_1 and Λ_2. Moreover, this test is now available for $T \geq m+1$, which means a significant reduction in the required i.i.d. sample size T compared with the general test (83), where T cannot be less than the dimension of the problem N.

Finally, given the ML estimate $\hat{R}_N^{(m)}$, the ML estimate of the time-frequency $\mathrm{TVAR}(m)$ function is

$$\hat{F}_t^{(m)}(w_t) = \frac{\hat{\sigma}_t^2}{\left|\sum_{k=0}^{m}\hat{a}_t(k)\exp(-iw_tk)\right|^2} \quad \text{for} \quad t = m+1,\ldots,N \tag{92}$$

with $\hat{a}_t(0) = 1$, and where $\hat{\sigma}_t^2$ and $\hat{a}_t(k)$ are uniquely specified by $\hat{R}_N^{(m)}$.

IV. $\mathrm{TVAR}(m)$ Modeling of HF Radar Returns Contaminated by Ionospheric Phase-Path Variations

As mentioned in the Introduction, one physical phenomenon that converts potentially stationary OTH radar clutter into a time-varying process is phase-path variation along the ionospheric propagation channel. A simplistic interpretation of this phenomenon is the variation of the virtual height of the ionospheric layer involved in the oblique backscattering during the coherent integration time (CIT) [30]. This variation in propagation path (radar slant range) introduces a multiplicative Doppler frequency modulation (FM) over the CIT. Therefore originally stationary sea clutter, that could be accurately modeled as an $\mathrm{AR}(m)$ process, is observed as a $\mathrm{TVAR}(m)$ process.

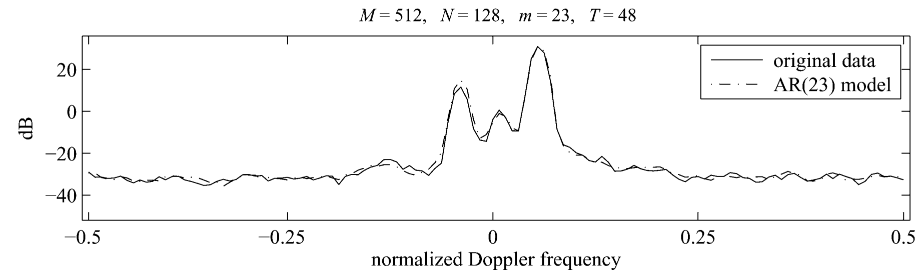

Fig. 1. Doppler power spectrum estimates of the original and $\mathrm{AR}(m)$-modeled data.

As a stationary clutter return, we used a clutter time-series data from a particular range cell collected by an experimental high-frequency (HF) surface-wave radar facility [31]. Unlike sky-wave radar, surface-wave radar observe clutter returns directly via surface wave (not involving ionospheric propagation), hence such clutter is unperturbed by the ionosphere and can be treated as "ideal" stationary clutter. The CIT comprises $M = 512$ repetition periods, *i.e.,* a single data vector consists of 512 complex values. Fig. 1 shows the conventionally averaged Doppler power spectrum for $N = 128$ samples, which involves forward- and backward-averaging of every 128-element partial Doppler spectrum. We used the 512 values to estimate the $m = 23$ stationary AR model, AR(23), in the following way. First, we used sliding-window averaging to form a number of 24-variate training vectors

$$\boldsymbol{x}_j \equiv [x_j, \ldots, x_{j+23}]^T \quad \text{for} \quad j = 1, \ldots, 488 \tag{93}$$

and computed the covariance matrix estimate

$$\hat{R}_{24} = \frac{1}{976} \sum_{j=1}^{488} \left(\boldsymbol{x}_j \boldsymbol{x}_j^H + J \boldsymbol{x}_j^* \boldsymbol{x}_j^T J \right). \tag{94}$$

