IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 17, 2018, accepted January 11, 2019. Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2019.2893110

Robust Spectrum-Based Comparison of Multivariate Complex Random Signals

JITENDRA K. TUGNAIT, (Life Fellow, IEEE)
Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849, USA
e-mail: tugnajk@auburn.edu

This work was supported by the National Science Foundation under Grant CCF-1617610.

ABSTRACT We consider the problem of comparing two complex multivariate random signal realizations, possibly contaminated with additive outliers, to ascertain whether they have identical power spectral densities. For clean data (i.e., known to be outlier free), a binary hypothesis testing formulation in frequency-domain, utilizing the estimated power spectral density matrices, has been proposed in the literature, and it results in a generalized likelihood ratio test (GLRT). In this paper, we first present an alternative, principled derivation of the existing GLRT using the asymptotic distribution of a frequency-domain sufficient statistic, based on the discrete Fourier transform of the two signal realizations. In order to robustify this GLRT in the presence of additive outliers, we first exploit an existing robust estimator of multivariate scatter to detect the outliers, and subsequently, to clean the data. The existing GLRT is then applied to the cleaned signal realizations. The approach is illustrated through simulations. The considered problem has applications in diverse areas, including user authentication in wireless networks with multi-antenna receivers.

INDEX TERMS Generalized likelihood ratio test, hypothesis testing, multichannel signal detection, multiple antennas, spectral analysis, wireless user authentication.

I. INTRODUCTION

We consider the problem of comparison of two realizations (sample functions) of some random signals (time series) to ascertain if they are realizations of the same random signal. This is formulated as a binary hypothesis testing problem. The following applications motivate consideration of this problem: (i) user authentication for wireless network security enhancement at the physical layer [1], [2], and (ii) spectrum sensing (looking for presence/absence of primary users (PUs) in spectral bands) in cognitive radio (CR) networks (based on the two-window approach of [3]) [4]. Recent general articles on comparison of random signals include [5]–[7], and references therein, where a variety of other applications have been mentioned: earthquake-explosion discrimination [8], financial portfolio management, clustering of environmental data [7], analysis of photometry data [5], and development of climate reference stations [6]. In [5] and [6], only real-valued scalar time series is investigated, and in [7] and [8], real-valued multivariate processes are addressed. In wireless user authentication at the physical layer [1], [2] and spectrum sensing based on the two-window approach [3], [4], the data are complex-valued.

Comparison of complex-valued multivariate random signals using their power spectral densities (PSDs) has been investigated in [9] to derive a generalized likelihood ratio test (GLRT), since the PSDs of the two signals are unknown. In [10] real-valued scalar signals of unequal lengths have been considered. In [11] the theory of [9] has been extended to complex random signals of unequal lengths (sample sizes).

In this paper we robustify the approach of [9] to allow additive outliers in the data. None of the prior works on comparing random signals, including [5]–[7], [9], consider the case of data contaminated with outliers. It is well known (see, e.g., [12]–[14]) that a small fraction of outliers (i.e., gross errors in a small fraction of the observations) when unaccounted for or not modeled, can dramatically affect the final result in any statistical inference (estimation or detection) problem. In the context of comparison of random signals formulated as a binary hypothesis testing problem, outliers can have a deleterious effect on the power of the test (probability of detection).

Another contribution of this paper is to derive the GLRT of [9] in a principled way. Our derivation of the GLRT in this paper is significantly different from that in [9], where the starting point is the estimated PSDs of the two signals, and the frequencies at which the PSDs are estimated, are picked in an ad hoc fashion. In this paper, we first develop a frequency-domain sufficient statistic (Sec. II-C), and then

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

under certain sufficient conditions ((A1)-(A3) in Sec. II-D), we derive the GLRT based on the frequency-domain sufficient statistic, in a principled way. The estimates of the two PSDs follow from the frequency-domain sufficient statistic as maximum likelihood (ML) estimates, as in (22) of Sec. III.

The paper is organized as follows. In Sec. II, we introduce notation and state the binary hypothesis testing problem of interest in this paper. We also develop a frequency-domain sufficient statistic (Sec. II-C), introduce certain sufficient conditions ((A1)-(A3) in Sec. II-D) needed later, and review asymptotic distribution of the sufficient statistic as number of measurement samples become large. We derive the GLRT in Sec. III based on the frequency-domain sufficient statistic and its asymptotic distribution, under the assumption that data are free of any outliers. Since the GLRT statistic is the same in this paper and in [9], all other details, such as analytical threshold calculation via [9, Th. 1] and corroboration of the theory via simulations, given in [9] apply here as well. In Sec. IV-A we first introduce the widely used additive outlier model for the measurements. This model is not essential to robust processing, but is one of the useful models frequently used in the literature [12], [13], and is used in our simulations. Then we review the method of [14] for robust estimation of multivariate scatter in Sec. IV-B. It is used in Sec. IV-C for detection of outliers in the two realizations of the random signals. In Sec. IV-D the detected outliers are ''cleaned'' via median filtering, or simply clipped. So cleaned datasets are then used for random signal comparison using the method of [9], as if the cleaned datasets are the original datasets. Simulation examples are presented in Sec. V in support of the proposed approach.

II. PRELIMINARIES AND SYSTEM MODEL

Here we first introduce notation in Sec. II-A, and state the binary hypothesis testing problem of interest in this paper in Sec. II-B. Then we motivate a frequency-domain sufficient statistic in Sec. II-C. Certain reasonable sufficient conditions are presented in Sec. II-D, for use later in the derivation of the GLRT. In Sec. II-E we review asymptotic distribution of the sufficient statistic as number of measurement samples become large.

A. NOTATION

We use $\boldsymbol{S} \succeq 0$ and $\boldsymbol{S} \succ 0$ to denote that Hermitian (or symmetric) matrix $\boldsymbol{S}$ is positive semi-definite and positive definite, respectively. For a square matrix $\boldsymbol{A}$, $|\boldsymbol{A}|$ and $\text{etr}(\boldsymbol{A})$ denote the determinant and the exponential of the trace of $\boldsymbol{A}$, respectively, i.e., $\text{etr}(\boldsymbol{A}) = \exp(\text{tr}(\boldsymbol{A}))$, $\boldsymbol{B}_{ij}$ is the ijth element of matrix $\boldsymbol{B}$, and $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. The superscripts $*$, $\top$ and H denote the complex conjugate, transpose and Hermitian (conjugate transpose) operations, respectively, and $\mathbb{E}$ denotes the expectation operation. The set of integers 0 through n is denoted by $[0, n]$. The sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. The notation $y = \mathcal{O}(g(x))$ means that there exists some finite real number $b > 0$ such that $\lim_{x\to\infty} |y/g(x)| \leq b$. For $\boldsymbol{x} \in \mathbb{C}^p$, $\text{Re}\{\boldsymbol{x}\} \in \mathbb{R}^p$ and $\text{Im}\{\boldsymbol{x}\} \in \mathbb{R}^p$ denote the real and the imaginary parts of $\boldsymbol{x}$.

The notation χ^2_n represents central chi-square distribution with n degrees of freedom. The notation $\boldsymbol{x} \sim \mathcal{N}_c(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes a random vector $\boldsymbol{x}$ that is proper complex Gaussian with mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$, and $\boldsymbol{x} \overset{a}{\sim} \mathcal{N}_c(\boldsymbol{m}, \boldsymbol{\Sigma})$ implies that $\boldsymbol{x}$ is asymptotically $\mathcal{N}_c(\boldsymbol{m}, \boldsymbol{\Sigma})$ as number of measurements tend to ∞. The notation $\overset{a}{\sim}$ applies to other distributions (χ^2, Wishart, etc) as well. Also, for real $\boldsymbol{x}$, $\boldsymbol{x} \sim \mathcal{N}_r(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes a real random vector $\boldsymbol{x}$ that is Gaussian with mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$. The abbreviations w.p.1. and i.i.d. stand for with probability one, and independent and identically distributed, respectively.

