



Two dimensional noncausal AR-ARCH model: Stationary conditions, parameter estimation and its application to anomaly detection

Saman Mousazadeh*, Israel Cohen

Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel



ARTICLE INFO

Article history:

Received 5 December 2012

Received in revised form

25 November 2013

Accepted 2 December 2013

Available online 12 December 2013

Keywords:

Image anomaly detection

Non-causal AR-ARCH

Parameter estimation

Matched subspace detector

ABSTRACT

Image anomaly detection is the process of extracting a small number of clustered pixels which are different from the background. The type of image, its characteristics and the type of anomalies depend on the application at hand. In this paper, we introduce a new statistical model called noncausal autoregressive–autoregressive conditional heteroscedasticity (AR-ARCH) model for background in sonar images. Based on this background model, we propose a novel anomaly detection technique in sonar images. This new statistical model (i.e. noncausal ARCH) is an extension of the conventional ARCH model. We provide sufficient stationarity conditions and develop a computationally efficient method for estimating the model parameters which reduces to solving two sets of linear equations. We show that this estimator is asymptotically consistent. Using matched subspace detector (MSD) along with noncausal AR-ARCH modeling of the background in the wavelet domain, we propose an anomaly detection algorithm for sonar images, which is computationally efficient and less dependent on the image orientation. Simulation results demonstrate the performance of the proposed parameter estimation and the anomaly detection algorithm.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Image anomaly detection refers to the problem of finding regions in an image which do not conform to the expected behavior. The characteristic of the image and the type of anomalies are application dependent. Anomaly detection has a wide range of applications, such as silicon wafer defect detection [1,2], medical imaging [3] and sea-mine detection in side-scan sonar images [4] just to name a few. In sea-mine detection, lethal targets must be detected with nearly 100% reliability. False detections are not disastrous but might slow down the demining process. Every anomaly detection algorithm consists of some or all of the following stages: selection of an appropriate feature space; choosing an appropriate

statistical model which represents the image background and selection of a detection algorithm.

A proper selection of a feature space, which allows distinction of anomalies from the background, is an important part of an anomaly detection algorithm. Features can be extracted from the image pixels themselves or from the image after passing through a transform. Kazantsev et al. [5] introduced a feature space based on two circular concentric windows W_1 and W_2 with radii R_1 and R_2 ($R_1 < R_2$), respectively. A similar approach was taken by Schweizer and Moura [6], where two concentric rectangles serve as the moving window. In these methods the features are extracted directly from the image itself. Features can be also extracted from the image in a transformed domain. Laine et al. [7] used a dyadic wavelet transform in mammography to emphasize mammographic features while reducing the noise. Strickland and Hahn [8] used an undecimated wavelet transform for detection of Gaussian objects in Markov noise. Xia et al. [9] used the

* Corresponding author.

E-mail addresses: smousazadeh@gmail.com,
smzadeh@tx.technion.ac.il (S. Mousazadeh),
icohen@ee.technion.ac.il (I. Cohen).

wavelet transform to insert an undetectable watermark into digital imagery. Noiboar and Cohen [4] used undecimated discrete wavelet transform (DWT) for anomaly detection in sonar images.

Once we have extracted our feature either from the image itself or from the transformed image, we must find a statistical model to describe the background in the selected feature space domain. A survey of current literature shows that the most popular models for the background are Gaussian and its extensions. These models are used mostly because of their mathematical tractability. Ashton [10] performed a sub-pixel anomaly detection in multispectral infrared imagery using a Gaussian distribution. Stein et al. [11] used a Gaussian mixture model (GMM) for modeling hyper-spectral imagery. Other extensions such as linear mixing model (LMM) and Gauss–Markov random field (GMRF) are used by several authors for modeling the background. A review of multi-resolution Markov models for signal and image processing can be found in [12]. A survey of results on the structure of two dimensional wide-sense stationary processes with special emphasis on finite order models can be found in [13]. The correlation structure of spatial linear and spatial moving average processes defined on a square lattice has been reviewed by Marc [14].

The generalized autoregressive conditional heteroscedasticity (GARCH) model was first introduced by Bollerslev [15] as an extension of the autoregressive conditional heteroscedasticity (ARCH) model developed by Engle [16] to model econometric data. Since then, many researchers have extended and used these models in several speech and image processing applications. Cohen [17] modeled the speech signal in the short time Fourier transform (STFT) domain as a complex GARCH process and used this model for speech enhancement. AR-GARCH model was utilized for modeling speech signal in the time domain and for developing voice activity detection (VAD) algorithms [18,19]. Abdolahi and Amindavar [20] used the parameters of the GARCH model for speech recognition in Persian isolated digits. Amirmazlaghani et al. [21] used two dimensional GARCH model for speckle suppression in SAR images. Two dimensional GARCH model is also used in image denoising [22]. Noiboar and Cohen [4] used causal GARCH model for anomaly detection in sonar images. The causality assumption incorporated into the GARCH model in [4] is unnatural for images. Developing a non-causal statistical model may lead to an improvement in detection performance by reducing the dependency of the detection procedure on image orientation.

In [23], we presented an anomaly detection method in sonar images based on noncausal AR-ARCH model. The background of the sonar image in the wavelet domain was modeled by a noncausal AR-ARCH model. Matched subspace detector (MSD) was used for detecting the anomaly in the image. In MSD it is assumed that the anomalies are within a subspace. This subspace is assumed to be known or can be estimated using training data.

In this paper, we provide sufficient stationary conditions for the model and propose an effective least squares method for estimating the model parameters. This estimator, which is shown to be asymptotically consistent, is

obtained by solving two sets of linear equations and have a closed-form expression. We also present the detection algorithm in more details. This algorithm is based on the noncausal AR-ARCH modeling of the background and MSD. The rest of the paper is organized as follows. In Section 2, we introduce a two dimensional noncausal ARCH model, provide sufficient stationary conditions and develop a novel technique for estimating the parameters of this model. We also show in the Appendix that this estimator is asymptotically consistent. In Section 3, we introduce our anomaly detection algorithm, which is based on noncausal autoregressive ARCH model and MSD. This section is a detailed description of the method previously proposed by the authors in [23]. In Section 4, the performances of the parameter estimation and anomaly detection are evaluated using simulations. We conclude the paper in Section 5.

2. Noncausal ARCH model

2.1. Two dimensional noncausal ARCH model and its stationary conditions

We define a two dimensional noncausal ARCH(p,q) model as follows:

$$x(t_1, t_2) = \sigma(t_1, t_2)\varepsilon(t_1, t_2), \quad (1)$$

$$\sigma^2(t_1, t_2) = c_0 + \sum_{i=-p}^p \sum_{j=0}^q a_{ij}(x^2(t_1-i, t_2-j) + x^2(t_1+i, t_2+j)), \quad (2)$$

$$c_0 > 0 \quad a_{ij} \geq 0, \quad (3)$$

where c_0 and a_{ij} 's are the two dimensional noncausal ARCH parameters, $\forall i \geq 0 \quad a_{i,0} = 0$, p and q are the model orders in the horizontal and vertical directions, respectively. $\varepsilon(t_1, t_2)$'s are zero mean independent identically distributed (IID) random variables with unit variance and $\sigma^2(t_1, t_2)$'s are called the conditional variances. The constraints (3) guarantee the positiveness of σ^2 . These equations simply state that each pixel in the image (i.e. $x(t_1, t_2)$) is a random variable whose conditional variance (conditioned on the neighboring pixels) is a weighted sum of the squared value of the neighboring pixels. The neighborhood is determined by p and q . This definition is explained graphically in Fig. 1. This figure shows that the conditional variance of the centered pixel is a weighted mean of the squared values of the neighboring pixels where the weights are symmetric (i.e. pixels dotted with the same color have the same weight). This condition is applied to the definition of the model in order to make it identifiable. More specifically, since our parameter estimation method is based on the LS method proposed in [24] for estimating the parameters of the noncausal AR model, and a noncausal AR model is identifiable by the LS method if and only if the weights are symmetric, we ought to assume that the weights are symmetric. Two dimensional noncausal AR model represents the gray-scale level at a specific pixel, as a linear combination of the gray-scale levels of neighboring pixels and an additive white noise. This model has been used in many applications in image

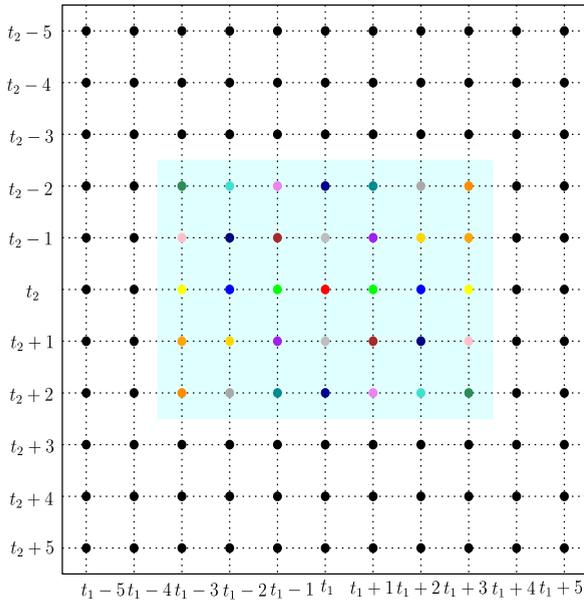


Fig. 1. The conditional variance of the centered pixel is a weighted mean of the squared value of the neighboring pixels where the weights are symmetric (i.e. pixels dotted with the same color have the same weight). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

processing and analysis. For example, it can be used for the design of image restoration algorithms [25,26], for multi-dimensional spectral estimation [27], and for texture analysis and synthesis [28].

Theorem. A two dimensional noncausal ARCH model is mean ergodic with finite second order moment if

$$\sup_{t_1, t_2} \varepsilon^2(t_1, t_2) = \kappa < \infty \quad \text{almost surely} \quad (4)$$

$$\sum_{i=-p}^p \sum_{j=0}^q a_{ij} < \frac{1}{2\kappa} \quad (5)$$

Proof. We define a semi-metric space (X, d) as

$$X = \left\{ x(t_1, t_2) \left| \lim_{N \rightarrow \infty} \frac{1}{(2N+1)^2} \sum_{i=-N}^N \sum_{j=-N}^N x^2(i, j) < \infty \right. \right\} \quad (6)$$

with the following semi-metric:

$$d(x, y) = \lim_{N \rightarrow \infty} \frac{1}{(2N+1)^2} \sum_{i=-N}^N \sum_{j=-N}^N |x^2(i, j) - y^2(i, j)|. \quad (7)$$

Let $T(x(t_1, t_2))$ be the following mapping:

$$\begin{aligned} T(x(t_1, t_2)) &= \left(c_0 + \sum_{i=-p}^p \sum_{j=0}^q a_{ij} (x^2(t_1 - i, t_2 - j) + x^2(t_1 + i, t_2 + j)) \right)^{1/2} \\ &\quad \times \varepsilon(t_1, t_2). \end{aligned} \quad (8)$$

If we show that (X, d) is a complete space and $T(x(t_1, t_2))$ is a contraction mapping from X to itself, then using Banach's fixed point theorem $x(t_1, t_2) = T(x(t_1, t_2))$ has at least one

fixed point. This means that $x(t_1, t_2)$ is mean ergodic with finite second order moment and hence stationary.

To show that (X, d) is a complete space we must show that every Cauchy sequence in X converges to an element in X with respect to the metric d . This can be done in a similar way of proving completeness of ℓ^2 which can be found in every classical textbook on functional analysis. See for example [29]. In order to show that $T(x(t_1, t_2))$ is a contraction mapping from X to itself, first we must show that if $x(t_1, t_2) \in X$ then $y(t_1, t_2) = T(x(t_1, t_2)) \in X$. Note that $\forall N \in \mathbb{N}$ we have

$$\begin{aligned} \alpha_N &= \frac{1}{(2N+1)^2} \sum_{t_1=-N}^N \sum_{t_2=-N}^N y^2(i, j) \\ &= \frac{1}{(2N+1)^2} \sum_{i=-N}^N \sum_{j=-N}^N c_0 \varepsilon^2(t_1, t_2) \\ &\quad + \frac{1}{(2N+1)^2} \sum_{i=-p}^p \sum_{j=0}^q a_{ij} \sum_{t_1=-N}^N \sum_{t_2=-N}^N \varepsilon^2(t_1, t_2) \\ &\quad \times (x^2(t_1 - i, t_2 - j) + x^2(t_1 + i, t_2 + j)) \\ &\leq \frac{\kappa}{(2N+1)^2} \sum_{i=-N}^N \sum_{j=-N}^N c_0 \\ &\quad + \frac{\kappa}{(2N+1)^2} \sum_{i=-p}^p \sum_{j=0}^q a_{ij} \sum_{t_1=-N}^N \sum_{t_2=-N}^N x^2(t_1 - i, t_2 - j) \\ &\quad + x^2(t_1 + i, t_2 + j) \\ &= \kappa c_0 + \frac{\kappa}{(2N+1)^2} \sum_{i=-p}^p \sum_{j=0}^q a_{ij} \\ &\quad \times \sum_{t_1=-N}^N \sum_{t_2=-N}^N x^2(t_1 - i, t_2 - j) + x^2(t_1 + i, t_2 + j). \end{aligned} \quad (9)$$

Taking a limit from both sides of the last inequality (9) and considering the fact that $x \in X$, it can easily be verified that

$$\lim_{N \rightarrow \infty} \alpha_N < \infty,$$

hence, T is a mapping from X to X .