Then we form the $(m+1)$-variate vector

$$[a_0, \ldots, a_{23}]^T \equiv \hat{R}_{24}^{-1} \boldsymbol{e}_{24} \tag{95}$$

and find the roots of the associated polynomial $a(z) = \sum_{k=0}^{23} a_k z^k$ to form the new polynomial [32]

$$p(z) = \sum_{k=0}^{23} p_k z^k = \exp(i\delta) \prod_{j=1}^{p} \frac{1 - z_j^* z}{z - z_j} a(z) \tag{96}$$

where

$$\delta \equiv \pi p + \sum_{j=1}^{p} \arg z_j \tag{97}$$

and z_j are the roots of $a(z)$ inside the unit disk, $|z| < 1$, taking multiplicity into account. The new $(m+1)$-variate vector $\boldsymbol{p}_{24} \equiv [p_0 > 0, p_1, \ldots, p_{23}]^T$ has no zeros inside the unit disk, and so can be presented as [32]

$$\boldsymbol{p}_{128} \equiv \begin{bmatrix} \boldsymbol{p}_{24} \\ 0 \end{bmatrix} = \mathcal{T}_{128}^{-1} \boldsymbol{e}_{128} \tag{98}$$

where $\mathcal{T}_{128}$ is the AR(23) 128-variate p.d. Toeplitz covariance matrix given by $\boldsymbol{p}_{128}$ via the Gohberg–Semencul formula (24). This technique differs from the well-known Burg estimation technique by using a p.d. Hermitian matrix ($\hat{R}_{24}$ in our case) to initialize Toeplitz covariance matrix estimation ("embedding"). Fig. 1 shows a sample power spectrum obtained by averaging over $2(m+1) = 48$ artificially generated $N = 128$ vectors of AR(23) random numbers produced by this model. We see that this AR(23) model generates a power spectrum that is indistinguishable from the observed surface-wave radar data.

Though experimentally derived, this AR(23) model will be used in following simulations as the clairvoyant (true) model of stationary (unperturbed) clutter. First, the effect of Doppler FM introduced by phase-path variation is simulated by the product

$$\boldsymbol{x}_j = D(k) \boldsymbol{y}_j \quad \text{for} \quad j = 1, \ldots, T \tag{99}$$

where $\boldsymbol{y}_j$ are the $N = 128$-variate vectors (snapshots) of stationary clutter generated by our AR(23) model, and $D(k)$ is the diagonal matrix

$$D(k) = \mathrm{diag}\left[\exp\left(\frac{i2\pi k}{N} \left[1 - \cos \frac{2\pi p t}{N\ell} \right] \right) \right] \quad \text{for} \quad t = 1, \ldots, N \tag{100}$$

where k is the index of the periodic FM, and p/ℓ is its (relative) frequency. In nature, FM is not necessarily periodic, so the FM periodicity will not be referenced in our simulations.

Our model means that if

$$\mathcal{E}\left\{ \boldsymbol{y}_j \boldsymbol{y}_j^H \right\} = \mathcal{T}_{128}^{(23)} \quad \text{for} \quad j = 1, \ldots, T \tag{101}$$

then

$$\mathcal{E}\left\{ \boldsymbol{x}_j \boldsymbol{x}_j^H \right\} = D(k) \mathcal{T}_{128}^{(23)} D^H(k) \quad \text{for} \quad j = 1, \ldots, T. \tag{102}$$

For the FM parameters $k = 40$, $\ell = 1$ and $p = 1$, Fig. 2 illustrates the exact TVAR(23) time-frequency function. We clearly see that the distinctive main maxima of the original AR(23) spectrum in Fig. 1 are reproduced in their time-varying fashion. The impact of phase-path variation on conventional OTH radar Doppler processing (which is just a weighted FFT) is shown in Fig. 3, where simulated Doppler spectra are plotted for both the original data $\boldsymbol{y}_j$ and the modulated vector $\boldsymbol{x}_j$. We see a significant broadening of the dominant spectral lines, known as the first-order Bragg lines [33]. Note that a target return (ionospherically uncontaminated) usually appears as a pure tone at a certain Doppler frequency.