B. BINARY HYPOTHESES

We consider two zero-mean, proper, complex multivariate (dimension p) stationary random signals $\{\boldsymbol{x}(t)\}$ and $\{\boldsymbol{y}(t)\}$ with $p \times p$ PSD matrices $\boldsymbol{S}_x(f)$ and $\boldsymbol{S}_y(f)$, respectively. Recall that $\{\boldsymbol{x}(t)\}$ is proper if $\mathbb{E}\{\boldsymbol{x}(t)\boldsymbol{x}^\top(t+\tau)\} \equiv 0$ [15]. We observe $\{\boldsymbol{x}(t)\}$ and $\{\boldsymbol{y}(t)\}$ for $t = 0, 1, \cdots, N-1$ (N samples each), and the two signals are assumed to be independent. Let $\mathcal{H}_0$ denote the null hypothesis that the PSDs of $\{\boldsymbol{x}(t)\}$ and $\{\boldsymbol{y}(t)\}$ are identical, and let $\mathcal{H}_1$ denote the alternative that $\boldsymbol{S}_x(f) \not\equiv \boldsymbol{S}_y(f)$. Then we have the following binary hypothesis testing problem

$$\begin{aligned} &\mathcal{H}_0 : \boldsymbol{S}_x(f) = \boldsymbol{S}_y(f), \quad 0 \leq f \leq 1.0 \\ &\mathcal{H}_1 : \mathcal{H}_0^c = \text{complement of } \mathcal{H}_0. \end{aligned} \tag{1}$$

C. FREQUENCY-DOMAIN SUFFICIENT STATISTIC

Consider the (normalized) discrete Fourier transform (DFT) $\boldsymbol{d}_x(f_n)$ of proper complex-valued $\boldsymbol{x}(t)$, $t = 1, 2, \cdots, N-1$, given by

$$\boldsymbol{d}_x(f_n) := \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \boldsymbol{x}(t) e^{-j2\pi f_n t} \tag{2}$$

where $f_n = n/N$, $n = 0, 1, \cdots, N-1$. Note that $\boldsymbol{d}_x(f_n)$ is periodic in n with period N, and is periodic in normalized frequency f_n with period 1. Hence $\boldsymbol{d}_x(f_n)$ for $n = 0, 1, \cdots, N-1$, completely determines $\boldsymbol{d}_x(f_n)$ for all integers n. Similarly define

$$\boldsymbol{d}_y(f_n) := \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \boldsymbol{y}(t) e^{-j2\pi f_n t}. \tag{3}$$

Here too $\boldsymbol{d}_y(f_n)$ for $n = 0, 1, \cdots, N-1$, completely determines $\boldsymbol{d}_y(f_n)$ for all integers n.

As proved in [16, p. 280, Sec. 6.2], for any statistical inference problem, the complete sample is a sufficient statistic, and so is any one-to-one function of a sufficient statistic. Since the inverse DFT yields (one-to-one transformation)

$$\boldsymbol{x}(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \boldsymbol{d}_x(f_n) e^{j2\pi f_n t}, \tag{4}$$

(and similarly for $\boldsymbol{y}(t)$), the set $\{\boldsymbol{d}_x(f_n),\ \boldsymbol{d}_y(f_n)\}_{n=0}^{N-1}$ is a sufficient statistic for our binary hypothesis testing problem.

D. MODEL ASSUMPTIONS

Here we state assumptions on $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$ used later in the paper.

(A1) *Assumption 2.6.1* [17]. Random sequences $\{\boldsymbol{x}(t)\}$ and $\{\boldsymbol{y}(t)\} \in \mathbb{C}^p$ are stationary with components $z_a(t)$, $a = 1, 2, \cdots, p, z \in \{\boldsymbol{x}, \boldsymbol{y}\}$, such that $E\{|z_a(t)|^k\} < \infty$, satisfying

$$\sum_{\tau_1,\tau_2,\cdots,\tau_{k-1}=-\infty}^{\infty} |c_{z:a_1,a_2,\cdots,a_k}(\tau_1, \tau_2, \cdots, \tau_{k-1})| < \infty \tag{5}$$

for $a_1, a_2, \cdots, a_k = 1, 2, \cdots, p$, and $k = 2, 3, \cdots$, where $c_{z:a_1,a_2,\cdots,a_k}(\tau_1, \tau_2, \cdots, \tau_{k-1})$ is the joint cumulant function of order k of stationary $\{z(t)\}, z \in \{\boldsymbol{x}, \boldsymbol{y}\}$.

(A2) The PSD matrices $\boldsymbol{S}_x(f) \succ 0$ and $\boldsymbol{S}_y(f) \succ 0$ for any $0 \le f \le 1$.

(A3) The PSD matrices $\boldsymbol{S}_x(f_n)$ and $\boldsymbol{S}_y(f_n)$ are locally smooth such that they are constant over $K\ (\ge p)$ consecutive frequency points, where $f_n = n/N$, $n \in [0, N-1]$.

See [15, Appendix A] (also [9, Appendix A]) for more details regarding satisfaction of Assumption 2.6.1 of [17] for a class of stationary sequences. If $\{\boldsymbol{x}(t)\}$ is Gaussian, then it is sufficient to verify (5) only for $k = 2$. Regarding assumption (A2), one can always add artificial proper white Gaussian noise to $\boldsymbol{x}(t)$ to achieve $\boldsymbol{S}_x(f) \succ 0$; similarly for $\{\boldsymbol{y}(t)\}$. Assumption (A3) is a standard assumption in PSD estimation literature [17].

E. ASYMPTOTIC DISTRIBUTION OF SUFFICIENT STATISTIC

Under (A1), exploiting [17, Th. 4.4.1], it has been shown in [9, Appendix A-C] that asymptotically, $\boldsymbol{d}_x(f_n)$, $n = 0, 1, \cdots, N-1$, are independent proper complex Gaussian $\mathcal{N}_c(\boldsymbol{0}, \boldsymbol{S}_x(f_n))$ random vectors, respectively, and similarly, $\boldsymbol{d}_y(f_n), n = 0, 1, \cdots, N-1$, are independent proper complex Gaussian $\mathcal{N}_c(\boldsymbol{0}, \boldsymbol{S}_y(f_n))$ random vectors, respectively. Since $\{\boldsymbol{x}(t)\}$ and $\{\boldsymbol{y}(t)\}$ are assumed to be independent, their DFTs $\boldsymbol{d}_x(f_n)$ and $\boldsymbol{d}_y(f_n)$ are also independent.

III. GENERALIZED LIKELIHOOD RATIO TEST

In this section we re-derive the GLRT of [9] in a principled way, using the sufficient statistic $\{\boldsymbol{d}_x(f_n),\ \boldsymbol{d}_y(f_n)\}_{n=0}^{N-1}$. In contrast, the starting point [9] is the estimated PSDs of $\{\boldsymbol{x}(t)\}$ and $\{\boldsymbol{y}(t)\}$.