For a mapping to be a contraction mapping on X , there must exist a constant $\beta < 1$ such that $d(Tx, Ty) < \beta d(x, y)$ for all $x, y \in X$. Now note that

$$\begin{aligned} d_N(Tx, Ty) &= \frac{1}{(2N+1)^2} \sum_{t_1=-N}^N \sum_{t_2=-N}^N |x^2(i, j) - y^2(i, j)| \\ &= \frac{1}{(2N+1)^2} \sum_{i=-p}^p \sum_{j=0}^q a_{ij} \sum_{t_1=-N}^N \sum_{t_2=-N}^N \varepsilon^2(t_1, t_2) \\ &\quad \times |x^2(t_1 - i, t_2 - j) - y^2(t_1 - i, t_2 - j)| \\ &\quad + \frac{1}{(2N+1)^2} \sum_{i=-p}^p \sum_{j=0}^q a_{ij} \sum_{t_1=-N}^N \sum_{t_2=-N}^N \varepsilon^2(t_1, t_2) \\ &\quad \times |x^2(t_1 + i, t_2 + j) - y^2(t_1 + i, t_2 + j)|. \end{aligned} \quad (10)$$

Using a similar approach we used to obtain (9), it can be shown that

$$d(Tx, Ty) = \lim_{N \rightarrow \infty} d_N(Tx, Ty) \leq \left(2\kappa \sum_{i=-p}^p \sum_{j=0}^q a_{ij} \right) d(x, y). \quad (11)$$

Hence, $\beta = (2\kappa \sum_{i=-p}^p \sum_{j=0}^q a_{ij})$ and a sufficient stationary condition is

$$\beta < 1 \Rightarrow \sum_{i=-p}^p \sum_{j=0}^q a_{ij} < \frac{1}{2\kappa}. \quad (12)$$

Note that the second condition can be interpreted as a condition on the probability distribution function (pdf) of $\varepsilon(t_1, t_2)$ to be almost surely compactly supported.

2.2. Parameter estimation of two dimensional noncausal ARCH model

Parameter estimation is an important part of any model-based signal processing algorithm. There are several parameter estimation methods for ARCH and GARCH models, for example quasi maximum likelihood (QML) [30], two stage least squares (TSLS) [31], constrained two stage least squares (CTSLS) [32] and multivariate two stage least squares (MTSLS) [33]. Since ARCH and GARCH models are mostly used in econometrics, where processes can be assumed to be causal, to the best of our knowledge, there was no attempt to define noncausal (G)ARCH models and finding a parameter estimation method for the noncausal case. Upon defining a two dimensional noncausal ARCH model in the previous section, in this section we develop a novel parameter estimation method for estimating the parameters of this model. The proposed method is mainly based on MTSLS and uses the same approach utilized in [24] for parameter estimation of the two dimensional noncausal AR model.

Let $y(t_1, t_2) = x^2(t_1, t_2)$. Then from the definition of the two dimensional noncausal ARCH model (1)–(2), we have

$$\begin{aligned} y(t_1, t_2) &= c_0 + \sum_{i=-p}^p \sum_{j=0}^q a_{ij}(y(t_1-i, t_2-j) + y(t_1+i, t_2+j)) \\ &\quad + \sigma^2(t_1, t_2)(\varepsilon^2(t_1, t_2) - 1) \\ &= c_0 + \sum_{i=-p}^p \sum_{j=0}^q a_{ij}(y(t_1-i, t_2-j) + y(t_1+i, t_2+j)) \\ &\quad + \sigma^2(t_1, t_2)\eta(t_1, t_2). \end{aligned} \tag{13}$$

Ignoring the dependence of $\sigma^2(t_1, t_2)\eta(t_1, t_2)$ on the parameters, (13) is similar to a two dimensional noncausal AR process. We utilize a similar approach employed in [24] for parameter estimation. Note that $y(t_1, t_2)$ can be divided into the following two respectively causal and anti-causal processes:

$$\frac{1}{2}y(t_1, t_2) = \frac{c_0}{2} + \sum_{i=-p}^p \sum_{j=0}^q a_{ij}y(t_1-i, t_2-j) + \sigma^2(t_1, t_2)\eta_L(t_1, t_2) \tag{14}$$

$$\frac{1}{2}y(t_1, t_2) = \frac{c_0}{2} + \sum_{i=-p}^p \sum_{j=0}^q a_{ij}y(t_1+i, t_2+j) + \sigma^2(t_1, t_2)\eta_U(t_1, t_2), \tag{15}$$

where $\eta_L(t_1, t_2)$ and $\eta_U(t_1, t_2)$ are the corresponding causal and anti-causal prediction errors such that $\eta_L(t_1, t_2) + \eta_U(t_1, t_2) = \eta(t_1, t_2)$. As in [24], we can assume that $\eta_L(t_1, t_2)$ and $\eta_U(t_1, t_2)$ are uncorrelated. Furthermore, it can also be assumed that $\eta_L(t_1, t_2)$ and $\eta_U(t_1, t_2)$ are uncorrelated with $y(t_1 \pm i, t_2 \pm j) : -p \leq i \leq p, 0 \leq j \leq q$. In order to check that if these assumptions hold, we simulate a noncausal ARCH(1, 1) model and computed the correlation between η, η_L, η_U , and y . In this experiment, we simulated a noncausal ARCH(1, 1) process of the size 200×200 using Banach fixed point

theorem with 15 iterations. The model parameters were set to $c_0 = 1, a_{-1,-1} = 0.03, a_{0,-1} = 0.04, a_{1,-1} = 0.05, a_{-1,0} = 0.06$.

The process noise $\varepsilon(t_1, t_2)$ was chosen to be a sequence of IID random variables driven from uniform distribution having zero mean and unity variance (i.e. $\varepsilon(t_1, t_2) \sim U([- \sqrt{3}, \sqrt{3}])$). The sample cross-correlation of column stack of η_L and y, η_U and y, η and y, η_L and η_U is depicted in Figs. 2(a), (b), (c), and (d), respectively. From these figures, it is obvious that these processes are uncorrelated for any lag greater than zero. Hence, the parameters of the model can be estimated by minimizing the following cost function:

$$\begin{aligned} C(\mathbf{a}) &= \sum_{t_1=p+1}^{N-p} \sum_{t_2=q+1}^{M-q} \left(\frac{1}{2}y(t_1, t_2) - \mathbf{a}^T \mathbf{y}_L(t_1, t_2) \right)^2 \\ &\quad + \left(\frac{1}{2}y(t_1, t_2) - \mathbf{a}^T \mathbf{y}_U(t_1, t_2) \right)^2 \end{aligned} \tag{16}$$

with respect to \mathbf{a} , where N and M are the number of available data in the horizontal and vertical directions, respectively, and $(\cdot)^T$ is the transpose operator. The vector of parameters $\mathbf{a}, \mathbf{y}_L(t_1, t_2)$ and $\mathbf{y}_U(t_1, t_2)$ is defined as follows:

$$\mathbf{a} = [c_0, a_{-p,0} \dots a_{-1,0}, a_{-p,1} \dots a_{p,1}, a_{-p,2} \dots a_{p,2}, \dots, a_{-p,q}, \dots a_{p,q}]^T \tag{17}$$

$$\begin{aligned} \mathbf{y}_L(t_1, t_2) &= [\frac{1}{2}y(t_1-1, t_2) \dots y(t_1-p, t_2), \\ &\quad y(t_1+p, t_2-1) \dots y(t_1-p, t_2-1), \\ &\quad y(t_1+p, t_2-2) \dots y(t_1-p, t_2-2), \dots, \\ &\quad y(t_1+p, t_2-q) \dots y(t_1-p, t_2-q)]^T \end{aligned} \tag{18}$$

$$\begin{aligned} \mathbf{y}_U(t_1, t_2) &= [\frac{1}{2}y(t_1+1, t_2) \dots y(t_1+p, t_2), y(t_1-p, t_2+1) \dots \\ &\quad y(t_1+p, t_2+1), y(t_1-p, t_2+2) \dots \\ &\quad y(t_1+p, t_2+2), \dots, y(t_1-p, t_2+q) \dots \\ &\quad y(t_1+p, t_2+q)]^T. \end{aligned} \tag{19}$$

Since it is difficult to obtain statistical properties of this estimator, we propose the following suboptimal parameter estimation method:

$$\begin{aligned} \hat{\mathbf{a}}_{pr} &= \frac{1}{2} \left(\arg \min_{\mathbf{a}} \sum_{t_1=p+1}^{N-p} \sum_{t_2=q+1}^{M-q} \left(\frac{1}{2}y(t_1, t_2) - \mathbf{a}^T \mathbf{y}_L(t_1, t_2) \right)^2 \right. \\ &\quad \left. + \arg \min_{\mathbf{a}} \left(\frac{1}{2}y(t_1, t_2) - \mathbf{a}^T \mathbf{y}_U(t_1, t_2) \right)^2 \right) \end{aligned} \tag{20}$$

which, as we show latter, is asymptotically unbiased and consistent. The minimum of each of these cost functions in (20) can be found by setting the gradient of the corresponding cost function with respect to the vector of parameters to zero. These cost functions have quadratic forms in parameters and therefore are convex and have a unique minimum. Hence a *primary* estimate of the parameters is obtained as follows:

$$\hat{\mathbf{a}}_{pr}^T = \frac{1}{2} \left(\mathbf{x}_L^T \mathbf{R}_L^{-1} + \mathbf{x}_U^T \mathbf{R}_U^{-1} \right), \tag{21}$$

where

$$\mathbf{R}_L = 2 \sum_{t_1=p+1}^{N-p} \sum_{t_2=q+1}^{M-q} \mathbf{y}_L(t_1, t_2) \mathbf{y}_L^T(t_1, t_2) \tag{22}$$

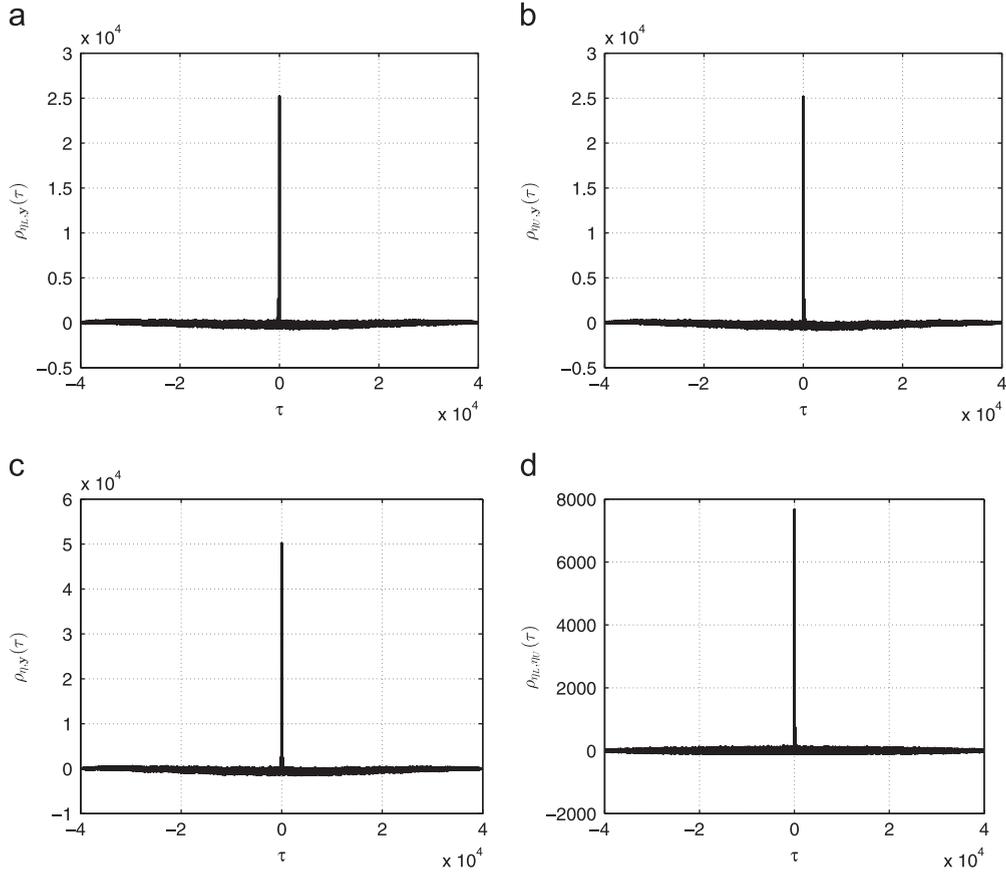


Fig. 2. The sample cross-correlation of: (a) η_L and y , (b) η_U and y , (c) η and y , and (d) η_L and η_U .