It is now obvious that conventional target detection will fail for a weak target located close to, or between, the Bragg lines due to this broadening. Indeed, Fig. 4 introduces spectra for stationary AR(23) data with an "artificial target" $\boldsymbol{y}_j + a\boldsymbol{s}(\Omega)$, and also for the data having undergone simulated FM $D(k)[\boldsymbol{y}_j +$

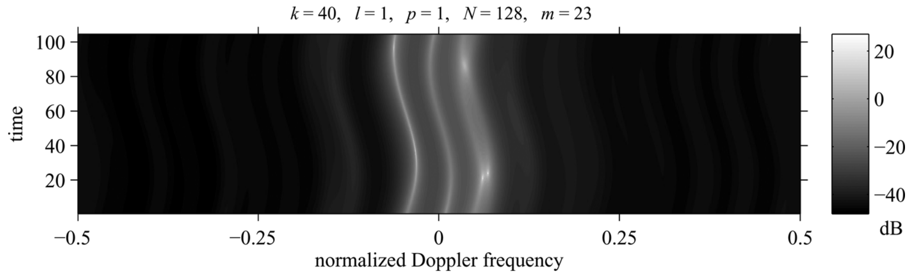

Fig. 2. Exact TVAR(23) time-frequency function.

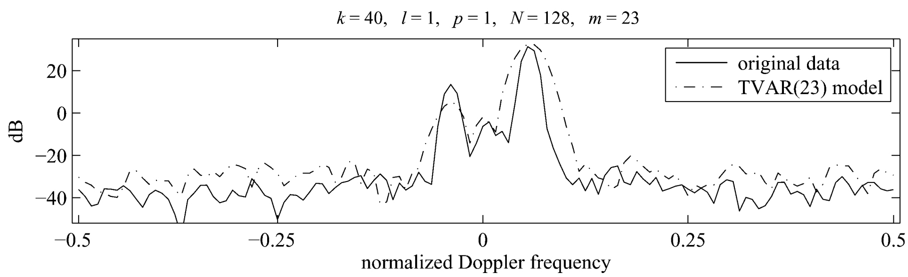

Fig. 3. Doppler spectra (weighted Fourier transforms) of a single $N = 128$ time sequence of a stationary surface-wave clutter return, and a single realization of a TVAR(23) model of sky-wave clutter.

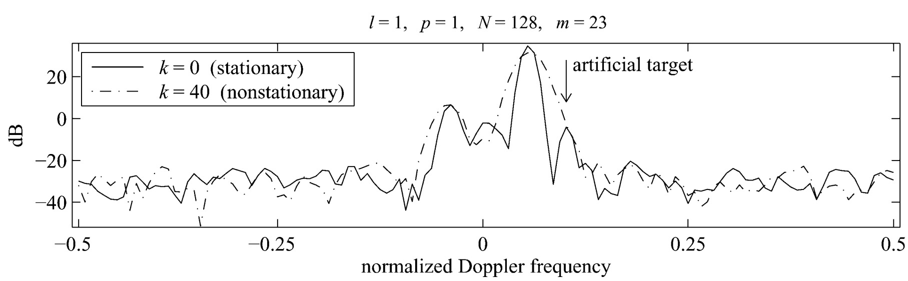

Fig. 4. Doppler spectra (weighted Fourier transforms) of a single realization of a AR(23) stationary model, and a TVAR(23) model of sky-wave clutter with an artificial target.

$a\boldsymbol{s}(\Omega)]$, where a is the target's amplitude and Ω is its spatial frequency. This clearly illustrates the viability of conventional processing in this case for the original data, but not for the contaminated data.

Specifically, the stationary clairvoyant filter is

$$\boldsymbol{w}_{AR}^{opt} = \left[\mathcal{T}_{128}^{(23)}\right]^{-1} \boldsymbol{s}(\omega_0),$$
$$\boldsymbol{s}(\omega) \equiv [1, \exp(i\omega), \ldots, \exp([N-1]i\omega)]^T \quad (103)$$

and for our example Doppler frequency $\omega_0 = 2\pi 12/N$, this filter gives a (normalized) output SNR (4) of

$$q\left(\boldsymbol{w}_{AR}^{opt}\right) = \boldsymbol{s}^H(\omega_0)\left[\mathcal{T}_{128}^{(23)}\right]^{-1}\boldsymbol{s}(\omega_0) = 16.0 \text{ dB}. \quad (104)$$