Define

$$\boldsymbol{D}_x = [\boldsymbol{d}_x(f_0)\, \boldsymbol{d}_x(f_1)\ \cdots\ \boldsymbol{d}_x(f_{N-1})]^\top \in \mathbb{C}^{N\times p}, \tag{6}$$

$$\boldsymbol{D}_y = [\boldsymbol{d}_y(f_0)\, \boldsymbol{d}_y(f_1)\ \cdots\ \boldsymbol{d}_y(f_{N-1})]^\top \in \mathbb{C}^{N\times p}, \tag{7}$$

and

$$\boldsymbol{D} = [\boldsymbol{D}_x\ \ \boldsymbol{D}_y] \in \mathbb{C}^{N\times(2p)}. \tag{8}$$

The asymptotic joint probability density function (pdf) of $\boldsymbol{D}$ is given by

$$f_{\boldsymbol{D}}(\boldsymbol{D}) = \prod_{n=0}^{N-1} \Big[\frac{\exp\left(-\boldsymbol{d}_x^H(f_n)\boldsymbol{S}_x^{-1}(f_n)\boldsymbol{d}_x(f_n)\right)}{\pi^p\, |\boldsymbol{S}_x(f_n)|} \times \frac{\exp\left(-\boldsymbol{d}_y^H(f_n)\boldsymbol{S}_y^{-1}(f_n)\boldsymbol{d}_y(f_n)\right)}{\pi^p\, |\boldsymbol{S}_y(f_n)|} \Big] \tag{9}$$

where we do not distinguish between a random vector/matrix and the values taken by them in our notation (for simplicity). The PSD matrices $\boldsymbol{S}_x(f_n)$ and $\boldsymbol{S}_y(f_n)$ are unknown. Under $\mathcal{H}_0$, $\boldsymbol{S}_x(f_n) = \boldsymbol{S}_y(f_n)\ \forall n$. Testing for equality of the two PSDs is then reformulated as the problem

$$\begin{aligned} &\mathcal{H}_0 : \boldsymbol{S}_x(f_n) = \boldsymbol{S}_y(f_n) \quad \forall n \in [0, N-1] \\ &\mathcal{H}_1 : \boldsymbol{S}_x(f_n) \not\equiv \boldsymbol{S}_y(f_n) \quad \text{for } n \in [0, N-1]. \end{aligned} \tag{10}$$

Now assume that $\boldsymbol{S}_x(f_n)$ and $\boldsymbol{S}_y(f_n)$ are locally smooth (assumption (A3) of Sec. II-D), so that $\boldsymbol{S}_x(f_n)$ and $\boldsymbol{S}_y(f_n)$ are (approximately) constant over $K = 2m_t + 1 \ge p$ consecutive frequency points for some $m_t > 0$. Pick

$$\tilde{f}_k = \frac{(k-1)K + m_t}{N}, \quad k = 1, 2, \cdots, M, \tag{11}$$

$$M = \left\lfloor \frac{N - m_t}{K} \right\rfloor, \tag{12}$$

leading to M equally spaced frequencies $\tilde{f}_k$ in the interval $[0, 1]$, at intervals of K/N. It is assumed that for each $\tilde{f}_k$ (local smoothness), and $\ell = -m_t, -m_t+1, \cdots, m_t$,

$$\boldsymbol{S}_x(\tilde{f}_{k,\ell}) = \boldsymbol{S}_x(\tilde{f}_k) \text{ and } \boldsymbol{S}_y(\tilde{f}_{k,\ell}) = \boldsymbol{S}_y(\tilde{f}_k) \tag{13}$$

where

$$\tilde{f}_{k,\ell} = \frac{(k-1)K + m_t + \ell}{N}. \tag{14}$$

Using (13) in (9), we have

$$\begin{aligned} f_D(D) &= \prod_{k=1}^{M} \frac{\text{etr}\left(-\boldsymbol{S}_x^{-1}(\tilde{f}_k)\tilde{\boldsymbol{D}}_x(\tilde{f}_k) - \boldsymbol{S}_y^{-1}(\tilde{f}_k)\tilde{\boldsymbol{D}}_y(\tilde{f}_k)\right)}{\pi^{2Kp}\, |\boldsymbol{S}_x(\tilde{f}_k)|^K\, |\boldsymbol{S}_y(\tilde{f}_k)|^K} \\ &= \prod_{k=1}^{M} f_{\check{\boldsymbol{D}}(\tilde{f}_k)}(\check{\boldsymbol{D}}(\tilde{f}_k)) \end{aligned} \tag{15}$$

where $K \times (2p)\ \check{\boldsymbol{D}}(\tilde{f}_k)$ is

$$\check{\boldsymbol{D}}(\tilde{f}_k) = \left[\check{\boldsymbol{D}}_x(\tilde{f}_k)\ \ \check{\boldsymbol{D}}_y(\tilde{f}_k)\right], \tag{16}$$

$K \times p\ \check{\boldsymbol{D}}_z(\tilde{f}_k), z \in \{\boldsymbol{x}, \boldsymbol{y}\}$, is

$$\check{\boldsymbol{D}}_z(\tilde{f}_k) = \left[\boldsymbol{d}_z(\tilde{f}_{k,-m_t})\, \boldsymbol{d}_z(\tilde{f}_{k,-m_t+1})\ \cdots\ \boldsymbol{d}_z(\tilde{f}_{k,m_t})\right]^\top, \tag{17}$$

and $p \times p\ \tilde{\boldsymbol{D}}_z(\tilde{f}_k), z \in \{\boldsymbol{x}, \boldsymbol{y}\}$, is

$$\tilde{\boldsymbol{D}}_z(\tilde{f}_k) = \sum_{\ell=-m_t}^{m_t} \boldsymbol{d}_z(\tilde{f}_{k,\ell})\boldsymbol{d}_z^H(\tilde{f}_{k,\ell}). \tag{18}$$

Note that $\left(\tilde{\boldsymbol{D}}_z(\tilde{f}_k)\right)^{-1}$ exists w.p.1 for $K \ge p$.

Explicitly indicating the dependence on the underlying hypothesis $\mathcal{H}_i$, $i = 0, 1$, the pdf of $\boldsymbol{D}$ under $\mathcal{H}_i$ is expressed as

$$f_{\boldsymbol{D}|\mathcal{H}_i}(\boldsymbol{D}|\mathcal{H}_i) = \prod_{k=1}^{M} f_{\check{\boldsymbol{D}}(\tilde{f}_k)|\mathcal{H}_i}(\check{\boldsymbol{D}}(\tilde{f}_k)|\mathcal{H}_i) \tag{19}$$

where

$$f_{\check{\boldsymbol{D}}(\tilde{f}_k)|\mathcal{H}_i}(\check{\boldsymbol{D}}(\tilde{f}_k)|\mathcal{H}_i) = \frac{\text{etr}\left(-\boldsymbol{S}_x^{-1}(\tilde{f}_k)\tilde{\boldsymbol{D}}_x(\tilde{f}_k\right)}{\pi^{Kp}\,|\boldsymbol{S}_x(\tilde{f}_k)|^K} \times \frac{\text{etr}\left(-\boldsymbol{S}_y^{-1}(\tilde{f}_k)\tilde{\boldsymbol{D}}_y(\tilde{f}_k)\right)}{\pi^{Kp}\,|\boldsymbol{S}_y(\tilde{f}_k)|^K}. \tag{20}$$

Under $\mathcal{H}_1$ where there is no specific structure to the PSD matrices, the unknowns in (19)-(20) are Hermitian positive definite matrices $\boldsymbol{S}_z(\tilde{f}_k)$, $z \in \{\boldsymbol{x}, \boldsymbol{y}\}$, $k = 1, 2, \cdots, M$. By [18, Th. 1.10.4], for any positive definite $m \times m$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and for any scalars $a > 0$ and $b > 0$, one has

$$|\boldsymbol{A}|^{-b}\text{etr}\left(-a\boldsymbol{A}^{-1}\boldsymbol{B}\right) \le |a\boldsymbol{B}/b|^{-b}\exp(-mb), \tag{21}$$

with equality if and only if $\boldsymbol{A} = a\boldsymbol{B}/b$. Applying (21) to (20) with $a = 1$, $b = K$, $\boldsymbol{A} = \boldsymbol{S}_z(\tilde{f}_k)$, $\boldsymbol{B} = \tilde{\boldsymbol{D}}_z(\tilde{f}_k)$, $z \in \{\boldsymbol{x}, \boldsymbol{y}\}$, and $m = p$, it follows that the maximum likelihood (ML) estimate $\hat{\boldsymbol{S}}_z(\tilde{f}_k)$ of $\boldsymbol{S}_z(\tilde{f}_k)$, under $\mathcal{H}_1$, is given by

$$\hat{\boldsymbol{S}}_z(\tilde{f}_k) = \frac{1}{K}\tilde{\boldsymbol{D}}_z(\tilde{f}_k) = \frac{1}{K}\sum_{l=-m_t}^{m_t} \boldsymbol{d}_z(\tilde{f}_{k,\ell})d_z^H(\tilde{f}_{k,\ell}). \tag{22}$$