$$\mathbf{R}_U = 2 \sum_{t_1=p+1}^{N-p} \sum_{t_2=q+1}^{M-q} \mathbf{y}_U(t_1, t_2) \mathbf{y}_U^T(t_1, t_2) \quad (23)$$

$$\mathbf{x}_L = \sum_{t_1=p+1}^{N-p} \sum_{t_2=q+1}^{M-q} y(t_1, t_2) \mathbf{y}_L(t_1, t_2) \quad (24)$$

$$\mathbf{x}_U = \sum_{t_1=p+1}^{N-p} \sum_{t_2=q+1}^{M-q} y(t_1, t_2) \mathbf{y}_U(t_1, t_2). \quad (25)$$

We denote this estimator by $\hat{\mathbf{a}}_{pr}$ where the subscript pr stands for ‘preliminary’. This is due to the fact that in deriving this estimator we have not considered the dependency of $\sigma^2(t_1, t_2)\eta(t_1, t_2)$ on the parameters. Using (2) and this preliminary estimate of the parameters, we can estimate the conditional variance $\sigma^2(t_1, t_2)$, in the following way:

$$\begin{aligned} \hat{\sigma}^2(t_1, t_2) &= \hat{c}_0^{pr} + \sum_{i=-p}^p \sum_{j=0}^q \hat{a}_{ij}^{pr} (y(t_1-i, t_2-j) + y(t_1+i, t_2+j)) \\ &= \hat{\mathbf{a}}_{pr}^T (\mathbf{y}_L(t_1, t_2) + \mathbf{y}_U(t_1, t_2)), \end{aligned} \quad (26)$$

where \hat{c}_0^{pr} and \hat{a}_{ij}^{pr} are the preliminary estimates of the parameters obtained in (21). It follows from (13) that

$$\frac{y(t_1, t_2)}{\hat{\sigma}^2}(t_1, t_2) \approx \frac{c_0}{\hat{\sigma}^2}(t_1, t_2)$$

$$\begin{aligned} &+ \sum_{i=-p}^p \sum_{j=0}^q a_{ij} \left(\frac{y(t_1-i, t_2-j)}{\hat{\sigma}^2}(t_1, t_2) \right. \\ &\left. + \frac{y(t_1+i, t_2+j)}{\hat{\sigma}^2}(t_1, t_2) \right) + (\varepsilon^2(t_1, t_2) - 1). \end{aligned} \quad (27)$$

Note that the two sides of (27) are not exactly equal because we have canceled out $\sigma^2(t_1, t_2)$ by its estimate, i.e. $\hat{\sigma}^2(t_1, t_2)$. Now, we use the above equation to find our final estimator. If we ignore the approximation used in deriving (27), we can derive the LS estimate of the parameters (i.e. \mathbf{a}) by solving the following optimization problem:

$$\begin{aligned} \hat{\mathbf{a}} &= \frac{1}{2} \left(\operatorname{argmin}_{\mathbf{a}} \sum_{t_1=p+1}^{N-p} \sum_{t_2=q+1}^{M-q} \left(\frac{1}{2} \bar{y}(t_1, t_2) - \mathbf{a}^T \bar{\mathbf{y}}_L(t_1, t_2) \right)^2 \right. \\ &\quad \left. + \operatorname{argmin}_{\mathbf{a}} \left(\frac{1}{2} \bar{y}(t_1, t_2) - \mathbf{a}^T \bar{\mathbf{y}}_U(t_1, t_2) \right)^2 \right), \end{aligned} \quad (28)$$

where

$$\bar{y}(t_1, t_2) = \frac{y(t_1, t_2)}{\hat{\sigma}^2}(t_1, t_2) \quad (29)$$

$$\bar{\mathbf{y}}_L(t_1, t_2) = \frac{\mathbf{y}_L(t_1, t_2)}{\hat{\sigma}^2}(t_1, t_2) \quad (30)$$

$$\bar{\mathbf{y}}_U(t_1, t_2) = \frac{\mathbf{y}_U(t_1, t_2)}{\hat{\sigma}^2}(t_1, t_2). \quad (31)$$

Again, since each of the cost function is a quadratic function of the vector \mathbf{a} , they are convex and have a unique minimum which can be found by setting the derivative of $\bar{C}(\mathbf{a})$ to zero. Hence, our final estimator is obtained as follows:

$$\hat{\mathbf{a}}^T = \frac{1}{2} \left(\bar{\mathbf{x}}_L^T \bar{\mathbf{R}}_L^{-1} + \bar{\mathbf{x}}_U^T \bar{\mathbf{R}}_U^{-1} \right), \quad (32)$$

where

$$\bar{\mathbf{R}}_L = 2 \sum_{t_1=p+1}^{N-p} \sum_{t_2=q+1}^{M-q} \bar{\mathbf{y}}_L(t_1, t_2) \bar{\mathbf{y}}_L^T(t_1, t_2) \quad (33)$$

$$\bar{\mathbf{R}}_U = 2 \sum_{t_1=p+1}^{N-p} \sum_{t_2=q+1}^{M-q} \bar{\mathbf{y}}_U(t_1, t_2) \bar{\mathbf{y}}_U^T(t_1, t_2) \quad (34)$$

$$\bar{\mathbf{x}}_L = \sum_{t_1=p+1}^{N-p} \sum_{t_2=q+1}^{M-q} \bar{\mathbf{y}}_L(t_1, t_2) \bar{\mathbf{y}}_L^T(t_1, t_2) \quad (35)$$

$$\bar{\mathbf{x}}_U = \sum_{t_1=p+1}^{N-p} \sum_{t_2=q+1}^{M-q} \bar{\mathbf{y}}_U(t_1, t_2) \bar{\mathbf{y}}_U^T(t_1, t_2). \quad (36)$$

We denote this estimator as two stage least squares (TSLS) estimator. In the Appendix, we show that under some moment conditions besides mean ergodicity, this estimator is asymptotically unbiased and consistent. It should be noted that the TSLS estimator does not guarantee that the estimated model is stationary. However, this method can be easily modified (by solving the least squares problems under stationary conditions) such that the estimated model be stationary, as was done for the univariate causal ARCH case in [32].

3. Anomaly detection based on noncausal AR-ARCH model

In this section, we introduce our anomaly detection algorithm which is based on noncausal AR-ARCH modeling of the image background in the wavelet domain. This section is mainly based on our previous work on anomaly detection [23]. We use a matched subspace detector (MSD) proposed in [34] along with noncausal AR introduced in [24] and noncausal ARCH model introduced in the previous section. We assume that the anomalies in the images are so rare that their influence on the parameter estimation algorithm can be ignored. The proposed algorithm is an extension of the method proposed in [4]. In [4], the authors introduced an anomaly detection algorithm based on causal AR-GARCH modeling of the images in the wavelet domain. Causality of the model in [4] leads to dependence of the anomaly detection procedure on the orientation of the image. This means that the algorithm is sensitive to the orientation of the image, so in order to use this method one must consider all four orientations of the image which is a time consuming process. Another disadvantage of the procedure proposed in [4] is its high computational load. That method uses causal AR-GARCH model in the wavelet domain, and the parameters of the GARCH part are estimated using a maximum likelihood

(ML) method, which has a very high computational load, since the maximum of the likelihood function is found by a search method. Our method overcomes both of these disadvantages. Using the proposed noncausal model, the sensitivity of our method to image orientation is reduced significantly and the computational load is decreased considerably using the proposed TSLS method for parameter estimation of the noncausal ARCH model.

In what follows, we introduce our noncausal AR-ARCH model for the image background in the wavelet domain. This model is to some extent similar to the model used in [4] with a major difference. The model used in [4] is a causal model whereas our model is noncausal. Let $z(t_1, t_2)$ be the original image. Using the two dimensional undecimated wavelet transform [35], we obtain a set of $2L+1$ images where L is the depth of the wavelet transform. These $2L+1$ images are obtained as follows. The undecimated wavelet transform yields four sub-band images at every analysis level. These four sub-band images are labeled by d_{LH}^i , d_{HL}^i , d_{HH}^i and s_{LL}^i , where the superscript i represents the layer index and the subscripts L and H stand for low-pass and high-pass filtering, respectively. d labels a detail sub-band, s represents the “smooth” sub-band, and the superscript ℓ specifies the analysis level. The undecimated wavelet transform yields a redundant representation. However, the same analysis and synthesis filters are used as in the undecimated wavelet transform, and since the transform preserves the spatial dimensions, it is easy to work with. Furthermore, the undecimated wavelet transform has an additional property, namely, translation invariance, which is important in the context of anomaly detection. As in [8], we use the sub-band images of an undecimated wavelet transform to create a $2L+1$ feature images at every spatial location (t_1, t_2) as follows:

$$y_{2i-1}(t_1, t_2) = d_{LH}^i(t_1, t_2) + d_{HL}^i(t_1, t_2); \quad 1 \leq i \leq L \quad (37)$$

$$y_{2i}(t_1, t_2) = d_{HH}^i(t_1, t_2); \quad 1 \leq i \leq L \quad (38)$$

$$y_{2L+1}(t_1, t_2) = s_{LL}^{2L+1}(t_1, t_2). \quad (39)$$

Let $y_\ell(t_1, t_2)$; $1 \leq \ell \leq 2L+1$ be the ℓ -th feature image in the wavelet domain, obtained by (37)–(39). We assume that the background image is a noncausal AR-ARCH process defined as follows:

$$y_\ell(t_1, t_2) = \sum_{i=-r}^r \sum_{j=0}^s b_{ij}^\ell (y_\ell(t_1-i, t_2-j) + y_\ell(t_1+i, t_2+j)) + x_\ell(t_1, t_2) \quad (40)$$

$$x_\ell(t_1, t_2) = \sigma_\ell(t_1, t_2) \varepsilon_\ell(t_1, t_2) \quad (41)$$

$$\sigma_\ell^2(t_1, t_2) = c_0^\ell + \sum_{i=-p}^p \sum_{j=0}^q a_{ij}^\ell (x_\ell^2(t_1-i, t_2-j) + x_\ell^2(t_1+i, t_2+j)), \quad (42)$$

where b_{ij}^ℓ 's are parameters of the AR part, $\forall i \geq 0$ $b_{i0}^\ell = 0$ and r and s are the order of AR model in the horizontal and vertical directions, respectively. c_0 and a_{ij} 's are the parameters of the ARCH part and $\forall i \geq 0$ $a_{i0} = 0$, p and q are the order of ARCH model in the horizontal and vertical directions, respectively. $\varepsilon_\ell(t_1, t_2)$'s are zero mean independent identically distributed (IID) random variables. We have

defined our two dimensional noncausal AR-ARCH model based on the definition of the two dimensional noncausal AR model [24] extensively used in image signal processing and the two dimensional noncausal ARCH model introduced in Section 2.