Note that the clairvoyant $\text{TVAR}(m)$ filter for the time-varying model has the same (normalized) output SNR, since in this case

$$\boldsymbol{s}_{TVAR} = D(k)\boldsymbol{s}(\omega_0) \quad (105)$$

and

$$\boldsymbol{w}_{TVAR}^{opt} = \left[D(k)\mathcal{T}_{128}^{(23)} D^H(k)\right]^{-1} D(k)\boldsymbol{s}(\omega_0). \quad (106)$$

Conventional Doppler processing for the stationary case ($k = 0$) gives, as expected, the reasonably high SNR for our target Doppler frequency of $q_{AR}^{conv} = 13.5$ dB, but for time-varying scenario ($k = 40$), the normalized SNR for $D(k)\mathcal{T}_{128}^{(23)}D^H(k)$ is much smaller at $q_{TVAR}^{conv} = 2.9$ dB, which makes detection of this target impossible, as illustrated by Fig. 4.

In [34], [35], various "desmearing" techniques were proposed based on the assumption of local stationarity for the model (99) with sufficiently slow FM. This allows us to estimate locally a (necessarily) low-order ($m = 3$ or 4) $\text{AR}(m)$ process, and then a direct estimation of the local frequency increment between adjacent "stationary" intervals. While those techniques led to a noticeable detection improvement in some cases, they suffers from an inherent contradiction: a quite high AR order m is needed to capture the real clutter spectrum behavior, but over this significant time interval (24 lags in our example) the assumption regarding local stationarity usually fails.

Our current approach again relies on the fact that both target and clutter returns are equally affected by the same ionospheric phase-path variations; but instead of using a low-order $\text{AR}(m)$ model over short time intervals, we can now estimate the TVAR(23) covariance matrix given $T \simeq 2(m+1)$ training samples (resolution cells). Fig. 5 shows a sample TVAR(23)

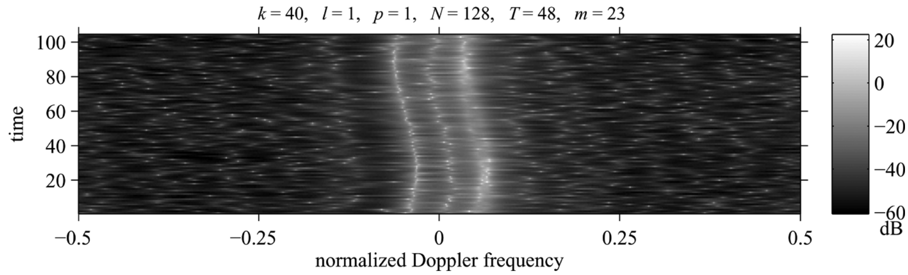

Fig. 5. Sample TVAR(23) time-frequency function for $T = 48$ training samples.

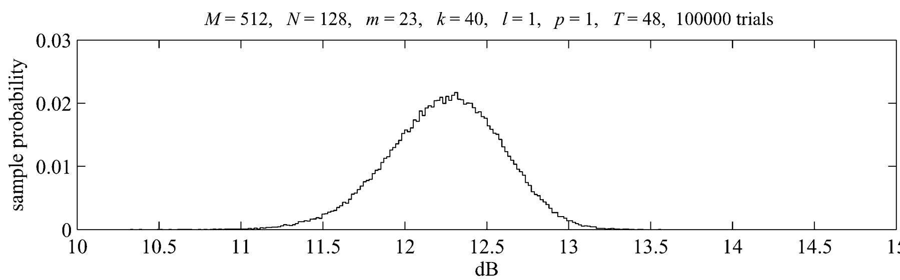

Fig. 6. Histogram of normalized SNR improvement achieved by adaptive TVAR processing.

time-frequency clutter function for $T = 2(m+1) = 48$ samples. Comparison of this estimate with the true function in Fig. 2 demonstrates that the time variations of the Bragg lines are accurately reproduced in the sample estimate, which enables the FM to be accurately estimated.