Setting $\boldsymbol{S}_x(\tilde{f}_k) = \hat{\boldsymbol{S}}_x(\tilde{f}_k)$ and $\boldsymbol{S}_y(\tilde{f}_k) = \hat{\boldsymbol{S}}_y(\tilde{f}_k)$ in (19) and (20) yields

$$\sup_{\boldsymbol{S}_x(\tilde{f}_k),\boldsymbol{S}_y(\tilde{f}_k)} f_{\boldsymbol{D}|\mathcal{H}_1}(\boldsymbol{D}|\mathcal{H}_1) = \frac{e^{-2MKp}}{\pi^{2MKp}}\prod_{k=1}^{M}|\hat{\boldsymbol{S}}_x(\tilde{f}_k)|^{-K}\,|\hat{\boldsymbol{S}}_y(\tilde{f}_k)|^{-K}. \tag{23}$$

Under $\mathcal{H}_0$, we have $\boldsymbol{S}_x(\tilde{f}_k) = \boldsymbol{S}_y(\tilde{f}_k)$. Setting $\boldsymbol{S}_0(\tilde{f}_k) = \boldsymbol{S}_x(\tilde{f}_k) = \boldsymbol{S}_y(\tilde{f}_k)$, the pdf of $\check{\boldsymbol{D}}(\tilde{f}_k)$ under $\mathcal{H}_0$ can be simplified as

$$f_{\check{\boldsymbol{D}}(\tilde{f}_k)|\mathcal{H}_0}(\check{\boldsymbol{D}}(\tilde{f}_k)|\mathcal{H}_0) = \frac{\text{etr}\left(-\boldsymbol{S}_0^{-1}(\tilde{f}_k)\left(\tilde{\boldsymbol{D}}_x(\tilde{f}_k) + \tilde{\boldsymbol{D}}_y(\tilde{f}_k)\right)\right)}{\pi^{2Kp}\,|\boldsymbol{S}_0(\tilde{f}_k)|^{2K}}. \tag{24}$$

Applying (21) to (24) with $a = 1$, $b = 2K$, $\boldsymbol{A} = \boldsymbol{S}_0(\tilde{f}_k)$, $\boldsymbol{B} = \tilde{\boldsymbol{D}}_x(\tilde{f}_k) + \tilde{\boldsymbol{D}}_y(\tilde{f}_k)$ and $m = 2$, it follows that the ML estimate $\hat{\boldsymbol{S}}_0(\tilde{f}_k)$ of $\boldsymbol{S}_0(\tilde{f}_k)$ is given by

$$\hat{\boldsymbol{S}}_0(\tilde{f}_k) = \frac{1}{2K}\left(\tilde{\boldsymbol{D}}_x(\tilde{f}_k) + \tilde{\boldsymbol{D}}_y(\tilde{f}_k)\right) = \left(\hat{\boldsymbol{S}}_x(\tilde{f}_k) + \hat{\boldsymbol{S}}_y(\tilde{f}_k)\right)/2\,. \tag{25}$$

Setting $\boldsymbol{S}_0(\tilde{f}_k) = \hat{\boldsymbol{S}}_0(\tilde{f}_k)$ in (19) and (24) leads to

$$\sup_{\boldsymbol{S}_x(\tilde{f}_k)=\boldsymbol{S}_y(\tilde{f}_k)} f_{\boldsymbol{D}|\mathcal{H}_0}(\boldsymbol{D}|\mathcal{H}_0) = \frac{e^{-2MKp}}{\pi^{2MKp}}\prod_{k=1}^{M}|\frac{1}{2}\left(\hat{\boldsymbol{S}}_x(\tilde{f}_k) + \hat{\boldsymbol{S}}_y(\tilde{f}_k)\right)|^{-2K}. \tag{26}$$

Using (23) and (26) we obtain the GLRT

$$\mathcal{L} := \frac{\sup_{\boldsymbol{S}_x(\tilde{f}_k),\boldsymbol{S}_y(\tilde{f}_k)} f_{\boldsymbol{D}|\mathcal{H}_1}(\boldsymbol{D}|\mathcal{H}_1)}{\sup_{\boldsymbol{S}_x(\tilde{f}_k)=\boldsymbol{S}_y(\tilde{f}_k)} f_{\boldsymbol{D}|\mathcal{H}_0}(\boldsymbol{D}|\mathcal{H}_0)} \tag{27}$$

$$= \prod_{k=1}^{M}\frac{|\frac{1}{2}\left(\hat{\boldsymbol{S}}_x(\tilde{f}_k) + \hat{\boldsymbol{S}}_y(\tilde{f}_k)\right)|^{2K}}{|\hat{\boldsymbol{S}}_x(\tilde{f}_k)|^K\,|\hat{\boldsymbol{S}}_y(\tilde{f}_k)|^K} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau_1 \tag{28}$$

where the threshold τ_1 is picked to achieve a pre-specified probability of false alarm $P_{fa} = P\{\mathcal{L} \ge \tau_1\,|\,\mathcal{H}_0\}$. This requires pdf of $\mathcal{L}$ under $\mathcal{H}_0$. A solution to this problem is offered in [9, Th. 1], and it works well. We recall [9, Th. 1] in the Appendix. Further remarks related to [9] are in Remark 1.

Note that $\hat{\boldsymbol{S}}_z(\tilde{f}_k)$ in (22) is an estimator of the PSD of $\boldsymbol{z}(t)$, $z \in \{\boldsymbol{x}, \boldsymbol{y}\}$, at frequency $\tilde{f}_k$, based on unweighted smoothing in frequency-domain, as given in [17, eq. (7.3.2)]. Based on [17, Th. 7.3.3], it is shown in [9, Appendix A-C] that as $N \to \infty$, $\hat{\boldsymbol{S}}_z(\tilde{f}_k)$ is distributed as $W_C\left(p, K, K^{-1}\boldsymbol{S}_z(\tilde{f}_k)\right)$ (denoted as $\hat{\boldsymbol{S}}_z(\tilde{f}_k) \overset{a}{\sim} W_C(p, K, K^{-1}\boldsymbol{S}_z(\tilde{f}_k))$) where $W_C\left(p, K, K^{-1}\boldsymbol{S}_z(\tilde{f}_k)\right)$ denotes the complex Wishart distribution of dimension p, degrees of freedom K, and mean value $\boldsymbol{S}_z(\tilde{f}_k)$. If a random matrix $\boldsymbol{X} \sim W_C(m, K, \boldsymbol{S}(f))$, then by [17, Sec. 4.2], $\mathbb{E}\{X\} = KS(f)$, $\text{cov}\left\{\boldsymbol{X}_{jk}, \boldsymbol{X}_{rs}\right\} = K\boldsymbol{S}_{jr}(f)\boldsymbol{S}_{ks}^*(f)$, and for $K \ge m$, the pdf of X is given by

$$f_{\boldsymbol{X}}(\boldsymbol{X}) = \frac{1}{\Gamma_m(K)}\frac{1}{|\boldsymbol{S}(f)|^K}|\boldsymbol{X}|^{K-m}\,\text{etr}\{-\boldsymbol{S}^{-1}(f)\boldsymbol{X}\} \tag{29}$$

where the pdf (29) is defined for $\boldsymbol{S}(f) \succ 0$ and $\boldsymbol{X} \succeq 0$, and is otherwise zero, and

$$\Gamma_m(K) := \pi^{m(m-1)/2}\prod_{j=1}^{m}\Gamma(K - j + 1) \tag{30}$$

where $\Gamma(n)$ denotes the (complete) Gamma function $\Gamma(z) := \int_0^\infty t^{z-1}e^{-t}\,dt$.