Now we propose a method for parameter estimation of the two dimensional noncausal AR-ARCH model. Considering the fact that the ARCH process is white, the parameters of the AR part can be easily estimated with the method proposed in [24]. Let this estimate of the parameters of the AR part be denoted by \hat{b}_{ij}^ℓ . Substituting this estimate of the parameters in (40), the residuals of the AR model (i.e. $x_\ell(t_1, t_2)$) can be estimated as follows:

$$\hat{x}_\ell(t_1, t_2) = y_\ell(t_1, t_2) - \sum_{i=-p}^r \sum_{j=0}^s \hat{b}_{ij}^\ell (y_\ell(t_1 - i, t_2 - j) + y_\ell(t_1 + i, t_2 + j)). \quad (43)$$

Using this estimate of the residuals, the parameters of the ARCH part can be estimated using the TSLS method introduced in the previous section. Substituting this estimate of the parameters of the ARCH part and the estimate of the residuals (i.e. $\hat{x}_\ell(t_1, t_2)$) in (42), the estimate of the conditional variance is obtained as follows:

$$\hat{\sigma}_\ell^2(t_1, t_2) = \hat{c}_0^\ell + \sum_{i=-p}^q \sum_{j=0}^q \hat{a}_{ij}^\ell (\hat{x}_\ell^2(t_1 - i, t_2 - j) + \hat{x}_\ell^2(t_1 + i, t_2 + j)), \quad (44)$$

where \hat{c}_0^ℓ and \hat{a}_{ij}^ℓ are the estimates of the parameters of the ARCH part obtained by the TSLS method. The order of the model (i.e. p , q , r and s) can be selected by Akaike information criterion (AIC) or Bayesian information criterion (BIC). In this work we select the order of the model experimentally. A more detailed discussion on order selection needs further research and is beyond the scope of this paper.

Using this estimate of the conditional variance, the anomaly detector is obtained as follows. For each pixel in each layer, we create a column vector $\hat{\mathbf{x}}_\ell(t_1, t_2)$ by column stacking an image chip of size $T_1 \times T_2$ of $\hat{x}_\ell(t_1, t_2)$ centered around the pixel (t_1, t_2) in the ℓ -th layer. Assume that there exists no interference and let $\boldsymbol{\psi}_\ell(t_1, t_2)$ be vector locating the anomaly within its subspace $\mathbf{H}_\ell = \text{span}(\mathbf{H}_\ell)$. These matrices, i.e. \mathbf{H}_ℓ ; $1 \leq \ell \leq 2L+1$, are obtained from a set of M training patches containing anomalies. Succinctly, suppose we are given a set of training images containing mines. Each of these images is transformed to the wavelet domain and for each image, $2L+1$ feature images are extracted using (37)–(39). In a specific layer ℓ , we used the corresponding M images to obtain the anomaly subspace in that layer. This task is done by first extracting all possible $T_1 \times T_2$ image chips (without overlap) from these

M images, column stacking each of these patches and finding the K principle components using principle component analysis (PCA), see [4] for a more detailed discussion. Let $\boldsymbol{\sigma}_\ell^2(t_1, t_2)$ be a vector obtained by column stacking of an image chip of size $T_1 \times T_2$ of $\sigma_\ell^2(t_1, t_2)$ centered around the pixel (t_1, t_2) in the ℓ -th layer. Let $\boldsymbol{\Sigma}_\ell(t_1, t_2)$ be a diagonal matrix whose diagonal is the vector $\boldsymbol{\sigma}_\ell^2(t_1, t_2)$. We define two hypotheses, \mathcal{H}_0 and \mathcal{H}_1 , which respectively represent the absence and presence of an anomaly as follows:

$$\mathcal{H}_0: \hat{\mathbf{x}}_\ell(t_1, t_2) = \boldsymbol{\Sigma}_\ell(t_1, t_2) \mathbf{e}_\ell(t_1, t_2) \quad (45)$$

$$\mathcal{H}_1: \hat{\mathbf{x}}_\ell(t_1, t_2) = \mathbf{H}_\ell \boldsymbol{\psi}_\ell(t_1, t_2) + \boldsymbol{\Sigma}_\ell(t_1, t_2) \mathbf{e}_\ell(t_1, t_2), \quad (46)$$

Under these two hypotheses, we assume that the sample conditional distribution of $\hat{\mathbf{x}}_\ell(t_1, t_2)$ is Gaussian with identical covariance matrices with different means, i.e.

$$\mathcal{H}_0: \hat{\mathbf{x}}_\ell(t_1, t_2) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\ell(t_1, t_2)) \quad (47)$$

$$\mathcal{H}_1: \hat{\mathbf{x}}_\ell(t_1, t_2) \sim \mathcal{N}(\mathbf{H}_\ell \boldsymbol{\psi}_\ell^\ell(t_1, t_2), \boldsymbol{\Sigma}_\ell(t_1, t_2)) \quad (48)$$

Under this assumption, the log-likelihood ratio in the ℓ -th layer is given by [4]

$$L_\ell(t_1, t_2) = ((\boldsymbol{\Sigma}_\ell(t_1, t_2))^{-1/2} \hat{\mathbf{x}}_\ell(t_1, t_2))^T P_{\mathbf{H}_\ell} ((\boldsymbol{\Sigma}_\ell(t_1, t_2))^{-1/2} \times \hat{\mathbf{x}}_\ell(t_1, t_2)) \quad (49)$$

where

$$P_{\mathbf{H}_\ell} = \mathbf{H}_\ell (\mathbf{H}_\ell^T \mathbf{H}_\ell)^{-1} \mathbf{H}_\ell^T \quad (50)$$

Since the true value of the conditional variance ($\boldsymbol{\Sigma}_\ell(t_1, t_2)$) is not available, by the generalized likelihood ratio test, it can be replaced by its estimate ($\hat{\boldsymbol{\Sigma}}_\ell(t_1, t_2)$). Our final detector is given by comparing the generalized likelihood ratio with a predefined threshold, i.e.

$$L_{GLR}(t_1, t_2) = \sum_{\ell=1}^{2L+1} ((\hat{\boldsymbol{\Sigma}}_\ell(t_1, t_2))^{-1/2} \hat{\mathbf{x}}_\ell(t_1, t_2))^T P_{\mathbf{H}_\ell} ((\hat{\boldsymbol{\Sigma}}_\ell(t_1, t_2))^{-1/2} \hat{\mathbf{x}}_\ell(t_1, t_2)) \stackrel{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\gtrless}} T_h \quad (51)$$

where $\hat{\boldsymbol{\Sigma}}_\ell(t_1, t_2)$ is a diagonal matrix whose diagonal elements are obtained by column stacking of an image chip of size $T_1 \times T_2$ of $\hat{\sigma}_\ell^2(t_1, t_2)$ centered around the pixel (t_1, t_2) in the ℓ -th layer. T_h is the threshold and its value selection is a tradeoff between the probability of detection and the probability of false alarm. Increasing T_h decreases both probability of detection and false alarm. The anomaly detection algorithm is summarized in Table 1.

Table 1
Anomaly detection algorithm using noncausal AR-ARCH model.

- (1) Transform the image to the wavelet domain and find the feature images using (37)–(39)
- (2) Find the anomaly subspace \mathbf{H}_ℓ for each layer
- (3) For each layer estimate the AR parameters and the residuals (i.e. $\hat{x}_\ell(t_1, t_2)$) using (43)
- (4) For each layer estimate the ARCH parameters and the conditional variance (i.e. $\hat{\sigma}_\ell(t_1, t_2)$) using (44)
- (5) Find the generalized likelihood ratio using (49)
- (6) Compare the generalized likelihood ratio to a threshold

4. Performance evaluation and simulation results

In this section, we evaluate the performance of our parameter estimation method and anomaly detection algorithm using simulations.

In the first simulation, we evaluate the performance of the parameter estimation method. We use a two dimensional noncausal ARCH model with the following vector of parameters:

$$\mathbf{a} = [1.00 \ 0.05 \ 0.01 \ 0.07 \ 0.04].$$

$\varepsilon(t_1, t_2)$'s are zero mean independent identically random variables uniformly distributed on the interval $[-\sqrt{3}, \sqrt{3}]$. The noncausal ARCH process is obtained using Banach's fixed point theorem with 15 iterations. The probability density function of the estimated parameters $\hat{\mathbf{a}}_j$ (an estimate of the j -th element of \mathbf{a}), obtained by 250 Monte-Carlo iterations for different image sizes, is depicted in Figs. 3–7. The normalized root mean square error (NRMSE) in the estimation of the parameters is depicted Fig. 8. Fig. 8 shows that the mean square error (MSE) in parameter

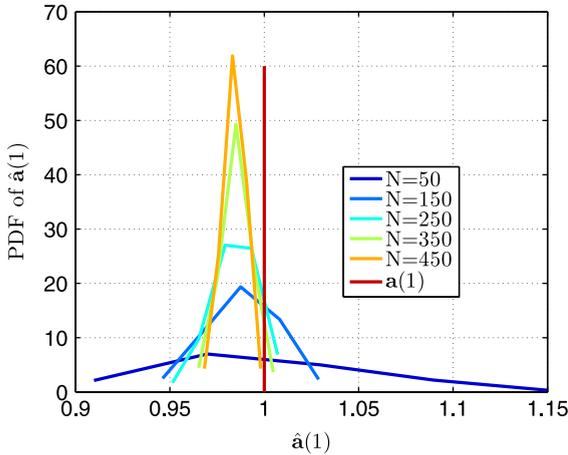


Fig. 3. PDF of $\hat{\mathbf{a}}_1$ for different values of N for an $N \times N$ noncausal ARCH(1, 1) process.

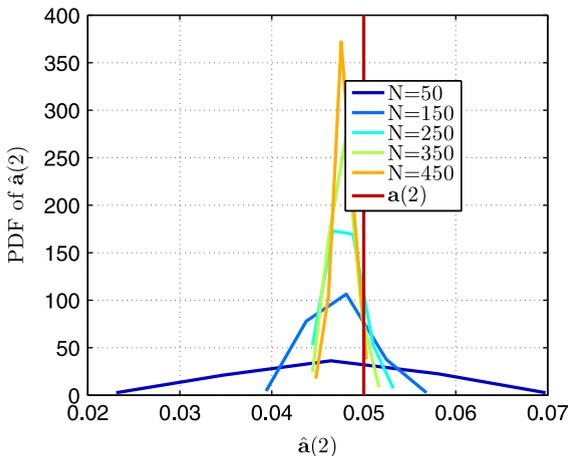


Fig. 4. PDF of $\hat{\mathbf{a}}_2$ for different values of N for an $N \times N$ noncausal ARCH(1, 1) process.

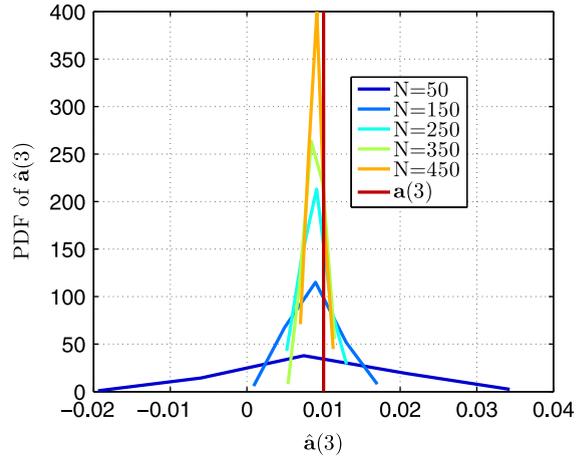


Fig. 5. PDF of $\hat{\mathbf{a}}_3$ for different values of N for an $N \times N$ noncausal ARCH(1, 1) process.

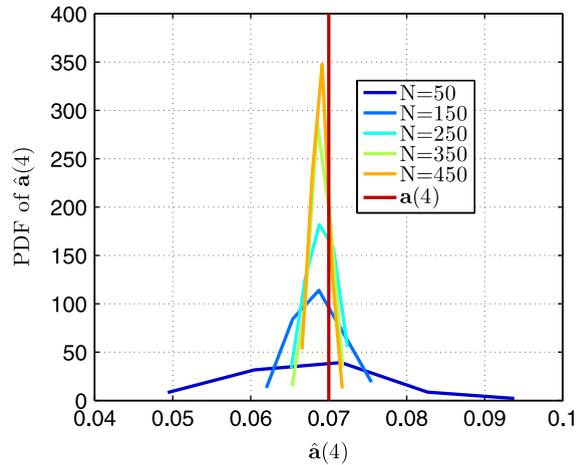


Fig. 6. PDF of $\hat{\mathbf{a}}_4$ for different values of N for an $N \times N$ noncausal ARCH(1, 1) process.

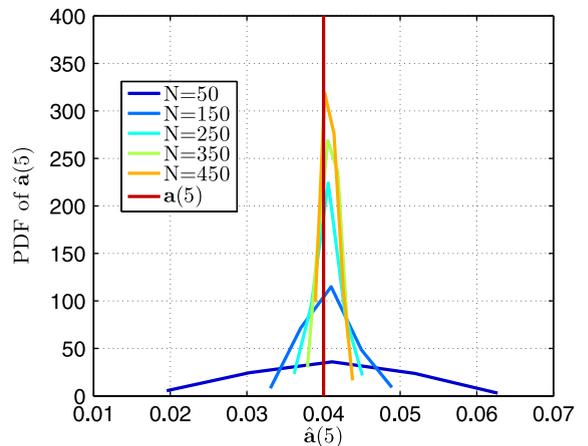


Fig. 7. PDF of $\hat{\mathbf{a}}_5$ for different values of N for an $N \times N$ noncausal ARCH(1, 1) process.

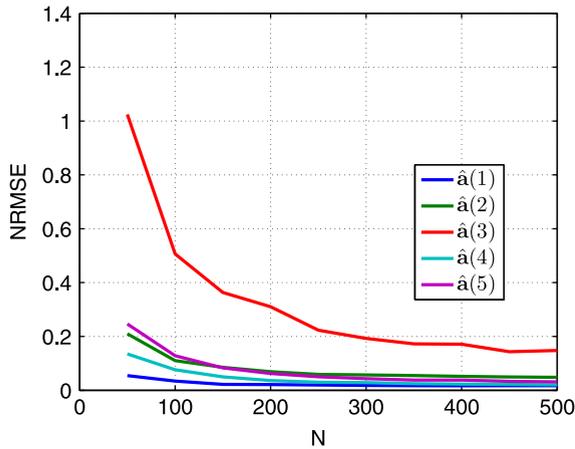


Fig. 8. Estimated normalized root mean square error in parameter estimation of an $N \times N$ noncausal ARCH(1, 1) process as a function of N .

estimation of the proposed estimator converges to zero as the number of data increases which was predicted theoretically in the previous section. From Figs. 3–7 it can be seen that the estimated parameters converge to Gaussian distributions as the number of data increases. This phenomenon was proved in [31,33] for causal one and multi-dimensional ARCH process, respectively. We conjecture that this is also true for noncausal case, where proving this needs further research.

In the second simulation, we compare the performance of the proposed detection algorithm using synthetic data. We have generated 200 noncausal ARCH(1,1) process with the following vector of parameters:

$$\mathbf{a} = [1.00 \ 0.03 \ 0.04 \ 0.05 \ 0.06],$$

using Banach's fixed point theorem with 15 iterations. $\varepsilon(t_1, t_2)$'s were zero mean independent identically random variables uniformly distributed on the interval $[-\sqrt{3}, \sqrt{3}]$. The anomaly was chosen to be a 5×5 image with changing location in each iteration. The anomaly subspace \mathbf{H} was chosen to be $\mathbf{H} = \text{span}(\{\mathbf{H}_i\}_{i=1}^4)$ where each \mathbf{H}_i was a 5×5 matrix consisting of IID zero mean Gaussian random variables with unity variance. The vector ψ locating the anomaly within anomaly subspace was chosen randomly such that each of its elements is distributed normally having zero mean and variance 0.5. We compared the performance of anomaly detection using noncausal ARCH modeling with that of GARCH modeling. The parameters of the noncausal ARCH model were estimated via the proposed TSLS method and the parameters of the GARCH model were estimated using maximum likelihood method. To check the sensitivity of the proposed method to anomalies, we estimated the parameters using both images with and without anomaly. We computed the probability of false alarm (P_{fa}) and the probability of detection (P_d) using 200 Monte-Carlo iterations. The receiver operating characteristic (ROC) curves are depicted in Fig. 9, where in this figure by theoretic we mean using the image without anomaly to estimate the parameters. As is expected, although the performance of detection is decreased if anomalies present, the performance degradation due to anomalies is not very serious in the proposed TSLS method while it is more critical for the ML method. An analytical sensitivity analysis of

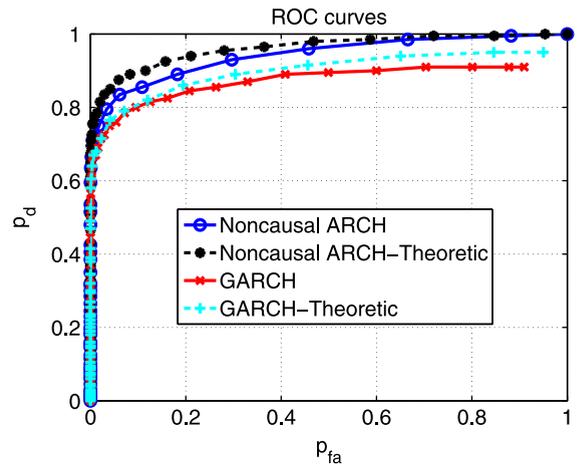


Fig. 9. ROC curves: comparison of detection performance between noncausal ARCH and GARCH modeling for synthetic data.

the proposed parameter estimation algorithm needs further research which is beyond the scope of this paper. It is also obvious that our presented noncausal ARCH model has higher performance comparing to GARCH modeling.

In the third simulation, we demonstrate the validity of the noncausal ARCH model in modeling the residual of the AR part. We also evaluate the performance of our anomaly detection algorithm. We use real side-scan sonar images. The side-scan sonar images presented in this simulation are from the sonar-5 database collected by the Naval Surface Warfare Center Coastal System Station (Panama City, FL). The images are 8-bit gray scale. Elongated sea mines (such as those presented in Fig. 10) are characterized by a bright line (the highlight or echo), corresponding to the scattering effect of mines to the acoustic insonification, and shadow behind them, corresponding to the blocking of sonar waves by mines. Further technical and navigational information about the specific database used is not available. Since the number of mines in our database was limited, it was impossible to estimate the anomaly subspace accurately. Instead we used the images in Fig. 10 to make artificial mines by combining these images linearly using random weights. More specifically, we constructed 100 images by linearly combining these images where the weights are chosen from uniform distribution. The anomaly subspace in each layer (i.e. \mathbf{H}_ℓ) is then estimated from these 100 training images using PCA where we used 10 principle components. Test images were constructed from a real mine-free background together with artificially added mines in random locations. An example of such test image is depicted in Fig. 11(a) together with true detection which is depicted in Fig. 11(b). The likelihood ratio (i.e. L_{CLR}) obtained using noncausal AR-ARCH and AR-GARCH models is represented in Fig. 11(c) and (d), respectively. The vertical and horizontal patch sizes (i.e. T_1 and T_2) are both set to 16. The wavelet depth L is set to 3. We used two dimensional undecimated Haar wavelet transform. Other simulation results show that the wavelet basis has no significant effect on the detection results. We used the AR-ARCH model with the following orders : $p = 1, q = 1, r = 2, s = 2$.

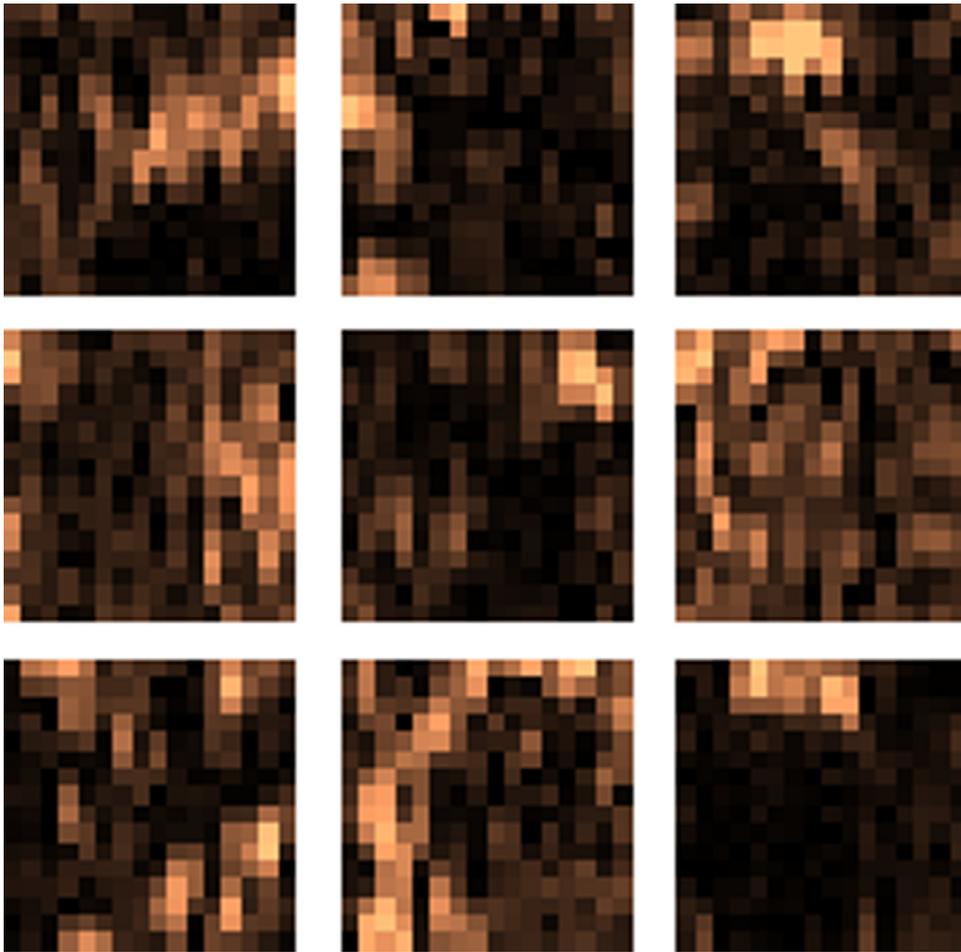


Fig. 10. Sample side scan sonar image with a sea mine (sea mines are characterized by a bright line (the highlight or echo), corresponding to the scattering effect of the mine to the acoustic insonification, and a shadow behind it, corresponding to the blocking of sonar waves by the mine). These images, each 16×16 pixel, are used as training image for obtaining the anomaly subspace.

Fig. 12 provides a comparison between the cumulative distribution function (CDF) of a column stack version of the $\hat{x}_2(t_1, t_2)$ obtained from a typical mine free image from our database and that of a column stack version of a simulated noncausal ARCH(2, 2) process. The parameters of this simulated image are set to those obtained by applying the TSL method to $\hat{x}_2(t_1, t_2)$. Fig. 13 provides quantile–quantile plot (QQ-plot) of these images. The purpose of the QQ plot is to determine whether the samples of two processes come from the same distribution. If the samples do come from the same distribution (same shape), even if one distribution is shifted and rescaled from the other (different location and scale parameters), the plot will be linear. A reference line passing through the first and third quartiles is helpful for judging whether the points are linear. These figures justify the use of noncausal ARCH model in the modeling of the AR residual of the background in the wavelet domain. The same phenomena have been observed in modeling the other layers. We also applied Kolmogorov–Smirnov test on a column stack version of the two dimensional fast Fourier transform (FFT) of $\hat{x}_2(t_1, t_2)$ and the FFT of noncausal ARCH

modeled $\hat{x}_2(t_1, t_2)$. The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The null distribution of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution. Extensive simulations showed that the test chooses the null hypothesis at the 5% significance level.

We compare the performance of our method to that of the method presented in [4]. The ROC curves of both methods are depicted in Fig. 14. These curves are obtained using 100 simulations. As it can be seen from these curves the proposed method has a slightly better performance compared with the method presented in [4]. As mentioned before, the advantage of the proposed method is its lower computational load compared to that of the method proposed in [4].

5. Conclusions

We have used the wavelet transform along with a noncausal AR-ARCH model for anomaly detection in side

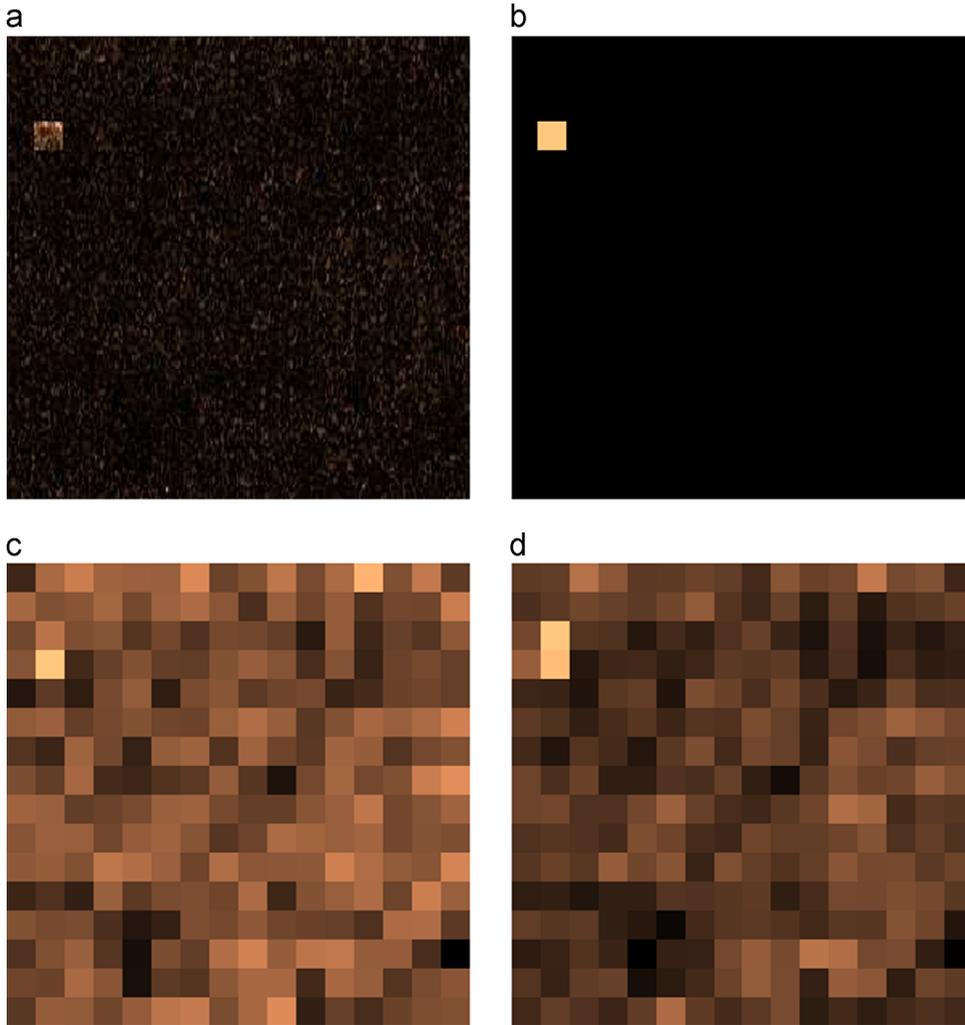


Fig. 11. A sample side-scan sonar image with artificial mine (a), true detection (b), generalized likelihood ratio L_{GLR} obtained by noncausal ARCH modeling (c) and GARCH modeling (d).

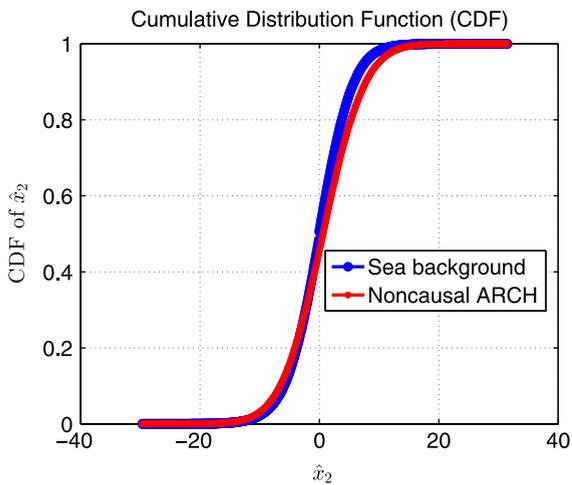


Fig. 12. Comparison of cumulative distribution function of a noncausal ARCH modeling and real sea mine data.

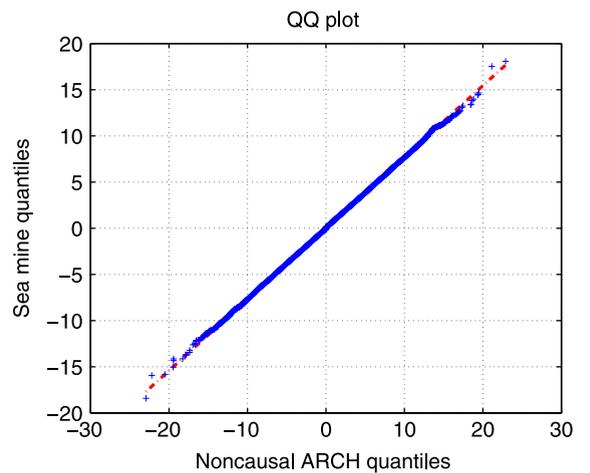


Fig. 13. Comparison of QQ-plot of a noncausal ARCH modeling and real sea mine data.

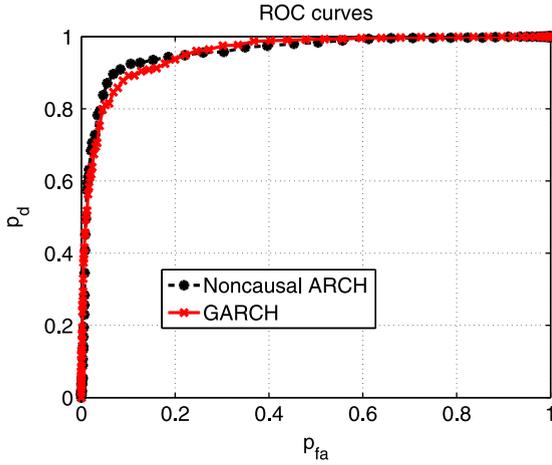


Fig. 14. ROC curves: comparison of detection performance between noncausal ARCH and GARCH modeling for real data.

scan sonar images. We introduced a novel background model (i.e. noncausal AR-ARCH model) and used matched subspace detector (MSD) to detect an anomaly in sonar images. We introduced the two dimensional noncausal ARCH model and obtained sufficient stationarity conditions. We also proposed the TSLs estimator for parameter estimation of this model. This estimator has a closed form obtained by solving two sets of linear equations. We have shown that this estimator is asymptotically consistent. The most important advantage of this parameter estimation method is its low computational complexity. We also presented an anomaly detection algorithm based on noncausal AR-ARCH modeling of the background. The detection algorithm utilizes a matched subspace detector [34] for the background noncausal AR-ARCH model. This model is an extension of the GARCH model used in [4] for anomaly detection. The advantages of the proposed method over that presented in [4] are lower computational load and lower dependency on the orientation of the image. Simulation results have demonstrated the performance of the proposed method.

Acknowledgments

The authors thank the anonymous reviewers for their constructive comments and useful suggestions. This research was supported by Robert H. Hillman Foundation for Global Security – Collaboration Technion and University Northeastern, and by the Israel Science Foundation (Grant no. 1130/11), and by Japan Technion Society Research Fund.

Appendix

In this Appendix, we show that both proposed estimators, i.e. $\hat{\mathbf{a}}_{pr}$ and $\hat{\mathbf{a}}$, are asymptotically unbiased and consistent estimators. In this section, without loss of generality we assume that N and M (i.e. the number of available data in the horizontal and vertical directions, respectively) are equal. For the sake of simplicity, we change the notation as

follows. Instead of double indexing the processes and double sum, we use single indexing and sum. More specifically, by t we mean (t_1, t_2) (hence $y(t) \triangleq y(t_1, t_2)$) and by $\sum_{t=1}^{N^2}$ we mean $\sum_{t_1=1}^N \sum_{t_2=1}^N$. We assume that the noncausal ARCH process satisfies the conditions (3)–(5), hence is mean ergodic. Besides mean ergodicity, we also assume that

$$\begin{aligned} \mathbb{E}\left\{\prod_{i=1}^j y(t_i)\right\} &< \infty \quad \forall t_i; j = 2, 3, 4 \\ \mathbb{E}\{\eta_L(t)^2\} &< \infty \quad \forall t \\ \mathbb{E}\{\eta_U(t)^2\} &< \infty \quad \forall t \end{aligned} \quad (52)$$

and $\mathbb{E}\{\mathbf{y}_L(1)\mathbf{y}_L^T(1)\}$ and $\mathbb{E}\{\mathbf{y}_U(1)\mathbf{y}_U^T(1)\}$ are positive definite matrices having finite elements where $\mathbb{E}\{\cdot\}$ is expectation.

In order to obtain the statistical properties of $\hat{\mathbf{a}}_{pr}$, note that

$$\hat{\mathbf{a}}_{pr}^T = \frac{1}{2} \left(\mathbf{x}_L^T \mathbf{R}_L^{-1} + \mathbf{x}_U^T \mathbf{R}_U^{-1} \right). \quad (53)$$

Using Eqs. (14) and (15) and the definitions of \mathbf{a} , \mathbf{y}_L and \mathbf{y}_U , it can be easily verified that

$$y(t) = 2\mathbf{a}^T \mathbf{y}_L(t) + 2\sigma^2(t)\eta_L(t), \quad (54)$$

$$y(t) = 2\mathbf{a}^T \mathbf{y}_U(t) + 2\sigma^2(t)\eta_U(t). \quad (55)$$

Substituting (22)–(25) into (53), we have

$$\begin{aligned} \hat{\mathbf{a}}_{pr}^T &= \frac{1}{2} \left(\mathbf{x}_L^T \mathbf{R}_L^{-1} + \mathbf{x}_U^T \mathbf{R}_U^{-1} \right) \\ &= \frac{1}{2} \left[\left(\sum_{t=1}^{N^2} y(t)\mathbf{y}_L^T(t) \right) \left(2 \sum_{t=1}^{N^2} \mathbf{y}_L(t)\mathbf{y}_L^T(t) \right)^{-1} \right. \\ &\quad \left. + \left(\sum_{t=1}^{N^2} y(t)\mathbf{y}_U^T(t) \right) \left(2 \sum_{t=1}^{N^2} \mathbf{y}_U(t)\mathbf{y}_U^T(t) \right)^{-1} \right]. \end{aligned} \quad (56)$$

Substituting (54)–(55) into (56), we have

$$\begin{aligned} \hat{\mathbf{a}}_{pr}^T &= \frac{1}{2} \left[\left(\sum_{t=1}^{N^2} (2\mathbf{a}^T \mathbf{y}_L(t) + 2\sigma^2(t)\eta_L(t))\mathbf{y}_L^T(t) \right) \right. \\ &\quad \times \left(2 \sum_{t=1}^{N^2} \mathbf{y}_L(t)\mathbf{y}_L^T(t) \right)^{-1} \\ &\quad \left. + \left(\sum_{t=1}^{N^2} (2\mathbf{a}^T \mathbf{y}_U(t) + 2\sigma^2(t)\eta_U(t))\mathbf{y}_U^T(t) \right) \right. \\ &\quad \left. \times \left(2 \sum_{t=1}^{N^2} \mathbf{y}_U(t)\mathbf{y}_U^T(t) \right)^{-1} \right] \\ &= \mathbf{a}^T + \frac{1}{2} \left[\left(\frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t)\eta_L(t)\mathbf{y}_L^T(t) \right) \right. \\ &\quad \times \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \mathbf{y}_L(t)\mathbf{y}_L^T(t) \right)^{-1} + \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t)\eta_U(t)\mathbf{y}_U^T(t) \right) \\ &\quad \left. \times \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \mathbf{y}_U(t)\mathbf{y}_U^T(t) \right)^{-1} \right]. \end{aligned} \quad (57)$$

Since we assumed that the process is mean ergodic, using the weak law of large numbers (WLLN) we have

$$\frac{1}{N^2} \sum_{t=1}^{N^2} \mathbf{y}_L(t)\mathbf{y}_L^T(t) \xrightarrow{p} \mathbb{E}\{\mathbf{y}_L(\tau)\mathbf{y}_L^T(\tau)\}, \quad (58)$$

$$\frac{1}{N^2} \sum_{t=1}^{N^2} \mathbf{y}_U(t) \mathbf{y}_U^T(t) \xrightarrow{p} \mathbb{E}\{\mathbf{y}_U(\tau) \mathbf{y}_U^T(\tau)\}, \quad (59)$$

where $\tau = [1, 1]$. Since $\sigma^2(t) = \mathbf{a}^T (\mathbf{y}_L(t) + \mathbf{y}_U(t))$, we have

$$\begin{aligned} \frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t) \eta_L(t) \mathbf{y}_L^T(t) &\xrightarrow{p} \mathbb{E}\{\sigma^2(\tau) \eta_L(\tau) \mathbf{y}_L^T(\tau)\} \\ &= \mathbb{E}\{\mathbf{a}^T (\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau)) \eta_L(\tau) \mathbf{y}_L^T(\tau)\} \\ &= \mathbb{E}\{\mathbf{a}^T \mathbf{y}_L(\tau) \eta_L(\tau) \mathbf{y}_L^T(\tau)\} + \mathbb{E}\{\mathbf{a}^T \mathbf{y}_U(\tau) \eta_L(\tau) \mathbf{y}_L^T(\tau)\} \\ &= \mathbb{E}\{\mathbf{a}^T \mathbf{y}_L(\tau) \mathbf{y}_L^T(\tau)\} \mathbb{E}\{\eta_L(\tau)\} + \mathbb{E}\{\mathbf{a}^T \mathbf{y}_U(\tau) \mathbf{y}_L^T(\tau)\} \mathbb{E}\{\eta_L(\tau)\} = \mathbf{0}, \end{aligned} \quad (60)$$

where we have used the fact that $\mathbb{E}\{\eta_L(1)\} = \mathbb{E}\{\eta_U(1)\} = 0$ and $\eta_L(t)$ is uncorrelated with $\mathbf{y}_U(\tau)$ and $\mathbf{y}_U(\tau)$, as discussed before. Similarly, it can be shown that

$$\frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t) \eta_U(t) \mathbf{y}_U^T(t) \xrightarrow{p} \mathbb{E}\{\sigma^2(\tau) \eta_U(\tau) \mathbf{y}_U^T(\tau)\} = \mathbf{0}. \quad (61)$$

Combining Eqs. (57)–(61) it follows that

$$\lim_{N \rightarrow \infty} \hat{\mathbf{a}}_{pr} - \mathbf{a} = \mathbf{0} \quad \text{almost surely} \quad (62)$$

and hence

$$\lim_{N \rightarrow \infty} \mathbb{E}\{\hat{\mathbf{a}}_{pr} - \mathbf{a}\} = \mathbf{0}, \quad (63)$$

which shows that $\hat{\mathbf{a}}_{pr}$ is an asymptotically unbiased estimator. In order to show that $\hat{\mathbf{a}}_{pr}$ is an asymptotically consistent estimator, it suffices to show that

$$\left(\frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t) \eta_L(t) \mathbf{y}_L(t) \right) \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t) \eta_L(t) \mathbf{y}_L^T(t) \right) \xrightarrow{p} \mathbf{0} \quad (64)$$

$$\left(\frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t) \eta_U(t) \mathbf{y}_U(t) \right) \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t) \eta_U(t) \mathbf{y}_U^T(t) \right) \xrightarrow{p} \mathbf{0} \quad (65)$$

$$\left(\frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t) \eta_L(t) \mathbf{y}_L(t) \right) \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t) \eta_U(t) \mathbf{y}_U^T(t) \right) \xrightarrow{p} \mathbf{0}. \quad (66)$$

In order to show (64), note that

$$\begin{aligned} &\left(\frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t) \eta_L(t) \mathbf{y}_L(t) \right) \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \sigma^2(t) \eta_L(t) \mathbf{y}_L^T(t) \right) \\ &= \frac{1}{N^4} \sum_{t_a=1}^{N^2} \sum_{t_b=1}^{N^2} \sigma^2(t_a) \sigma^2(t_b) \eta_L(t_a) \eta_L(t_b) \mathbf{y}_L(t_a) \mathbf{y}_L^T(t_b) \\ &= \frac{1}{N^4} \left(\sum_{t=1}^{N^2} \sigma^4(t) \eta_L^2(t) \mathbf{y}_L(t) \mathbf{y}_L^T(t) \right) \\ &\quad + \sum_{t_a=1}^{N^2} \sum_{t_b=1, t_b \neq t_a}^{N^2} \sigma^2(t_a) \sigma^2(t_b) \eta_L(t_a) \eta_L(t_b) \mathbf{y}_L(t_a) \mathbf{y}_L^T(t_b) \\ &= \frac{1}{N^4} \sum_{t=1}^{N^2} (\mathbf{a}^T (\mathbf{y}_L(t) + \mathbf{y}_U(t)))^2 \eta_L^2(t) \mathbf{y}_L(t) \mathbf{y}_L^T(t) \\ &\quad + \frac{1}{N^4} \sum_{t_a=1}^{N^2} \sum_{t_b=1, t_b \neq t_a}^{N^2} \sigma^2(t_a) \sigma^2(t_b) \eta_L(t_a) \eta_L(t_b) \mathbf{y}_L(t_a) \mathbf{y}_L^T(t_b). \end{aligned} \quad (67)$$

In order to show that the first term in Eq. (67) converges to zero in probability, note that because of WLLN

$$\begin{aligned} \frac{1}{N^2} \sum_{t=1}^{N^2} (\mathbf{a}^T (\mathbf{y}_L(t) + \mathbf{y}_U(t)))^2 \eta_L^2(t) \mathbf{y}_L(t) \mathbf{y}_L^T(t) \\ \xrightarrow{p} \mathbb{E}\{(\mathbf{a}^T (\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau)))^2 \eta_L^2(\tau) \mathbf{y}_L(\tau) \mathbf{y}_L^T(\tau)\} < \infty, \end{aligned} \quad (68)$$

and is finite since we assumed that the process $\mathbf{y}(t)$ satisfies conditions (52). Hence, the first term in Eq. (67) converges to zero in probability. To show that the second term in Eq. (67) converges to zero in probability, note that

$$\begin{aligned} \frac{1}{N^4} \sum_{t_a=1}^{N^2} \sum_{t_b=1, t_b \neq t_a}^{N^2} \sigma^2(t_a) \eta_L(t_a) \eta_L(t_b) \mathbf{y}_L(t_a) \mathbf{y}_L^T(t_b) \\ = \frac{1}{N^2} \sum_{t_a=1}^{N^2} \sigma^2(t_a) \eta_L(t_a) \mathbf{y}_L(t_a) \frac{1}{N^2} \sum_{t_b=1, t_b \neq t_a}^{N^2} \sigma^2(t_b) \eta_L(t_b) \mathbf{y}_L^T(t_b), \end{aligned} \quad (69)$$

where both sum can be shown to converge to zero using the same technique that resulted in (60). Similarly, it can be easily verified that (65) and (66) hold, hence $\hat{\mathbf{a}}_{pr}$ is an asymptotically consistent estimate of the parameters.

To obtain the statistical properties of $\hat{\mathbf{a}}$ note that

$$\hat{\mathbf{a}}^T = \frac{1}{2} (\bar{\mathbf{x}}_L^T \mathbf{R}_L^{-1} + \bar{\mathbf{x}}_U^T \mathbf{R}_U^{-1}). \quad (70)$$

Substituting (33)–(36) into (70) and using (29)–(31), we have

$$\begin{aligned} \hat{\mathbf{a}}^T = \frac{1}{2} \left[\left(\sum_{t=1}^{N^2} \frac{\mathbf{y}(t) \mathbf{y}_L^T(t)}{\hat{\sigma}^4} (t) \right) \left(2 \sum_{t=1}^{N^2} \frac{\mathbf{y}_L(t) \mathbf{y}_L^T(t)}{\hat{\sigma}^4} (t) \right)^{-1} \right. \\ \left. + \left(\sum_{t=1}^{N^2} \frac{\mathbf{y}(t) \mathbf{y}_U^T(t)}{\hat{\sigma}^4} (t) \right) \left(2 \sum_{t=1}^{N^2} \frac{\mathbf{y}_U(t) \mathbf{y}_U^T(t)}{\hat{\sigma}^4} (t) \right)^{-1} \right]. \end{aligned} \quad (71)$$

Inserting (54)–(55) into (71), we have

$$\begin{aligned} \hat{\mathbf{a}}^T = \frac{1}{2} \left[\left(\sum_{t=1}^{N^2} \frac{(\mathbf{a}^T \mathbf{y}_L(t) + \sigma^2(t) \eta_L(t)) \mathbf{y}_L^T(t)}{\hat{\sigma}^4} (t) \right) \right. \\ \times \left(\sum_{t=1}^{N^2} \frac{\mathbf{y}_L(t) \mathbf{y}_L^T(t)}{\hat{\sigma}^4} (t) \right)^{-1} \\ \left. + \left(\sum_{t=1}^{N^2} \frac{(\mathbf{a}^T \mathbf{y}_U(t) + \sigma^2(t) \eta_U(t)) \mathbf{y}_U^T(t)}{\hat{\sigma}^4} (t) \right) \right. \\ \times \left. \left(\sum_{t=1}^{N^2} \frac{\mathbf{y}_U(t) \mathbf{y}_U^T(t)}{\hat{\sigma}^4} (t) \right)^{-1} \right] \\ = \mathbf{a}^T + \frac{1}{2} \left[\left(\sum_{t=1}^{N^2} \frac{\sigma^2(t) \eta_L(t) \mathbf{y}_L^T(t)}{\hat{\sigma}^4} (t) \right) \left(\sum_{t=1}^{N^2} \frac{\mathbf{y}_L(t) \mathbf{y}_L^T(t)}{\hat{\sigma}^4} (t) \right)^{-1} \right. \\ \left. + \left(\sum_{t=1}^{N^2} \frac{\sigma^2(t) \eta_U(t) \mathbf{y}_U^T(t)}{\hat{\sigma}^4} (t) \right) \left(\sum_{t=1}^{N^2} \frac{\mathbf{y}_U(t) \mathbf{y}_U^T(t)}{\hat{\sigma}^4} (t) \right)^{-1} \right]. \end{aligned} \quad (72)$$

Using a very similar approach as used in [33] it can be shown that, in (72) we can substitute $\hat{\sigma}^2(t)$ with $\sigma^2(t)$ and we get

$$\hat{\mathbf{a}}^T - \mathbf{a}^T = \frac{1}{2} \left[\left(\sum_{t=1}^{N^2} \frac{\eta_L(t) \mathbf{y}_L^T(t)}{\sigma^2(t)} \right) \left(\sum_{t=1}^{N^2} \frac{\mathbf{y}_L(t) \mathbf{y}_L^T(t)}{\sigma^2(t)} \right)^{-1} \right.$$

$$+ \left(\sum_{t=1}^{N^2} \frac{\eta_U(t) \mathbf{y}_U^T(t)}{\sigma^2(t)} \right) \left(\sum_{t=1}^{N^2} \frac{\mathbf{y}_U(t) \mathbf{y}_U^T(t)}{\sigma^2(t)} \right)^{-1} \Big] + o_p(1). \quad (73)$$

Since $\sigma^2(t) = \mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))$ we have

$$\begin{aligned} \hat{\mathbf{a}}^T - \mathbf{a}^T &= \frac{1}{2} \left[\left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_L(t) \mathbf{y}_L^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right) \right. \\ &\quad \times \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\mathbf{y}_L(t) \mathbf{y}_L^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right)^{-1} \\ &\quad + \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_U(t) \mathbf{y}_U^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right) \\ &\quad \left. \times \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\mathbf{y}_U(t) \mathbf{y}_U^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right)^{-1} \right] + o_p(1). \quad (74) \end{aligned}$$

With mean ergodicity assumption and use of WLLN, we have

$$\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\mathbf{y}_L(t) \mathbf{y}_L^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \xrightarrow{p} \mathbb{E} \left\{ \frac{\mathbf{y}_L(\tau) \mathbf{y}_L^T(\tau)}{\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau))} \right\}, \quad (75)$$

$$\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\mathbf{y}_U(t) \mathbf{y}_U^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \xrightarrow{p} \mathbb{E} \left\{ \frac{\mathbf{y}_U(\tau) \mathbf{y}_U^T(\tau)}{\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau))} \right\}. \quad (76)$$

Note that since we assumed that if c_0 is absolutely positive, then $\sigma^2(\tau) > 0$ and because of the moment constraints (52) we assumed that these quantities are finite. Furthermore, we have

$$\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_L(t) \mathbf{y}_L^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \xrightarrow{p} \mathbb{E} \left\{ \frac{\eta_L(\tau) \mathbf{y}_L^T(\tau)}{\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau))} \right\}, \quad (77)$$

$$\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_U(t) \mathbf{y}_U^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \xrightarrow{p} \mathbb{E} \left\{ \frac{\eta_U(\tau) \mathbf{y}_U^T(\tau)}{\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau))} \right\}. \quad (78)$$

Using the fact that $\mathbb{E}\{\eta_L(1)\} = \mathbb{E}\{\eta_U(1)\} = 0$ and $\eta_L(t)$ is uncorrelated with $\mathbf{y}_U(\tau)$ and $\mathbf{y}_U(\tau)$, we have

$$\mathbb{E} \left\{ \frac{\eta_L(\tau) \mathbf{y}_L^T(\tau)}{\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau))} \right\} = \mathbb{E} \left\{ \frac{\mathbf{y}_L^T(\tau)}{\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau))} \right\} \mathbb{E}\{\eta_L\} = \mathbf{0}, \quad (79)$$

$$\mathbb{E} \left\{ \frac{\eta_U(\tau) \mathbf{y}_U^T(\tau)}{\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau))} \right\} = \mathbb{E} \left\{ \frac{\eta_U(\tau) \mathbf{y}_U^T(\tau)}{\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau))} \right\} \mathbb{E}\{\eta_U\} = \mathbf{0}. \quad (80)$$

Combining Eqs. (72)–(80) it follows that

$$\lim_{N \rightarrow \infty} \hat{\mathbf{a}} - \mathbf{a} = \mathbf{0}, \quad (81)$$

and hence

$$\lim_{N \rightarrow \infty} \mathbb{E}\{\hat{\mathbf{a}} - \mathbf{a}\} = \mathbf{0}, \quad (82)$$

which shows that $\hat{\mathbf{a}}$ is an asymptotically unbiased estimator.

In order to show that $\hat{\mathbf{a}}$ is an asymptotically consistent estimator, it suffices to show that

$$\left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_L(t) \mathbf{y}_L(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right) \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_L(t) \mathbf{y}_L^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right) \xrightarrow{p} \mathbf{0}, \quad (83)$$

$$\left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_L(t) \mathbf{y}_U(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right) \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_L(t) \mathbf{y}_U^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right) \xrightarrow{p} \mathbf{0}, \quad (84)$$

$$\left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_L(t) \mathbf{y}_U(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right) \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_L(t) \mathbf{y}_L^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right) \xrightarrow{p} \mathbf{0}. \quad (85)$$

In order to show that (83) is correct, note that

$$\begin{aligned} &\left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_L(t) \mathbf{y}_L(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right) \left(\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_L(t) \mathbf{y}_L^T(t)}{\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t))} \right) \\ &= \frac{1}{N^4} \sum_{t_a=1}^{N^2} \sum_{t_b=1}^{N^2} \frac{\eta_L(t_a) \mathbf{y}_L(t_a)}{\mathbf{a}^T(\mathbf{y}_L(t_a) + \mathbf{y}_U(t_a))} \frac{\eta_L(t_b) \mathbf{y}_L^T(t_b)}{\mathbf{a}^T(\mathbf{y}_L(t_b) + \mathbf{y}_U(t_b))} \\ &= \frac{1}{N^4} \sum_{t=1}^{N^2} \frac{\eta_L(t)^2 \mathbf{y}_L(t) \mathbf{y}_L^T(t)}{(\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t)))^2} \\ &\quad + \frac{1}{N^4} \sum_{t_a=1}^{N^2} \sum_{t_b=1, t_b \neq t_a}^{N^2} \frac{\eta_L(t_a) \mathbf{y}_L(t_a)}{\mathbf{a}^T(\mathbf{y}_L(t_a) + \mathbf{y}_U(t_a))} \frac{\eta_L(t_b) \mathbf{y}_L^T(t_b)}{\mathbf{a}^T(\mathbf{y}_L(t_b) + \mathbf{y}_U(t_b))}. \quad (86) \end{aligned}$$

In order to show that (86) converges to zero in probability, note that, for the first term in (86) using WLLN we have

$$\frac{1}{N^2} \sum_{t=1}^{N^2} \frac{\eta_L(t)^2 \mathbf{y}_L(t) \mathbf{y}_L^T(t)}{(\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t)))^2} \xrightarrow{p} \mathbb{E} \left\{ \frac{\eta_L(\tau)^2 \mathbf{y}_L(\tau) \mathbf{y}_L^T(\tau)}{(\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau)))^2} \right\}, \quad (87)$$

which is finite because of the moment constraints (i.e. (52)) we assumed and the fact that $(\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau)))^2 > 0$ since $c_0 > 0$. Hence,

$$\frac{1}{N^4} \sum_{t=1}^{N^2} \frac{\eta_L(t)^2 \mathbf{y}_L(t) \mathbf{y}_L^T(t)}{(\mathbf{a}^T(\mathbf{y}_L(t) + \mathbf{y}_U(t)))^2} \xrightarrow{p} \mathbf{0}. \quad (88)$$

In order to show that the second term in (86) converges to zero in probability, note that,

$$\begin{aligned} &\frac{1}{N^4} \sum_{t_a=1}^{N^2} \sum_{t_b=1, t_b \neq t_a}^{N^2} \frac{\eta_L(t_a) \mathbf{y}_L(t_a)}{\mathbf{a}^T(\mathbf{y}_L(t_a) + \mathbf{y}_U(t_a))} \frac{\eta_L(t_b) \mathbf{y}_L^T(t_b)}{\mathbf{a}^T(\mathbf{y}_L(t_b) + \mathbf{y}_U(t_b))} \\ &= \frac{1}{N^2} \sum_{t_a=1}^{N^2} \left(\frac{\eta_L(t_a) \mathbf{y}_L(t_a)}{\mathbf{a}^T(\mathbf{y}_L(t_a) + \mathbf{y}_U(t_a))} \right) \\ &\quad \times \frac{1}{N^2} \sum_{t_b=1, t_b \neq t_a}^{N^2} \left(\frac{\eta_L(t_b) \mathbf{y}_L^T(t_b)}{\mathbf{a}^T(\mathbf{y}_L(t_b) + \mathbf{y}_U(t_b))} \right), \quad (89) \end{aligned}$$

and using WLLN and the fact that $\mathbb{E}\{\eta_L(1)\} = 0$ and $\eta_L(t)$ is uncorrelated with $\mathbf{y}_U(\tau)$ and $\mathbf{y}_U(\tau)$, we have

$$\begin{aligned} &\frac{1}{N^2} \sum_{t_a=1}^{N^2} \left(\frac{\eta_L(t_a) \mathbf{y}_L(t_a)}{\mathbf{a}^T(\mathbf{y}_L(t_a) + \mathbf{y}_U(t_a))} \right) \xrightarrow{p} \mathbb{E} \left\{ \frac{\eta_L(\tau) \mathbf{y}_L(\tau)}{\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau))} \right\} \\ &= \mathbb{E} \left\{ \frac{\mathbf{y}_L(\tau)}{\mathbf{a}^T(\mathbf{y}_L(\tau) + \mathbf{y}_U(\tau))} \right\} \mathbb{E}\{\eta_L(\tau)\} = \mathbf{0}. \quad (90) \end{aligned}$$

Hence (86) converges to zero in probability and therefore (83) holds. Using the similar approach it can be verified that (84)–(85) hold, and hence $\hat{\mathbf{a}}$ is an asymptotically consistent estimator. This is the main result of this paper.

References

[1] M. Zontak, I. Cohen, Defect detection in patterned wafers using multichannel scanning electron microscope, *Signal Process.* 89 (8) (2009) 1511–1520.

- [2] M. Zontak, I. Cohen, Defect detection in patterned wafers using anisotropic kernels, *Mach. Vis. Appl.* 21 (2) (2010) 129–141.
- [3] C. Spence, L. Parra, P. Sajda, Detection, synthesis and compression in mammographic image analysis with a hierarchical image probability model, in: *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, IEEE Computer Society, 2001.
- [4] A. Noiboar, I. Cohen, Anomaly detection based on wavelet domain GARCH random field modeling, *IEEE Trans. Geosci. Remote Sens.* 45 (2007) 1361–1373.
- [5] I.G. Kazantsev, I. Lemahieu, G.I. Salov, R. Denys, Statistical detection of defects in radiographic images in nondestructive testing, *Signal Process.* 82 (2002) 791–801.
- [6] S.M. Schweizer, J.M.F. Moura, Hyperspectral imagery: clutter adaptation in anomaly detection, *IEEE Trans. Inf. Theory* 46 (2002) 1855–1871.
- [7] A.F. Laine, S. Schuler, J. Fan, W. Huda, Mammographic feature enhancement by multiscale analysis, *IEEE Trans. Med. Imaging* 13 (1994) 725–740.
- [8] R.N. Strickland, H.I. Hahn, Wavelet transform methods for object detection and recovery, *IEEE Trans. Image Process.* 6 (1997) 724–735.
- [9] X.G. Xia, C.G. Boncelet, G.R. Arce, A multiresolution watermark for digital images, in: *Proceedings of the International Conference on Image Processing (ICIP '97)*, IEEE Computer Society, Washington DC, USA, 1997, pp. 497–511.
- [10] E.A. Ashton, Detection of subpixel anomalies in multispectral infrared imagery using an adaptive Bayesian classifier, *IEEE Trans. Geosci. Remote Sens.* 36 (1998) 506–517.
- [11] D.W.J. Stein, S.G. Beaven, L.E. Hoff, E.M. Winter, A.P. Schaum, A. D. Stocker, Anomaly detection from hyperspectral imagery, *IEEE Signal Process. Mag.* 19 (2002) 58–69.
- [12] A.S. Willsky, Multiresolution Markov models for signal and image processing, *Proc. IEEE* 90 (2002) 1396–1458.
- [13] P. Loubaton, Champs stationnaires au sens large sur Z^2 : Propriétés structurelles et modèles paramétriques, *Ecole nationale supérieure des télécommunications*, 1989.
- [14] M. Marc, Spatial linear processes, *Stoch. Models* 4 (1) (1988) 45–75.
- [15] T. Bollerslev, Generalized autoregressive conditional heteroscedasticity, *J. Econom.* 31 (1986) 307–327.
- [16] Engle R.F., Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica: J. Econom. Soc.* (1982) 987–1007.
- [17] I. Cohen, Speech spectral modeling and enhancement based on autoregressive conditional heteroscedasticity models, *Signal Process.* 86 (2006) 698–709.
- [18] H.K. Solvanga, K. Ishizukac, M. Fujimotoc, Voice activity detection based on adjustable linear prediction and GARCH models, *Speech Commun.* 50 (2008) 476–486.
- [19] S. Mousazadeh, I. Cohen, AR-GARCH in presence of noise: parameter estimation and its application to voice activity detection, *IEEE Trans. Audio Speech Lang. Process.* 19 (4) (2011) 916–926.
- [20] M. Abdolahi, H. Amindavar, GARCH coefficients as feature for speech recognition in persian isolated digit, in: *Proceedings of 30th IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2005*, pp. 957–960.
- [21] M. Amirmazlaghani, H. Amindavar, A. Moghaddamjoo, Speckle suppression in SAR images using the 2-D GARCH model, *IEEE Trans. Image Process.* 18 (2) (2009) 250–259.
- [22] M. Amirmazlaghani, H. Amindavar, Image denoising using two-dimensional GARCH model, in: *14th International Workshop on Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference Focused on Speech and Image Processing, Multimedia Communications and Services, IEEE, 2007*, pp. 397–400.
- [23] S. Mousazadeh, I. Cohen, Anomaly detection in sonar images based on wavelet domain noncausal AR-ARCH random field modeling, in: *Proceedings of 26th IEEE Convention of Electrical and Electronics Engineers in Israel, IEEE-2010, Eilat, Israel, 17–20 November 2010*, 2010.
- [24] P.Y. Zhao, D.R. Yu, An unbiased and computationally efficient LS estimation method for identifying parameters of 2-d noncausal SAR models, *IEEE Trans. Signal Process.* 41 (1993) 849–857.
- [25] A.K. Jain, J.R. Jain, Partial differential equations and finite difference methods in image processing, part ii: image restoration, *IEEE Trans. Autom. Control* 23 (1978) 817–834.
- [26] A.M. Tekalp, H. Kaufman, J.W. Woods, Identification of image and blur parameters for the restoration of noncausal blurs, *IEEE Trans. Acoust. Speech Signal Process.* 34 (1986) 963–972.
- [27] J.H. McClellan, Multidimensional spectral estimation, *Proc. IEEE* 70 (1982) 1029–1039.
- [28] C.W. Therrien, T.F. Quatieri, D.E. Dudgeon, Statistical model-based algorithms for image analysis, *Proc. IEEE* 74 (1986) 532–607.
- [29] E. Kreyszig, *Introductory Functional Analysis with Applications*, first ed. Wiley, New York, 1989.
- [30] A.A. Weiss, Asymptotic theory of ARCH models estimation and testing, *Econom. Theory* 2 (1989) 107–131.
- [31] A. Bose, K. Mukherjee, Estimating the ARCH parameters by solving linear equations, *J. Time Ser. Anal.* 24 (2003) 127–136.
- [32] S. Mousazadeh, M. Karimi, ARCH parameter estimation via constrained two stage least squares method, in: *Proceedings of IEEE International Symposium on Signal Processing and its Application (ISSPA), 2007*, pp. 1–4.
- [33] S. Mousazadeh, M. Karimi, Estimating multivariate ARCH parameters by two-stage least-squares method, *Signal Process.* 89 (2009) 921–932.
- [34] L.L. Scharf, B. Friedlander, Matched subspace detectors, *IEEE Trans. Signal Process.* 42 (1994) 2146–2157.
- [35] D.F. Walnut, *An Introduction to Wavelet Analysis*, Birkhäuser, Boston, 2004.