Finally, Fig. 6 shows a sample histogram of the SNR improvement achieved by the adaptive TVAR(23) filter (4)

$$q\left(\hat{\boldsymbol{w}}_{TVAR}^{(23)}\right) = \frac{\left|\hat{\boldsymbol{w}}_{TVAR}^{(23)H}\boldsymbol{s}\right|^2}{\hat{\boldsymbol{w}}_{TVAR}^{(23)H}\left[D(k)\mathcal{T}_{128}^{(23)}D^H(k)\right]^{-1}\hat{\boldsymbol{w}}_{TVAR}^{(23)}} \tag{107}$$

where

$$\hat{\boldsymbol{w}}_{TVAR}^{(23)} \equiv \left[\hat{R}_{128}^{(23)}\right]^{-1} D(k)\boldsymbol{s}(\omega_0). \tag{108}$$

We see that for $T = 2(m+1) = 48$ training samples, the average SNR degradation (due to the finite-sample support in estimating the $\mathrm{TVAR}(m)$ covariance matrix) relative to the clairvoyant filter value (16.0 dB) is slightly above 3 dB, which permits efficient target detection, something that is not possible by conventional processing.

Recall that in order to achieve a similar SNR loss factor by adaptive filtering that does not exploit the TVAR(23) model, the sample size $T \simeq 2N = 256$ would be required [8], which is considerably more than our $T = 48$ in this example. For HF OTH radar applications with a relatively small number of homogeneous range cells available, this reduction in sample support with $\mathrm{TVAR}(m)$ modeling is very important.

V. Summary and Conclusions

We have considered TVAR models for particular radar applications where multiple i.i.d. observations of the same nonstationary process are available. Typical examples with identically affected clutter returns backscattered by different radar range resolution cells are introduced and examined. Such multiple observations allow us to not impose any additional restrictions on the nature of the time variations (such as "local stationarity"). We showed that the only necessary and sufficient condition for a p.d. Hermitian matrix to be a covariance matrix of a $\mathrm{TVAR}(m)$ model is that its inverse is a $(2m+1)$-wide band matrix. We also reintroduced results on solving the band-extension problem (Theorems 1–3) provided by Dym and Gohberg. Here the goal is to find the specific extension (completion) to the given $(2m+1)$-wide band of a Hermitian matrix whose inverse has zeros outside this band. This problem is the time-varying generalization of Burg's famous ME extension to a $(2m+1)$-wide band Toeplitz Hermitian matrix. As in the stationary (Burg) case, the Dym–Gohberg extension has maximum possible determinant (and, hence, ME) for the completed Hermitian matrix. Dym and Gohberg also showed that the only necessary and sufficient condition for the extension to exist is that all $(m+1)$-variate submatrices in the band are p.d. For the usual sample covariance matrix $\hat{R}$ built with $T \geq m+1$ i.i.d. training samples, this condition is satisfied.

We demonstrated that the Dym–Gohberg extension of the $(2m+1)$-wide band of $\hat{R}$ is the ML estimate of the $\mathrm{TVAR}(m)$ Hermitian covariance matrix for $T > m+1$ i.i.d. complex Gaussian observations. We also introduced LR tests for the null hypothesis that a specific $\mathrm{TVAR}(m)$ covariance matrix is the covariance matrix of the training data.

Our $\mathrm{TVAR}(m)$ estimation technique was then applied to a particular HF OTH radar problem where ionospheric phase-path variations introduce unwanted Doppler FM. We first showed that the underlying stationary clutter is accurately represented by an AR(23) model, with the ionospheric contamination causing significant broadening of the main Doppler spectrum peaks, thus masking targets that are nearby in Doppler frequency. Whereas conventional processing (a weighted FFT) fails to detect such targets, we demonstrated that TVAR(23) modeling with adaptive (ML) covariance matrix estimation using only $T = 48$ training samples can provide a significant detection improvement. A similar improvement using conventional (unstructured) clutter covariance matrix estimation requires a sample size T that significantly exceeds the dimension of the problem ($N = 128$ in our example). Thus, the relatively high order of the $\mathrm{TVAR}(m)$ model has enough flexibility to capture even subtle features of the clutter return, yet gives a great reduction in required sample support (48 instead of about 256 in our example).

This example is only one demonstration of the practical importance of our $\mathrm{TVAR}(m)$ modeling for multiple observations of nonstationary phenomena. Moreover, we have not addressed all the problems and issues that are associated with this model, such as order estimation. These will be addressed in our next paper.