Remark 1: The GLRT (28) was first derived in [9]. Our derivation of the GLRT in this paper is significantly different from that in [9], where the starting point is the estimated PSDs of signals $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$. In this paper, we first developed a frequency-domain sufficient statistic (Sec. II-C), and then under certain sufficient conditions ((A1)-(A3) in Sec. II-D), we derived the GLRT based on the frequency-domain sufficient statistic, in a principled way. The estimates of the two PSDs follow from the frequency-domain sufficient statistic as ML estimates, as in (22). Since the GLRT statistic is the same in this paper and in [9], all other details, such as analytical threshold calculation via [9, Th. 1] and corroboration of the theory via simulations, given in [9] apply here as well.

IV. ROBUSTIFICATION

The GLRT of Sec. III is based on clean data (i.e., no outliers). We now turn to robustifying this GLRT against a class of outliers.

In Sec. IV-A we first introduce the widely used additive outlier model for measurements. This model is not essential to robust processing, but is one of the useful models frequently used in the literature [12], [13], and is used in our simulations. The method of [14] for robust estimation of multivariate scatter is reviewed in Sec. IV-B. It is used in Sec. IV-C for detection of outliers in the two sets of measurements. In Sec. IV-D the detected outliers are "cleaned" via median filtering. So cleaned datasets are then used for random signal comparison using the method of [9], as summarized in Sec. IV-E.

A. ADDITIVE OUTLIERS

Now $\{\boldsymbol{x}(t)\}$ and $\{\boldsymbol{y}(t)\}$ are assumed to be corrupted by additive outliers [12, p. 253], to yield corrupted signals $\{\tilde{\boldsymbol{x}}(t)\}$ and $\{\tilde{\boldsymbol{y}}(t)\}$:

$$\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t) + \boldsymbol{v}_x(t), \quad \tilde{\boldsymbol{y}}(t) = \boldsymbol{y}(t) + \boldsymbol{v}_y(t), \tag{31}$$

where, with a "small" probability p_o,

$$\boldsymbol{v}_x(t) \sim \mathcal{N}_c(0, \sigma_{vx}^2 \boldsymbol{I}_p), \quad \boldsymbol{v}_y(t) \sim \mathcal{N}_c(0, \sigma_{vy}^2 \boldsymbol{I}_p), \tag{32}$$

$$\sigma_{vx}^2 \gg \mathbb{E}\{|\boldsymbol{x}(t)|^2\}, \quad \sigma_{vy}^2 \gg \mathbb{E}\{|\boldsymbol{y}(t)|^2\}, \tag{33}$$

the outliers are i.i.d. mutually independent, and independent of the clean signals, and with probability $1 - p_o$,

$$\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t), \quad \tilde{\boldsymbol{y}}(t) = \boldsymbol{y}(t). \tag{34}$$

Recall that $\mathcal{N}_c(\boldsymbol{m}, \boldsymbol{\Sigma})$ denotes proper, complex Gaussian distribution with mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$. The model (31) is not essential to robust processing, but is one of the useful models frequently used in the literature [12], [13], and is used in our simulations.

B. ROBUST ESTIMATION OF MULTIVARIATE SCATTER

There is a large body of work on robust statistics and robust signal processing [12], [13]. We will use the computationally efficient algorithm DetS of [14]. The approach of [14] is described below.

Given a sample $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_n \in \mathbb{R}^q$ of real-valued random vectors, a multivariate S-estimator of location (mean) $\boldsymbol{m}$ and scatter (covariance matrix) $\boldsymbol{S}$ is defined as the couple $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ which minimizes $|\boldsymbol{S}|$ under the constraint

$$\frac{1}{n}\sum_{t=1}^{n} \rho\left(\sqrt{(\boldsymbol{z}_t - \boldsymbol{m})^T \boldsymbol{S}^{-1}(\boldsymbol{z}_t - \boldsymbol{m})}\right) = b \tag{35}$$

over all $(\boldsymbol{m}, \boldsymbol{S})$, where $\boldsymbol{m} \in \mathbb{R}^q$, $\boldsymbol{S}$ is a $q \times q$ symmetric positive definite matrix, ρ is a smooth bounded "$\rho-$function," typically picked to be Tukey's bisquare $\rho-$function, and the constant $b \in (0, 1)$ influences the "breakdown" value of the estimator under the nominal (Gaussian) model [12], [14]. Thus, an S-estimator of multivariate location and scale minimizes the determinant of the covariance matrix, subject to a constraint on the magnitudes of the corresponding Mahalanobis distances. Tukey's bisquare function is given by

$$\rho(x) = \begin{cases} \dfrac{x^2}{2} - \dfrac{x^4}{2c^2} + \dfrac{x^6}{4c^4} & \text{for } |x| \le c \\ \dfrac{c^2}{6} & \text{for } |x| > c \end{cases} \tag{36}$$

where c is a tuning constant that, together with b in (35), influences the breakdown point (BP). The BP of an estimator characterizes its quantitative robustness. It is the maximal fraction of outliers in the observations which an estimate can handle without breaking down [13]. The BP takes values between 0 and 50%, with higher value indicating larger quantitative robustness. The BP of a multivariate S-estimator is $b/\rho(c)$ [14]. If the nominal (i.e., outlier free) model for the data generates Gaussian vectors, then [19]

$$b = \frac{q}{2}\bar{\chi}_{q+2}^2(c^2) - \frac{q(q+2)}{2c^2}\bar{\chi}_{q+4}^2(c^2) + \frac{q(q+2)(q+4)}{6c^4}\bar{\chi}_{q+6}^2(c^2) + \frac{c^2}{6}(1 - \bar{\chi}_q^2(c^2)) \tag{37}$$

where $\bar{\chi}_\nu^2$ denotes the cumulative distribution function (CDF) of a χ_ν^2 random variable. For a given BP between 0 and 50%, one can derive the value of the corresponding tuning parameter c in (36) [14].

Tukey's bisquare $\rho-$function is a non-convex function, hence, to obtain a global minimum that satisfies (35), random subsampling methods are used to obtain a good initial guess [14]. The end result is a robust scatter estimate $\hat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ that is "close" to the true value under no outliers, but is "robust" to outliers. Software in MATLAB for computationally efficient algorithm DetS of [14] is available from the author's website (`http://wis.kuleuven.be/stat/robust/LIBRA`), and was used with the default BP $b/\rho(c) = 0.5$.

C. OUTLIER DETECTION

In our case, we have complex-valued data. Therefore, to invoke [14], we represent $\tilde{\boldsymbol{x}}(t)$ as two real-valued random vectors $\mathrm{Re}\{\tilde{\boldsymbol{x}}(t)\}$ and $\mathrm{Im}\{\tilde{\boldsymbol{x}}(t)\}$; similarly for $\tilde{\boldsymbol{y}}(t)$. We have

$$\boldsymbol{z}_x(t) = \begin{bmatrix} \mathrm{Re}\{\tilde{\boldsymbol{x}}(t)\} \\ \mathrm{Im}\{\tilde{\boldsymbol{x}}(t)\} \end{bmatrix} \in \mathbb{R}^{2p}, \tag{38}$$

$$\boldsymbol{z}_y(t) = \begin{bmatrix} \mathrm{Re}\{\tilde{\boldsymbol{y}}(t)\} \\ \mathrm{Im}\{\tilde{\boldsymbol{y}}(t)\} \end{bmatrix} \in \mathbb{R}^{2p} \tag{39}$$

and $q = 2p$. Now apply DetS algorithm of [14] to $\boldsymbol{z}_x(t)$ and $\boldsymbol{z}_y(t)$ (separately) to obtain $(2p) \times (2p)$ robust scatter S-estimates $\hat{\boldsymbol{\Sigma}}_{zx}$ and $\hat{\boldsymbol{\Sigma}}_{zy}$, respectively.