References

[1] A. Kolosov, *Over-the-Horizon Radar*. Boston, MA: Artech House, 1987, English translation by W. F. Barton.

[2] R. Klemm, *Space-Time Adaptive Processing: Principles and Applications*. London, U.K.: IEE, 1998.

[3] J. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Dep. Elect. Eng., Stanford University, Stanford, CA, 1975.

[4] C. Hawkes and S. Haykin, "Modeling of clutter for coherent pulsed radar," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 703–707, Nov. 1975.

[5] S. Haykin, B. Currie, and S. Kesler, "Maximum-entropy spectral analysis of radar clutter," *Proc. IEEE*, vol. 70, no. 9, pp. 953–962, 1982.

[6] W. Melvin, "Space-time adaptive radar performance in heterogeneous clutter," *IEEE Trans. Aerosp. Elect. Syst.*, vol. 36, no. 2, pp. 621–633, Apr. 2000.

[7] T. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1958.

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

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

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

[11] S. Anderson and A. Mahoney, "Spectral estimation techniques for HF sea clutter analysis," in *Proc. Int. Symp. Signal Processing Its Applications (ISSPA)*, Gold Coast, Australia, Aug. 1990, pp. 123–128.

[12] O. Cheremisin, "Efficiency of adaptive algorithms with regularised sample covariance matrix," *Radio Eng. Electron. Phys.*, vol. 27, no. 10, pp. 69–77, 1982.

[13] M. Daniels and R. Kass, "Shrinkage estimators for covariance matrices," *Biometrics*, vol. 57, pp. 1173–1184, 2001.

[14] R. Dahlhaus, "Fitting time series models to nonstationary processes," *Ann. Statist.*, vol. 25, pp. 1–37, 1997.

[15] S. Mallat, G. Papanicolaou, and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *Ann. Statist.*, vol. 26, pp. 1–47, 1998.

[16] K. Eom, "Analysis of acoustic signatures from moving vehicles using time-varying autoregressive models," *Multisignal Syst. Signal Process.*, vol. 10, pp. 357–378, 1999.

[17] C. Sodsri, "Time-varying autoregressive modelling for nonstationary acoustic signal and its frequency analysis," Ph.D. dissertation, Pennsylvania State Univ., University Park, PA, 2003.

[18] H. Dym and I. Gohberg, "Extensions of band matrices with band inverses," *Linear Algebra Appl.*, vol. 36, pp. 1–24, 1981.

[19] R. Grone, C. Johnson, E. M. de Sá, and H. Wolkowicz, "Positive definite completions of partial Hermitian matrices," *Linear Algebra Appl.*, vol. 58, pp. 109–124, 1984.

[20] H. Dym and I. Gohberg, "A new class of contractive interpolants and maximum entropy principles," in *Topics in Operator Theory and Interpolation*, I. Gohberg, Ed. Basel, Switzerland: Birkhäuser Verlag, 1988, pp. 117–150.

[21] H. Woerdeman, "Matrix and operator extensions," Ph.D. dissertation, Vrije Universiteit, Amsterdam, The Netherlands, 1989.

[22] S. Rotheram, "Ground wave propagation: I theory for short distances," *Inst. Elect. Eng. Proc. Part F: Radar, Sonar Navig.*, vol. 128, no. 5, pp. 275–284, 1981.

[23] B. Porat, *Digital Processing of Random Signals*, 5th ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[24] I. Gohberg and A. Semencul, "On the inversion of finite Toeplitz matrices and their continuous analogues," (in Russian) *Mat. Issled.*, vol. 7, no. 2, pp. 201–223, 1972.

[25] M. Krein and A. Nudelman, *The Markov Moment Problem and Extremal Problems*. Providence, RI: Amer. Math. Soc., 1977, translation of mathematical monographs vol. 50.

[26] A. Artemenko, "Hermitian positive functions and positive functionals," (in Russian) *Theory of Funct., Funct. Anal. Their Appl.*, vol. 1:41 & 2:42, pp. 3–16, 1984, & 3-21.

[27] Y. Abramovich, D. Arov, and V. Kachur, "Adaptive cancellation filters for stationary interference with a Toeplitz correlation matrix," *Soviet J. Commun. Technol. Electron.*, vol. 33, no. 4, pp. 54–61, 1988, English translation of Radioteknika i Elektronika.

[28] M. Siotani, T. Hayakawa, and Y. Fujikoshi, *Modern Multivariate Statistical Analysis*. Columbus, OH: Amer. Sci., 1985.

[29] R. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.

[30] J. Parent and A. Bourdillon, "A method to correct HF skywave backscattered signals for ionospheric frequency modulation," *IEEE Trans. Antennas Propag.*, vol. 36, no. 1, pp. 127–135, 1988.

[31] Y. Abramovich and P. Turcaj, "Impulsive noise mitigation in spatial and temporal domains for surface-wave over-the-horizon radar," in *Proc. ASAP-2001*, MA, 2001, pp. 19–24, MIT Lincoln Lab.

[32] D. Arov, Y. Abramovich, J. Böhme, and D. Maiwald, "Spectral analysis via analytic continuation problems," in *Proc. 3rd IMA Conf. Mathemat. Signal Process.*, Warwick, 1992, pp. 367–374.

[33] D. Barrick and J. Snider, "The statistics of HF sea-echo Doppler spectra," *IEEE Trans. Antennas Propag.*, vol. 25, no. 1, pp. 19–28, 1977.

[34] Y. Abramovich, S. Anderson, G. Frazer, and I. Solomon, "Measurement and interpretation of phase fluctuations in HF radar echoes propagated via sporadic E layers," in *Proc. Conf. Solar and Terrestrial Physics (STP)*, Adelaide, Australia, 1995, pp. 152–155.

[35] S. Anderson and Y. Abramovich, "A unified approach to detection, classification, and correction of ionospheric distortion in HF skywave radar systems," *Radio Sci.*, vol. 33, no. 4, pp. 1055–1067, 1998.

Yuri I. Abramovich (M'96–SM'06) received the Dipl. Eng. (Hons.) degree in radio electronics in 1967 and the Cand. Sci. degree (Ph.D. equivalent) in theoretical radio techniques in 1971, both from the Odessa Polytechnic University, Odessa (Ukraine), U.S.S.R., and in 1981, he received the D.Sc. degree in radar and navigation from the Leningrad Institute for Avionics, Leningrad (Russia), U.S.S.R.

From 1968 to 1994, he was with the Odessa State Polytechnic University, Odessa, Ukraine, as a Research Fellow, Professor, and ultimately as Vice-Chancellor of Science and Research. From 1994 to 2006, he was with the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP), Adelaide, Australia. Since 2000, he has been with the Australian Defence Science and Technology Organisation (DSTO), Adelaide, as Principal Research Scientist, seconded to CSSIP until its closure. His research interests are in signal processing (particularly spatio-temporal adaptive processing, beamforming, signal detection, and estimation), its application to radar (particularly over-the-horizon radar), electronic warfare, and communication.

Prof. Abramovich served as Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2005.

Nicholas K. Spencer received the B.Sc. (Hons.) degree in applied mathematics in 1985 and the M.Sc. degree in computational mathematics in 1992, both from the Australian National University, Canberra.

He has been with the Australian Department of Defence, Canberra; the Flinders University of South Australia, Adelaide; the University of Adelaide; and the Australian Centre for Remote Sensing, Canberra; and the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP), Adelaide, in the areas of computational and mathematical sciences. He is currently a Senior Researcher at Adelaide Research & Innovation Pty. Ltd. (ARI), Australia. His research interests include array signal processing, parallel and supercomputing, software best-practice, human–machine interfaces, multilevel numerical methods, modeling and simulation of physical systems, theoretical astrophysics, and cellular automata.

Michael D. E. Turley received the Ph.D. degree in plasma physics from the Flinders University of South Australia, Adelaide, in 1986.

Since 1986, he has been employed on HF skywave and surface wave radar projects, with the Defence Science and Technology Organisation, Edinburgh, South Australia. As a Principal Research Scientist, he presently heads the Signal Processing and Propagation Group in the Intelligence, Surveillance & Reconnaissance Division. His research interests are in the area of ionospheric physics and signal processing techniques, and include clutter mitigation, signal detection, CFAR, EW, adaptive beamforming, and spectral analysis.