Having obtained outlier-resistant estimates of the covariances matrices of $\{\tilde{\boldsymbol{x}}(t)\}$ and $\{\tilde{\boldsymbol{y}}(t)\}$, we now consider detection of outliers. Note that, by assumption, clean data are zero-mean. Hence, our nominal models for the data are $\boldsymbol{z}_x(t) \sim \mathcal{N}_r(\boldsymbol{0}, \hat{\boldsymbol{\Sigma}}_{zx})$ and $\boldsymbol{z}_y(t) \sim \mathcal{N}_r(\boldsymbol{0}, \hat{\boldsymbol{\Sigma}}_{zy})$. Define

$$w_{zx}(t) = \boldsymbol{z}_x^\top(t)\hat{\boldsymbol{\Sigma}}_{zx}^{-1}\boldsymbol{z}_x(t), \quad w_{zy}(t) = \boldsymbol{z}_y^\top(t)\hat{\boldsymbol{\Sigma}}_{zy}^{-1}\boldsymbol{z}_y(t). \tag{40}$$

Under the nominal Gaussian models, both $w_{zx}(t) \sim \chi^2_{2p}$ and $w_{zy}(t) \sim \chi^2_{2p}$. For a significance level (probability of false alarm) α, we set the threshold τ at $\tau = (\chi^2_{2p})^{-1}(1-\alpha)$, i.e., $1-\alpha = \bar{\chi}^2_{2p}(\tau)$, and declare $\tilde{\boldsymbol{x}}(t)$ to be an outlier if the corresponding $w_{zx}(t) > \tau$. For our simulations, we picked $\alpha = 0.025$. The same procedure is applied to $\tilde{\boldsymbol{y}}(t)$. This method is robust since the estimates of the covariance matrices are robust.

D. DATA CLEANING

Once an outlier is detected, it has to be "replaced." We first replace it with the median over a 5-point window, centered at the location of the detected outlier unless one is at the edge of the data block, in which case the window is "unbalanced." In the unbalanced case, we still consider a 5-point window which includes the detected outlier and four additional contiguous data points. This is done for $\tilde{\boldsymbol{x}}(t)$, at $t = 0, 1, \cdots, N-1$, one sample at a time. In the final step, the selectively median filtered data is again checked for outliers to guard against "patchy" outliers. Any outlier detected at this stage is clipped (i.e., scaled with a positive scalar) to yield $\boldsymbol{z}_x^\top(t)\hat{\boldsymbol{\Sigma}}_{zx}^{-1}\boldsymbol{z}_x(t) = \tau$. The sequence obtained at the end of this stage is our cleaned $\tilde{\boldsymbol{x}}(t)$, labeled $\check{\boldsymbol{x}}(t)$. We obtain $\check{\boldsymbol{y}}(t)$ from $\tilde{\boldsymbol{y}}(t)$ similarly. While the median filtered sequences in the first step are well-motivated, the clipping step (if needed) is heuristic.

E. ROBUST COMPARISON

Given two contaminated realizations $\{\tilde{\boldsymbol{x}}(t)\}$ and $\{\tilde{\boldsymbol{y}}(t)\}$ for $t = 0, 1, \cdots, N-1$, our proposed robust comparison method is as follows.

(i) Obtain robust scatter S-estimates $\hat{\boldsymbol{\Sigma}}_{zx}$ and $\hat{\boldsymbol{\Sigma}}_{zy}$ using the DetS algorithms of [14], as discussed in Secs. IV-B and IV-C.
(ii) Detect outliers in $\{\tilde{\boldsymbol{x}}(t)\}$ and $\{\tilde{\boldsymbol{y}}(t)\}$ using $\hat{\boldsymbol{\Sigma}}_{zx}$ and $\hat{\boldsymbol{\Sigma}}_{zy}$, as discussed in Sec. IV-C.
(iii) Clean $\{\tilde{\boldsymbol{x}}(t)\}$ and $\{\tilde{\boldsymbol{y}}(t)\}$ of the detected outliers by a combination of selective median filtering and clipping, as discussed in Sec. IV-D. Denote the respective cleaned sequences as $\check{\boldsymbol{x}}(t)$ and $\check{\boldsymbol{y}}(t)$.
(iv) Now compare $\check{\boldsymbol{x}}(t)$ and $\check{\boldsymbol{y}}(t)$ via the PSD-based approach discussed in Sec. III.

Remark 2: The robustification approach outlined in this section, based on robust estimation of multivariate scatter, is novel, even though the various components (Secs. IV-B or IV-C) have been used before in various contexts. Data cleaning via dynamic modeling of the underlying signals (time series) has a long history; see, e.g., [12, Ch. 8], [20], and references therein. Robust dynamic modeling (autoregressive, autoregressive moving average, or state-space models) of vector time series is significantly more complicated and computationally demanding than the relatively simple robust estimation of multivariate scatter. Our attempt to use [20] to robustly fit vector autoregressive models and to clean contaminated data based on the fitted model (as noted on [20, p. 72]) has been unsuccessful for the simulation examples presented later in Sec. V.

V. SIMULATION EXAMPLES

We now present some computer simulation examples to illustrate application of the proposed approach to wireless user authentication when the data may be corrupted with outliers.

We generate stationary $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t) \in \mathbb{C}^p$, $p = 2$, as $\boldsymbol{x}(t) = \boldsymbol{s}_0(t) + \boldsymbol{n}(t)$, $\boldsymbol{y}(t) = \boldsymbol{s}_0(t) + \boldsymbol{n}(t)$ under $\mathcal{H}_0$, and $\boldsymbol{x}(t) = \boldsymbol{s}_0(t) + \boldsymbol{n}(t)$, $\boldsymbol{y}(t) = \boldsymbol{s}_1(t) + \boldsymbol{n}(t)$ under $\mathcal{H}_1$, where $\{\boldsymbol{n}(t)\}$ is spatially uncorrelated, colored, proper complex Gaussian noise, and $\{\boldsymbol{s}_i(t)\}$, $i = 0, 1$, are the signal sequences. Noise sequences $\{\boldsymbol{n}(t)\}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ are independent of each other, but identically distributed. The noise sequence $\{\boldsymbol{n}(t)\}$ is generated as

$$\boldsymbol{n}(t) = \boldsymbol{n}_c(t) + \boldsymbol{n}_w(t), \tag{41}$$

where $\boldsymbol{n}_w(t) \sim \mathcal{N}_c(\boldsymbol{0}, \sigma_w^2\boldsymbol{I})$ is i.i.d., and $\boldsymbol{n}_c(t)$ is generated as follows. The various components of $\boldsymbol{n}_c(t)$ are i.i.d., and each component is generated by filtering an i.i.d. scalar sequence, distributed as $\mathcal{N}_c(0, \sigma_c^2)$, through a linear filter with impulse response $\{0.4575, 0.7625, 0.4575\}$, where $\sqrt{(0.4575)^2 + (0.7625)^2 + (0.4575)^2} = 1$. Therefore, $\mathbb{E}\{\|\boldsymbol{n}_c(t)\|^2\} = p\sigma_c^2$, leading to $\mathbb{E}\{\|\boldsymbol{n}(t)\|^2\} = p(\sigma_c^2 + \sigma_w^2) = p\sigma_n^2$. We pick $\sigma_w^2 = 0.2\sigma_n^2$ for a given value of σ_n^2. The signal $\{\boldsymbol{s}(t)\}$ is a filtered digital communications signal generated by passing an information sequence through a frequency-selective Rayleigh fading channel as follows:

$$\boldsymbol{s}_i(t) = \sum_{l=0}^{4} \boldsymbol{h}_i(l)d(t-l), \quad i = 0, 1, \tag{42}$$

where $d(t)$ is a scalar i.i.d. QPSK sequence, filtered through a random time-invariant, frequency-selective Rayleigh fading $p \times 1$ vector channel $\boldsymbol{h}_i(l)$ with 5 taps, equal power delay profile, mutually independent components, which are identically distributed zero-mean proper complex Gaussian random variables. For different ls, $\boldsymbol{h}_i(l)$s are mutually independent and identically distributed as $\boldsymbol{h}_i(l) \sim \mathcal{N}_c(\boldsymbol{0}, \sigma_h^2\boldsymbol{I})$; they are also independent for $i = 0$ and 1. The signal $\boldsymbol{s}_i(t)$ was scaled to achieve a given SNR $\mathbb{E}\{\|\boldsymbol{s}_i(t)\|^2\}/\mathbb{E}\{\|\boldsymbol{n}(t)\|^2\}$.

Additive outliers were added as in (31) with $\sigma_{vx}^2 = 100\mathbb{E}\{|\boldsymbol{x}(t)|^2\}/p$ and $\sigma_{vy}^2 = 100\mathbb{E}\{|\boldsymbol{y}(t)|^2\}/p$.

We picked N=256 with K=11 ($m_t = 5$), and the probability of outlier addition $p_o = 0$ or 0.1 (10% contamination). The approach of [9] (also Sec. III) was applied after detecting outliers and cleaning the data, where the threshold (τ in (43)) was picked for a false alarm rate of 0.005 (0.5%). Figs. 1-3 show the "authentication" results based on 1000 runs, under various scenarios. In user authentication in wireless networks with multi-antenna receivers, we suppose that $\boldsymbol{x}(t)\}$ originates from an authenticated user, and the problem is to ascertain if $\boldsymbol{y}(t)\}$ also originates from the same user with identical channel, hence, identical PSD, or if it is from a "spoofer" at a different location, hence with a different channel, therefore, different PSD.

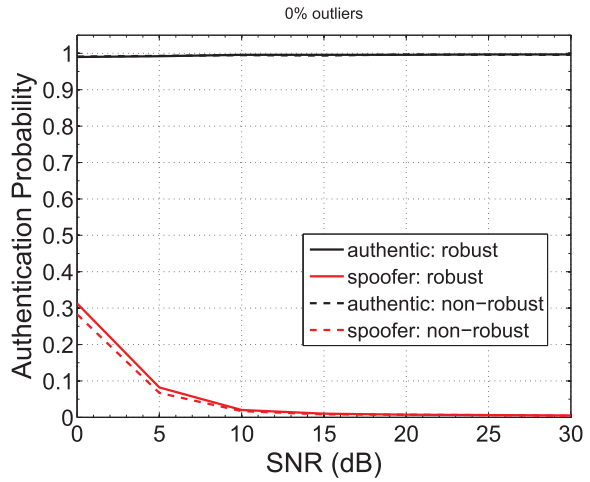

FIGURE 1. Authentication probability results based on 1000 runs, $p = 2$, design $P_{fa} = 0.005$, $N = 256$, $K = 11$, no outliers ($p_o = 0$). (The curves for "authentic: robust" and "authentic: non-robust" are very close with no discernible difference.) Authentication probability refers to fraction of runs (out of 1000) in which the user was declared to be authentic. The curve labeled "authentic" refers to the case where the second message $\tilde{y}(t)$ originated from the author of the first message $\tilde{x}(t)$. The curve labeled "spoofer" refers to the case where the second message originated from a spoofer. For curves labeled "non-robust," the method of [9] was applied without looking for any outliers, and curves labeled "robust" refer to the case where the method of [9] was applied after detecting outliers (if any) and cleaning the data.

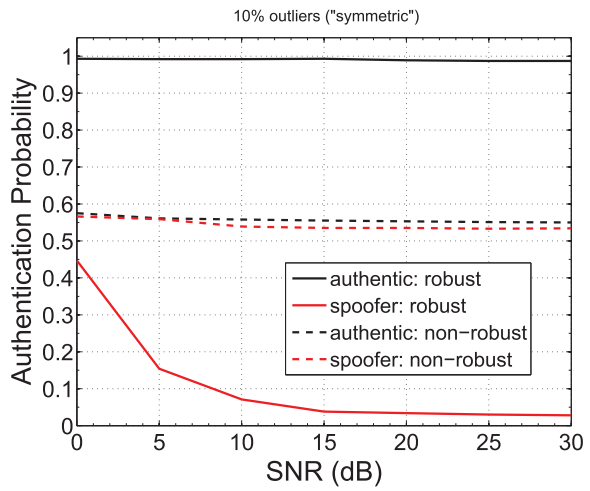

FIGURE 2. Authentication probability results based on 1000 runs, $p = 2$, design $P_{fa} = 0.005$, $N = 256$, $K = 11$, 10% outliers in both $x(t)$ and $y(t)$ ($p_o = 0.1$).

Figs. 1-3 show our authentic user detection probability results where authentic user detection probability refers to fraction of runs (out of 1000) in which the authentic user was selected by the proposed test, i.e., $\{\tilde{\boldsymbol{x}}(t)\}$ and $\{\tilde{\boldsymbol{y}}(t)\}$ both have the same signal $\boldsymbol{s}_0(t)$. Two cases are depicted in Figs. 1-3: the curves labeled "authentic" is obtained when $\{\tilde{\boldsymbol{y}}(t)\}$ originates from the authentic user; the curves labeled "spoofer" are obtained when $\{\tilde{\boldsymbol{y}}(t)\}$ originates from a spoofer. If $\mathcal{H}_0$ is accepted, then one declares that the second message ($\{\tilde{\boldsymbol{y}}(t)\}$) originates from the author of the first message ($\{\tilde{\boldsymbol{x}}(t)\}$). If $\mathcal{H}_0$ is rejected, then one declares that the second message

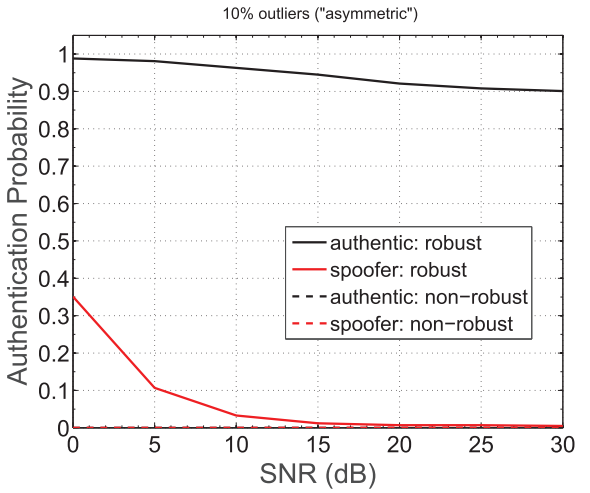

FIGURE 3. Authentication probability results based on 1000 runs, $p = 2$, design $P_{fa} = 0.005$, $N = 256$, $K = 11$, 10% outliers in $x(t)$ only and no outliers in $y(t)$ ($p_o = 0.1$ for $x(t)$ and $p_o = 0$ for $y(t)$). The two curves for non-robust case are very close with no discernible difference.

originates from other than the author of the first message (we have a potential spoofer). Two approaches were considered: We applied the method of [9] without looking for any outliers, this is the "non-robust" case, and we applied the method of [9] after detecting outliers (if any) and cleaning the data (as in Sec. IV), this is the "robust" case.

It is seen from Fig. 1 that when there are no outliers, our proposed method yields results very close to the "optimal" results of [9]. When 10% outliers are added to the two noisy signals (Fig. 2), the non-robust method of [9] is unable to distinguish between the authentic and spoofer signals as they ware buried in outliers, whereas the proposed robust approach yields results that are not significantly different from the results for the outlier-free case shown in Fig. 1. Finally, we show the "asymmetric" case in Fig. 3 where outliers are present only in $\tilde{\boldsymbol{x}}(t)$. Here too the proposed approach works well whereas the non-robust method always finds the two signals to be distinct, resulting in close-to-zero authentication probability.

VI. CONCLUSIONS

We considered the problem of comparing two complex multivariate random signal realizations, possibly contaminated with additive outliers, to ascertain whether they have identical power spectral densities. We exploited an existing robust estimator of multivariate scatter to detect the outliers, and subsequently to clean the data. An existing GLRT of [9] was then applied to the cleaned signal realizations. We also derived the GLRT of [9] in a principled way. The approach was illustrated via simulations in the context of user authentication in wireless networks with multi-antenna receivers.

APPENDIX

Here we recall the result from [9] needed to compute the threshold for the hypothesis testing problem (10).

Theorem 1: [9, Th. 1]. The GLRT for the binary hypothesis testing problem (10) is given by

$$2\rho\ \ln(\mathcal{L}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{43}$$

where

$$\rho = 1 - \frac{2p^2-1}{4pK}, \tag{44}$$

$$\ln(\mathcal{L}) = K\Bigg(-2pM\ln(2) + \sum_{k=1}^{M}\Big\{2\ln(|\hat{\boldsymbol{S}}_x(\tilde{f}_k) + \hat{\boldsymbol{S}}_y(\tilde{f}_k)| - \ln(|\hat{\boldsymbol{S}}_x(\tilde{f}_k)| - \ln(|\hat{\boldsymbol{S}}_y(\tilde{f}_k)|\Big\}\Bigg). \tag{45}$$

The threshold τ is picked to achieve a pre-specified probability of false alarm $P_{fa} = P\{2\rho\ \ln(\mathcal{L}) > \tau\,|\,\mathcal{H}_0\} = 1 - P\{2\rho\ \ln(\mathcal{L}) \le \tau\,|\,\mathcal{H}_0\}$. The probability $P\{2\rho\ \ln(\mathcal{L}) \le \tau\,|\,\mathcal{H}_0\}$ is given by

$$\begin{aligned} &P\{2\rho\ \ln(\mathcal{L}) \le z\,|\,\mathcal{H}_0\} \\ &= P\{\chi^2_\nu \le z\} + \omega_2\Big[P\{\chi^2_{\nu+4} \le z\} - P\{\chi^2_\nu \le z\}\Big] + \omega_3\Big[P\{\chi^2_{\nu+6} \le z\} - P\{\chi^2_\nu \le z\}\Big] \\ &\quad + \Big\{\omega_4\Big[P\{\chi^2_{\nu+8} \le z\} - P\{\chi^2_\nu \le z\}\Big] + \frac{1}{2}\omega_2^2\Big[P\{\chi^2_{\nu+8} \le z\} - 2P\{\chi^2_{\nu+4} \le z\} + P\{\chi^2_\nu \le z\}\Big]\Big\} + \mathcal{O}(K^{-5}) \end{aligned} \tag{46}$$

where

$$\omega_r = \frac{(-1)^{r+1}M}{r(r+1)(\rho K)^r}\sum_{l=1}^{p}\Big\{2B_{r+1}((1-\rho)K+1-l) - \frac{1}{2^r}B_{r+1}(2(1-\rho)K+1-l)\Big\}, \tag{47}$$

$B_r(h)$ denotes the Bernoulli polynomial of degree r and order unity,

$$\nu = p^2M, \tag{48}$$

and χ^2_n denoting a random variable with central chi-square distribution with n degrees of freedom (as well as the distribution itself). •

For details, please see [9].

REFERENCES

[1] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Using the physical layer for wireless authentication in time-variant channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2571–2579, Jul. 2008.
[2] J. K. Tugnait, "Wireless user authentication via comparison of power spectral densities," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1791–1802, Sep. 2013.
[3] A. Tani and R. Fantacci, "A low-complexity cyclostationary-based spectrum sensing for UWB and WiMAX coexistence with noise uncertainty," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2940–2950, Jul. 2010.
[4] J. K. Tugnait, "Multiple antenna spectrum sensing in colored noise," in *Handbook of Cognitive Radio*, W. Zhang, Ed. Singapore: Springer, May 2017, pp. 1–28.
[5] K. Fokianos and A. Savvides, "On comparing several spectral densities," *Technometrics*, vol. 50, no. 3, pp. 317–331, 2008.
[6] R. Lund, H. Bassily, and B. Vidakovic, "Testing equality of stationary autocovariances," *J. Time Series Anal.*, vol. 30, pp. 332–348, May 2009.
[7] N. Ravishanker, J. R. M. Hosking, and J. Mukhopadhyay, "Spectrum-based comparison of stationary multivariate time series," *Methodol. Comput. Appl. Probab.*, vol. 12, no. 4, pp. 749–762, 2010.
[8] Y. Kakizawa, R. H. Shumway, and M. Taniguchi, "Discrimination and clustering for multivariate time series," *J. Amer. Stat. Assoc.*, vol. 93, no. 441, pp. 328–340, 1998.
[9] J. K. Tugnait, "Comparing multivariate complex random signals: Algorithm, performance analysis and application," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 934–947, Feb. 2016.
[10] P. Preuß and T. Hildebrandt, "Comparing spectral densities of stationary time series with unequal sample sizes," *Statist. Probab. Lett.*, vol. 83, no. 4, pp. 1174–1183, Apr. 2013.
[11] J. K. Tugnait, "Spectrum-based comparison of multivariate complex random signals of unequal lengths," in *Proc. 51st Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Oct./Nov. 2017, pp. 757–761.
[12] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*. Hoboken, NJ, USA: Wiley, 2006.
[13] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, Jul. 2012.
[14] M. Hubert, P. Rousseeuw, D. Vanpaemel, and T. Verdonck, "The DetS and DetMM estimators for multivariate location and scatter," *Comput. Statist. Data Anal.*, vol. 81, pp. 64–75, Jan. 2015.
[15] J. K. Tugnait and S. A. Bhaskar, "On testing for impropriety of multivariate complex-valued random sequences," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2988–3003, Jun. 2017.
[16] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Pacific Grove, CA, USA: Duxbury, 2002.
[17] D. R. Brillinger, *Time Series: Data Analysis and Theory*. New York, NY, USA: McGraw Hill, 1981.
[18] M. S. Srivastava and C. G. Khatri, *An Introduction to Multivariate Statistical Analysis*. New York, NY, USA: Elsevier, 1979.
[19] N. A. Campbell, H. P. Lopuhaä, and P. J. Rousseeuw, "On the calculation of a robust S-estimator of a covariance matrix," *Statist. Med.*, vol. 17, pp. 2685–2695, Dec. 1998.
[20] N. Muler and V. J. Yohai, "Robust estimation for vector autoregressive models," *Comput. Statist. Data Anal.*, vol. 65, pp. 68–79, Sep. 2013.

JITENDRA K. TUGNAIT (M'79–SM'93–F'94–LF'16) received the B.Sc. degree (Hons.) in electronics and electrical communication engineering from the Punjab Engineering College, Chandigarh, India, in 1971, the M.S. and E.E. degrees from Syracuse University, Syracuse, NY, USA, and the Ph.D. degree from the University of Illinois, Urbana-Champaign, IL, USA, in 1973, 1974, and 1978, respectively, all in electrical engineering.

From 1978 to 1982, he was an Assistant Professor with the Electrical and Computer Engineering, University of Iowa, Iowa City, IA, USA. He was with the Long Range Research Division, Exxon Production Research Company, Houston, TX, USA, from 1982 to 1989. He joined the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA, in 1989, as a Professor. He currently holds the title of James B. Davis Professor. His current research interests are in statistical signal processing, wireless communications, and multiple target tracking.

Dr. Tugnait was an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was a Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and a Senior Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS.