A Frequency Domain Test for Propriety of Complex-Valued Vector Time Series

Swati Chandna and A. T. Walden, *Senior Member, IEEE*

Abstract

This paper proposes a frequency domain approach to test the hypothesis that a stationary complex-valued vector time series is proper, i.e., for testing whether the vector time series is uncorrelated with its complex conjugate. If the hypothesis is rejected, frequency bands causing the rejection will be identified and might usefully be related to known properties of the physical processes. The test needs the associated spectral matrix which can be estimated by multitaper methods using, say, K tapers. Standard asymptotic distributions for the test statistic are of no use since they would require $K \to \infty$, but, as K increases so does resolution bandwidth which causes spectral blurring. In many analyses K is necessarily kept small, and hence our efforts are directed at practical and accurate methodology for hypothesis testing for small K. Our generalized likelihood ratio statistic combined with exact cumulant matching gives very accurate rejection percentages. We also prove that the statistic on which the test is based is comprised of canonical coherencies arising from our complex-valued vector time series. Frequency specific tests are combined using multiple hypothesis testing to give an overall test. Our methodology is demonstrated on ocean current data collected at different depths in the Labrador Sea. Overall this work extends results on propriety testing for complex-valued vectors to the complex-valued vector time series setting.

Index Terms

Generalized likelihood ratio test (GLRT), improper complex time series, multichannel signal, multiple hypothesis test, spectral analysis.

I. Introduction

There has long been an interest in time series motions on the complex plane: the rotary analysis method decomposes such motions into counter-rotating components which have proved particularly useful in the study of geophysical flows influenced by the rotation of the Earth [11], [12], [23], [37], [38].

S. Chandna is with Dept. of Statistical Science, University College London, London WC1E 6BT, UK (e-mail: s.chandna@ucl.ac.uk). A. T. Walden is with the Dept. of Mathematics, Imperial College London, London SW7 2AZ, UK (e-mail: a.walden@imperial.ac.uk).

Let a complex-valued p-vector-valued discrete time series be denoted $\{\boldsymbol{Z}_t\}$. This has as t-th element, $(t \in \mathbb{Z})$, the column vector $\boldsymbol{Z}_t = [Z_{1,t}, \ldots, Z_{p,t}]^T$. A length-$N$ realization of $\{\boldsymbol{Z}_t\}$ namely $\boldsymbol{z}_0, \ldots, \boldsymbol{z}_{N-1}$ has $\boldsymbol{z}_t \in \mathbb{C}^p$. In this paper we assume the p processes are jointly second-order stationary.

We propose a frequency domain approach to testing the hypothesis that a complex-valued p-vector-valued time series is proper, i.e., for testing whether the vector time series $\{\boldsymbol{Z}_t\}$ is uncorrelated with its complex conjugate $\{\boldsymbol{Z}_t^*\}$. If we denote the covariance sequence between these terms by $\{\boldsymbol{r}_{\boldsymbol{Z},\tau}\}$ then propriety corresponds to $\boldsymbol{r}_{\boldsymbol{Z},\tau} = \boldsymbol{0}$ for all $\tau \in \mathbb{Z}$, or $\boldsymbol{R}_{\boldsymbol{Z}}(f) = \boldsymbol{0}$ over the Nyquist frequency range, where $\boldsymbol{R}_{\boldsymbol{Z}}(f)$ is the Fourier transform of $\{\boldsymbol{r}_{\boldsymbol{Z},\tau}\}$. Otherwise the time series is said to be improper; the practical importance and occurrence of improper processes is discussed in, e.g., [1], [23], [26], [27] and [32].

In this paper we take as an example a multi-component complex-valued ocean current time series recorded in the Labrador Sea. Frequency domain analysis is particularly useful in a scientific setting: if the hypothesis is rejected, frequency bands causing the rejection can be identified and quite possibly related to known properties of the physical processes.

Analogous tests applicable to complex-valued random vectors — rather than time series — are given by, e.g., [33] and [39]. However, we must consider new methodology suitable for very limited degrees of freedom. Our test uses the associated spectral matrix which can be estimated by multitaper methods using, say, K tapers. Standard asymptotic distributions for the test statistic are useless as they require $K \to \infty$, but, as K increases so does resolution bandwidth which causes spectral blurring. In many analyses K is necessarily kept small, and hence our efforts are directed at practical and accurate methodology for hypothesis testing for small K. Our generalized likelihood ratio statistic combined with exact cumulant matching gives very accurate rejection percentages.

For the scalar case, $(p = 1)$, a parametric hypothesis test for propriety of complex time series is given in [34], [35]. This is based on the series being well-modelled by a Matérn process in [34] or complex autoregressive process of order one in [35], and utilises the χ^2 distribution for the test statistic, an asymptotic result. This is in contrast to our approach which (i) is suitable for $p > 1$, (ii) is nonparametric, so does not rely on a good fit to a parametric model, and (iii) develops a suitable non-asymptotic distribution for the test statistic.

Our frequency-specific test statistic is comprised of canonical coherencies arising from the complex-valued vector time series, analogous to the situation for complex-valued random vectors. Canonical analysis of real-valued vector time series has been extensively studied and utilised (e.g., [24], [31]), mostly in the context of parametric autoregressive moving-average (ARMA) models. Miyata [25] looked at real-valued vector time series, and developed canonical correlations through linear functions of discrete vector Fourier transforms of two sets of time series. Rather than work with the Fourier transforms, which

are sample values, we instead work with the orthogonal processes underlying the complex-valued vector time series, and whose variances and cross-covariances correspond exactly to the spectral components. We are thus able to define population — as well as sample — canonical coherencies for complex-valued vector time series.

We use multiple hypothesis testing to develop an overall —rather than frequency-specific — test of propriety for the time series. Our methodology is demonstrated on ocean current data collected at different depths in the Labrador Sea.

At the refereeing stage our attention was drawn to the contemporaneous conference paper [36]. Unlike [36] the current paper shows that our test statistic — which appeared first in [7, Chapter 5] — is comprised of canonical coherencies and we give a simple scaled F distribution (obtained via cumulant matching) for deriving rejection percentages. The way the tests are constructed and used in the two cases are also different, and this is discussed in Section IX.

A. Contributions

Following some background in Section II on complex-valued time series, and the statistical properties of their spectral matrix estimators under the Gaussian stationary assumption for $\{\boldsymbol{Z}_t\}$, the contributions of this paper are as follows:

1) In Section III we formally derive the canonical coherencies for $\{\boldsymbol{Z}_t\}$ and $\{\boldsymbol{Z}_t^*\}$ and show in Section IV how a test statistic $T(f)$ for testing $\boldsymbol{R}_{\boldsymbol{Z}}(f) = \boldsymbol{0}$ arises from the sample canonical coherencies.
2) After giving further research context in Section V, we carefully study the statistical properties of $M(f) = -2K \log T(f)$ in Section VI, concentrating on the small K case. We show that Box's scaled chi-square approximation is exact for $p = 1$ but not for $p > 1$, and we derive the cumulants of $M(f)$.
3) In Section VII we show that for $p > 1$ and small K matching the first three cumulants of $M(f)$ exactly to a scaled F distribution performs at least as well as competitor methods.
4) A simulation study is given in Section VIII which supports the use of the scaled F approximation for $M(f)$ for the complex-valued vector time series setting.
5) In Section IX we describe the use of multiple hypothesis tests for developing an overall — rather than frequency-specific — test of propriety for the time series. Simulations show good behaviour under both the null hypothesis (Section IX-D) and under alternatives (Section IX-E).
6) A data analysis using vector valued oceanographic time series is given in Section X which shows that when propriety is rejected, the frequency domain approach usefully shows which frequency

bands cause the rejection, which may be linked to the physical processes involved.

II. Background

We will first show how propriety is linked to matrices in the frequency domain and then discuss the estimation of these matrices from complex vector-valued Gaussian stationary time series.

A. Some Definitions

We consider a complex-valued *p-vector-valued discrete time stochastic process* $\{\boldsymbol{Z}_t\}$ whose tth element, $t \in \mathbb{Z}$, is the column vector $\boldsymbol{Z}_t = [Z_{1,t}, \ldots, Z_{p,t}]^T$, and without loss of generality take each component process to have zero mean. The sample interval is Δ_{t} and the Nyquist frequency is $f_{\mathcal{N}} = 1/(2\Delta_{\mathrm{t}})$. We assume the p processes are *jointly* second-order stationary (SOS), i.e., the covariance $\mathrm{cov}\{Z_{l,t+\tau}, Z_{m,t}\} \stackrel{\mathrm{def}}{=} E\{Z_{l,t+\tau} Z^*_{m,t}\}$ and the complementary-covariance $\mathrm{ccov}\{Z_{l,t+\tau}, Z_{m,t}\} \stackrel{\mathrm{def}}{=} E\{Z_{l,t+\tau} Z_{m,t}\}$, $1 \le l, m \le p$, are functions of τ only. Note that $\mathrm{ccov}\{Z_{l,t+\tau}, Z_{m,t}\} = \mathrm{cov}\{Z_{l,t+\tau}, Z^*_{m,t}\}$, the covariance between one process and the complex conjugate of the other.

A matrix covariance sequence is then given by $\boldsymbol{s}_{\boldsymbol{Z},\tau} = E\{\boldsymbol{Z}_{t+\tau} \boldsymbol{Z}_t^H\}$, $\tau \in \mathbb{Z}$, where superscript H denotes Hermitian (complex-conjugate) transpose; we define the (l,m)th element as $s_{\boldsymbol{Z},lm,\tau} \stackrel{\mathrm{def}}{=} (\boldsymbol{s}_{\boldsymbol{Z},\tau})_{lm}$. A matrix complementary-covariance follows as $\boldsymbol{r}_{\boldsymbol{Z},\tau} = E\{\boldsymbol{Z}_{t+\tau} \boldsymbol{Z}_t^T\}$, $\tau \in \mathbb{Z}$, with (l,m)th element $r_{\boldsymbol{Z},lm,\tau} \stackrel{\mathrm{def}}{=} (\boldsymbol{r}_{\boldsymbol{Z},\tau})_{lm}$. From their definitions we see that

$$s_{\boldsymbol{Z},lm,\tau} = s^*_{\boldsymbol{Z},ml,-\tau};\ r_{\boldsymbol{Z},lm,\tau} = r_{\boldsymbol{Z},ml,-\tau},\ 1 \le l, m \le p.$$

We assume $\sum_{\tau=-\infty}^{\infty} |s_{\boldsymbol{Z},lm,\tau}| < \infty$ and $\sum_{\tau=-\infty}^{\infty} |r_{\boldsymbol{Z},lm,\tau}| < \infty$, for $1 \le l \le m \le p$, which means that the Fourier transforms $S_{\boldsymbol{Z},lm}(f)$ and $R_{\boldsymbol{Z},lm}(f)$ for $1 \le l, m \le p$, exist and are bounded and continuous. For $|f| \le f_{\mathcal{N}}$, the corresponding matrices are $\boldsymbol{S}_{\boldsymbol{Z}}(f) = \Delta_{\mathrm{t}} \sum_{\tau=-\infty}^{\infty} \boldsymbol{s}_{\boldsymbol{Z},\tau}\, \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}}$ and $\boldsymbol{R}_{\boldsymbol{Z}}(f) = \Delta_{\mathrm{t}} \sum_{\tau=-\infty}^{\infty} \boldsymbol{r}_{\boldsymbol{Z},\tau} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}}$, respectively. Then

$$\boldsymbol{r}_{\boldsymbol{Z},\tau} = \boldsymbol{r}^T_{\boldsymbol{Z},-\tau} \implies \boldsymbol{R}_{\boldsymbol{Z}}(f) = \boldsymbol{R}^T_{\boldsymbol{Z}}(-f). \tag{1}$$

The covariance stationarity means that there exists an orthogonal process $\boldsymbol{Z}(f)$ such that $\boldsymbol{Z}_t = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi f t} \mathrm{d}\boldsymbol{Z}(f)$ [41, p. 317] and $\boldsymbol{Z}^*_t = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi f t} \mathrm{d}\boldsymbol{Z}^*(-f)$, with [38]

$$E\{\mathrm{d}\boldsymbol{Z}(f) \mathrm{d}\boldsymbol{Z}^H(f')\} = \begin{cases} \boldsymbol{S}_{\boldsymbol{Z}}(f)\mathrm{d}f, & f = f' \\ 0, & \text{otherwise,} \end{cases}$$

and

$$E\{\mathrm{d}\boldsymbol{Z}(f) \mathrm{d}\boldsymbol{Z}^T(f')\} = \begin{cases} \boldsymbol{R}_{\boldsymbol{Z}}(f)\mathrm{d}f, & f = -f' \\ 0, & \text{otherwise.} \end{cases}$$

B. Proper Processes

If $\boldsymbol{r}_{\boldsymbol{Z},\tau} = \boldsymbol{0}$ for all $\tau \in \mathbb{Z}$, or $\boldsymbol{R}_{\boldsymbol{Z}}(f) = \boldsymbol{0}$ for all $|f| \leq f_{\mathcal{N}}$, then the process $\{\boldsymbol{Z}_t\}$ is said to be *proper*. Equivalently we see that if $\{\boldsymbol{Z}_t\}$ is uncorrelated with its complex conjugate $\{\boldsymbol{Z}_t^*\}$, then the vector-valued process is proper. This paper considers the problem of testing that the vector process is proper. From (1) we see that if $\boldsymbol{R}_{\boldsymbol{Z}}(f) = \boldsymbol{0}$ for $f > 0$ then it is also $\boldsymbol{0}$ for $f < 0$. Hence we need to test:

$$H_0 : \boldsymbol{R}_{\boldsymbol{Z}}(f) = \boldsymbol{0} \text{ for all } f \leq f_{\mathcal{N}}. \tag{2}$$

Remark 1: Based on the naming convention adopted in [32, p. 41] for complex-valued vectors, an alternative would be to call the component processes 'jointly proper.'

C. Spectral Matrices

Here we define spectral matrices for both the so-called composite real and augmented complex representations of the series. Let

$$Z_{l,t} = X_{l,t} + \mathrm{i}Y_{l,t}, \tag{3}$$

with $\{X_{l,t}\}$ and $\{Y_{l,t}\}$ real-valued, for $l = 1, \ldots, p$. The composite real representation of the series is given by $\boldsymbol{V}_t = [\boldsymbol{X}_t^T, \boldsymbol{Y}_t^T]^T = [X_{1,t}, \ldots, X_{p,t}, Y_{1,t}, \ldots, Y_{p,t}]^T$, a real $2p$-dimensional vector-valued Gaussian stationary process. If

$$\boldsymbol{T} \stackrel{\text{def}}{=} \begin{bmatrix} \boldsymbol{I}_p & \mathrm{i}\boldsymbol{I}_p \\ \boldsymbol{I}_p & -\mathrm{i}\boldsymbol{I}_p \end{bmatrix}, \tag{4}$$

we see that

$$\boldsymbol{T}\boldsymbol{V}_t = \begin{bmatrix} \boldsymbol{X}_t + \mathrm{i}\boldsymbol{Y}_t \\ \boldsymbol{X}_t - \mathrm{i}\boldsymbol{Y}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z}_t \\ \boldsymbol{Z}_t^* \end{bmatrix} = \boldsymbol{U}_t, \tag{5}$$

where $\boldsymbol{U}_t = [\boldsymbol{Z}_t^T, \boldsymbol{Z}_t^H]^T = [Z_{1,t}, \ldots, Z_{p,t}, Z_{1,t}^*, \ldots, Z_{p,t}^*]^T$ is the augmented complex representation of the series, a complex $2p$-dimensional vector-valued Gaussian stationary process.

The spectral matrix for $\boldsymbol{V}_t$ is given by

$$\boldsymbol{S}_{\boldsymbol{V}}(f) = \begin{bmatrix} \boldsymbol{S}_{\boldsymbol{X}\boldsymbol{X}}(f) & \boldsymbol{S}_{\boldsymbol{X}\boldsymbol{Y}}(f) \\ \boldsymbol{S}_{\boldsymbol{Y}\boldsymbol{X}}(f) & \boldsymbol{S}_{\boldsymbol{Y}\boldsymbol{Y}}(f) \end{bmatrix} \in \mathbb{C}^{2p \times 2p}. \tag{6}$$

The spectral matrix for $\boldsymbol{U}_t$ is $\boldsymbol{S}_{\boldsymbol{U}}(f) = \boldsymbol{T}\boldsymbol{S}_{\boldsymbol{V}}(f)\boldsymbol{T}^H$ and has the form

$$\boldsymbol{S}_{\boldsymbol{U}}(f) = \begin{bmatrix} \boldsymbol{S}_{\boldsymbol{Z}}(f) & \boldsymbol{R}_{\boldsymbol{Z}}(f) \\ \boldsymbol{R}_{\boldsymbol{Z}}^H(f) & \boldsymbol{S}_{\boldsymbol{Z}}^T(-f) \end{bmatrix} \in \mathbb{C}^{2p \times 2p}. \tag{7}$$

The matrix $\boldsymbol{S}_{\boldsymbol{U}}(f)$ can be written in the alternative covariance matrix form $E\{\boldsymbol{U}(f)\,\boldsymbol{U}^H(f)\} = \boldsymbol{S}_{\boldsymbol{U}}(f)\mathrm{d}f$, where

$$\boldsymbol{U}(f) \stackrel{\text{def}}{=} [\mathrm{d}\boldsymbol{Z}^T(f), \mathrm{d}\boldsymbol{Z}^H(-f)]^T. \tag{8}$$

Remark 2: Consider (7). We see that testing $\boldsymbol{R}_{\boldsymbol{Z}}(f) = \boldsymbol{0}$ is the same as testing the independence of the two complex Gaussian p-vectors, $\mathrm{d}\boldsymbol{Z}(f)$ and $\mathrm{d}\boldsymbol{Z}^*(-f)$. This simple fact enables us to utilise results from the propriety testing of complex-valued vectors.

D. Estimation

With $\boldsymbol{S}_{\boldsymbol{U}}(f)$ in (7) seen to be of central importance we now turn to its estimation and related statistical properties.

Given a length-N sample $\boldsymbol{V}_0, \ldots, \boldsymbol{V}_{N-1}$, form $h_{k,t}\boldsymbol{V}_t$ using a suitable set of K length-N orthonormal data taper sequences $\{h_{k,t}\}, k = 0, \ldots, K-1$, and compute $\boldsymbol{J}_{\boldsymbol{V},k}(f) = \Delta_{\mathrm{t}}^{1/2} \sum_{t=0}^{N-1} h_{k,t}\boldsymbol{V}_t \mathrm{e}^{-\mathrm{i}2\pi f t \Delta_{\mathrm{t}}}$. In this work we use sine tapers (e.g., [40]).

As $N \to \infty$, with the number of degrees of freedom, K fixed, and with the given taper properties, $\{\boldsymbol{J}_{\boldsymbol{V},k}(f), k = 0, 1, \ldots, K-1\}$ are proper, independent and identically distributed random vectors distributed as

$$\boldsymbol{J}_{\boldsymbol{V},k}(f) \overset{\mathrm{d}}{=} \mathcal{N}_{2p}^{C}(\boldsymbol{0}, \boldsymbol{S}_{\boldsymbol{V}}(f)), \;\; 0 < |f| < f_{\mathcal{N}}, \tag{9}$$

for $k = 0, \ldots, K-1$ (e.g., [8]). As $\boldsymbol{J}_{\boldsymbol{U},k}(f) = \boldsymbol{T}\boldsymbol{J}_{\boldsymbol{V},k}(f)$, as $N \to \infty$, with K fixed, $\{\boldsymbol{J}_{\boldsymbol{U},k}(f), k = 0, 1, \ldots, K-1\}$ are also a set of proper, independent and identically distributed random vectors each of which are distributed as

$$\boldsymbol{J}_{\boldsymbol{U},k}(f) \overset{\mathrm{d}}{=} \mathcal{N}_{2p}^{C}(\boldsymbol{0}, \boldsymbol{S}_{\boldsymbol{U}}(f)), \;\; 0 < |f| < f_{\mathcal{N}}. \tag{10}$$

The probability density function (PDF) of $\boldsymbol{J}_{\boldsymbol{U},k}(f)$ — a proper Gaussian vector in $\mathbb{C}^{2p}$ is given by [29]

$$\pi^{-p}[\det\{\boldsymbol{S}_{\boldsymbol{U}}(f)\}]^{-1} \exp\left\{-\boldsymbol{J}_{\boldsymbol{U},k}^{H}(f)\boldsymbol{S}_{\boldsymbol{U}}^{-1}(f)\boldsymbol{J}_{\boldsymbol{U},k}(f)\right\}. \tag{11}$$

The independence of $\boldsymbol{J}_{\boldsymbol{U},k}(f)$'s allows us to write the joint PDF of $\boldsymbol{J}_{\boldsymbol{U},0}(f), \ldots, \boldsymbol{J}_{\boldsymbol{U},K-1}(f)$ as the product of their marginal densities given by (11). So the likelihood function, $g_{\boldsymbol{J}}(\boldsymbol{S}_{\boldsymbol{U}}(f)|\boldsymbol{J}_{\boldsymbol{U},0}(f), \ldots, \boldsymbol{J}_{\boldsymbol{U},K-1}(f))$, of $\boldsymbol{S}_{\boldsymbol{U}}(f)$ given $\boldsymbol{J}_{\boldsymbol{U},0}(f), \ldots, \boldsymbol{J}_{\boldsymbol{U},K-1}(f)$, is given by

$$[\pi^p \det\{\boldsymbol{S}_{\boldsymbol{U}}(f)\}]^{-K} \exp\left\{-\sum_{k=0}^{K-1} \boldsymbol{J}_{\boldsymbol{U},k}^{H}(f)\boldsymbol{S}_{\boldsymbol{U}}^{-1}(f)\boldsymbol{J}_{\boldsymbol{U},k}(f)\right\}. \tag{12}$$

Now $\hat{\boldsymbol{S}}_{\boldsymbol{U}}(f)$ is the sample covariance matrix of $\{\boldsymbol{J}_{\boldsymbol{U},k}(f); k = 0, 1, \ldots, K-1\}$, i.e.,

$$\hat{\boldsymbol{S}}_{\boldsymbol{U}}(f) = \frac{1}{K}\sum_{k=0}^{K-1} \boldsymbol{J}_{\boldsymbol{U},k}(f)\boldsymbol{J}_{\boldsymbol{U},k}^{H}(f) = \begin{bmatrix} \hat{\boldsymbol{S}}_{\boldsymbol{Z}}(f) & \hat{\boldsymbol{R}}_{\boldsymbol{Z}}(f) \\ \hat{\boldsymbol{R}}_{\boldsymbol{Z}}^{H}(f) & \hat{\boldsymbol{S}}_{\boldsymbol{Z}}^{T}(-f) \end{bmatrix}. \tag{13}$$

Noting that the argument of $\exp\{\cdot\}$ in (12) is scalar, and so is equal to its trace, and recalling the linearity and cyclicity of the trace operator, we can write

$$g_{\boldsymbol{J}} = [\pi^p \det\{\boldsymbol{S}_{\boldsymbol{U}}(f)\}]^{-K} \exp\left\{-K\mathrm{tr}\{\boldsymbol{S}_{\boldsymbol{U}}^{-1}(f)\hat{\boldsymbol{S}}_{\boldsymbol{U}}(f)\}\right\}; \tag{14}$$

dependence of $g_{\boldsymbol{J}}$ on its arguments are not shown explicitly.

For a finite value of N, $\{\boldsymbol{J}_{\boldsymbol{U},k}(f); k = 0, 1, \ldots, K-1\}$ are proper random variables with

$$\boldsymbol{J}_{\boldsymbol{U},k}(f) \stackrel{\mathrm{d}}{=} \mathcal{N}_{2p}^{C}(\boldsymbol{0}, \boldsymbol{S}_{\boldsymbol{U}}(f)), \quad W_N < |f| < f_{\mathcal{N}} - W_N, \tag{15}$$

where $[-W_N, W_N]$ is the extent of the spectral window induced by tapering [8]. For sine tapers

$$W_N = (K+1)/[2(N+1)\Delta_{\mathrm{t}}], \tag{16}$$

(e.g., [40]). Therefore, in practice, we have to restrict interest to frequencies in the range $W_N < |f| < f_{\mathcal{N}} - W_N$.

III. Canonical Coherencies

From Remark 2 we know that the structure of the testing problem will be related to measures of coherence between vector-valued processes, and so we next turn our attention to the idea of canonical coherence. We start by developing the framework required for complex-valued time series.

Let $\{\boldsymbol{\xi}_t\}$ be the cross-correlation of complex-valued deterministic matrix sequence $\{\boldsymbol{A}_t\}$ with time series $\{\boldsymbol{Z}_t\}$:

$$\boldsymbol{\xi}_t = \boldsymbol{A}^* \star \boldsymbol{Z}_t \stackrel{\mathrm{def}}{=} \sum_{u=-\infty}^{\infty} \boldsymbol{A}_u^* \boldsymbol{Z}_{t+u}.$$

Next let $\{\boldsymbol{\eta}_t\}$ be the cross-correlation of complex-valued deterministic matrix sequence $\{\boldsymbol{B}_t\}$ with time series $\{\boldsymbol{Z}_t^*\}$:

$$\boldsymbol{\eta}_t = \boldsymbol{B}^* \star \boldsymbol{Z}_t^* \stackrel{\mathrm{def}}{=} \sum_{u=-\infty}^{\infty} \boldsymbol{B}_u^* \boldsymbol{Z}_{t+u}^*.$$

Component-wise we have

$$\begin{bmatrix} \xi_{1,t} \\ \xi_{2,t} \\ \vdots \\ \xi_{p,t} \end{bmatrix} = \sum_u \begin{bmatrix} a_{11,u}^* & \cdots & \cdots & a_{1p,u}^* \\ a_{21,u}^* & \cdots & \cdots & a_{2p,u}^* \\ \vdots & & & \vdots \\ a_{p1,u}^* & \cdots & \cdots & a_{pp,u}^* \end{bmatrix} \begin{bmatrix} Z_{1,t+u} \\ Z_{2,t+u} \\ \vdots \\ Z_{p,t+u} \end{bmatrix}. \tag{17}$$

So, for $j = 1, \ldots, p$,

$$\xi_{j,t} = \sum_u a_{j1,u}^* Z_{1,t+u} + \cdots + \sum_u a_{jp,u}^* Z_{p,t+u}. \tag{18}$$

The spectral representation theorem allows us to write $\xi_{j,t}, j = 1, \ldots, p$ and $Z_{l,t}$, $l = 1, \ldots, p$, as

$$\xi_{j,t} = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathrm{e}^{\mathrm{i}2\pi f t\,\Delta_{\mathrm{t}}} \mathrm{d}Z_{\xi_j}(f); \;\; Z_{l,t} = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathrm{e}^{\mathrm{i}2\pi f t\,\Delta_{\mathrm{t}}} \mathrm{d}Z_l(f).$$

Substituting the spectral representation for $Z_{l,t}$ in the lth sum on the right of (18), we get

$$\sum_u a_{jl,u}^* Z_{l,t+u} \;\; = \;\; \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathrm{e}^{\mathrm{i}2\pi f t\,\Delta_{\mathrm{t}}} A_{jl}^*(f)\, \mathrm{d}Z_l(f),$$

where $A_{jl}(f) = \sum_u a_{jl,u} \mathrm{e}^{-\mathrm{i}2\pi f u \Delta_t}$. Proceeding in analogous fashion, and using the fact that the orthogonal process in a spectral representation is unique [10, p. 34], we obtain

$$\begin{aligned} \mathrm{d}Z_{\xi_j}(f) &= A^*_{j1}(f)\mathrm{d}Z_1(f) + \ldots + A^*_{jp}(f)\mathrm{d}Z_p(f) \\ &\stackrel{\text{def}}{=} \boldsymbol{A}^H_j(f)\mathrm{d}\boldsymbol{Z}(f). \end{aligned}$$

So

$$\xi_{j,t} = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathrm{e}^{\mathrm{i}2\pi f t \, \Delta_t} \boldsymbol{A}^H_j(f)\mathrm{d}\boldsymbol{Z}(f). \tag{19}$$

For $\{\boldsymbol{\eta}_t\}$ a similar procedure gives

$$\begin{aligned} \mathrm{d}Z_{\eta_j}(f) &= B^*_{j1}(f)\mathrm{d}Z^*_1(-f) + \ldots + B^*_{jp}(f)\mathrm{d}Z^*_p(-f) \\ &\stackrel{\text{def}}{=} \boldsymbol{B}^H_j(f)\mathrm{d}\boldsymbol{Z}^*(-f), \end{aligned}$$

and

$$\eta_{j,t} = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathrm{e}^{\mathrm{i}2\pi f t \, \Delta_t} \boldsymbol{B}^H_j(f)\mathrm{d}\boldsymbol{Z}^*(-f). \tag{20}$$

The usual definition of the (magnitude squared) coherencies $\gamma^2_j(f)$ between series $\{\xi_{j,t}\}$ and $\{\eta_{j,t}\}$ is

$$\begin{aligned} \gamma^2_j(f) &= \frac{|E\{\mathrm{d}Z_{\xi_j}(f)\mathrm{d}Z^H_{\eta_j}(f)\}|^2}{E\{|\mathrm{d}Z_{\xi_j}(f)|^2\}E\{|\mathrm{d}Z_{\eta_j}(f)|^2\}} \\ &= |\mathrm{corr}\{\mathrm{d}Z_{\xi_j}(f), \mathrm{d}Z_{\eta_j}(f)\}|^2. \end{aligned}$$

Remark 3: It should be emphasized that throughout we use the usual definition of coherence as a magnitude squared quantity, basically a squared correlation coefficient.

Consider the following procedure. Find $\boldsymbol{A}_1(f)$ and $\boldsymbol{B}_1(f)$ such that $|K_{11}(f)| = |\mathrm{corr}\{\mathrm{d}Z_{\xi_1}(f), \mathrm{d}Z_{\eta_1}(f)\}|$ is maximized. Next find $\boldsymbol{A}_2(f)$ and $\boldsymbol{B}_2(f)$ such that $|K_{22}(f)| = |\mathrm{corr}\{\mathrm{d}Z_{\xi_2}(f), \mathrm{d}Z_{\eta_2}(f)\}|$ is maximized, subject to $\mathrm{d}Z_{\xi_2}(f), \mathrm{d}Z_{\eta_2}(f)$ being uncorrelated with $\mathrm{d}Z_{\xi_1}(f), \mathrm{d}Z_{\eta_1}(f)$. In general, at step j for $j = 2, \ldots, p$, $\boldsymbol{A}_j(f)$ and $\boldsymbol{B}_j(f)$ are found such that $|K_{jj}(f)| = |\mathrm{corr}\{\mathrm{d}Z_{\xi_j}(f), \mathrm{d}Z_{\eta_j}(f)\}|$ is maximized subject to $\mathrm{d}Z_{\xi_j}(f), \mathrm{d}Z_{\eta_j}(f)$ being uncorrelated with $\mathrm{d}Z_{\xi_k}(f), \mathrm{d}Z_{\eta_k}(f)$ for $1 \le k < j$.

Lemma 1: The canonical coherencies $l^2_j(f) \stackrel{\text{def}}{=} |K_{jj}(f))|^2, \quad j = 1, \ldots, p$ and $\boldsymbol{A}_j(f)$ and $\boldsymbol{B}_j(f)$ for $j = 1, \ldots, p$, defined above are eigenvalues and eigenvectors as follows:

$$\begin{aligned} \boldsymbol{S}^{-1}_{\boldsymbol{Z}}(f)\boldsymbol{R}_{\boldsymbol{Z}}(f)\boldsymbol{S}^{-T}_{\boldsymbol{Z}}(-f)\boldsymbol{R}^H_{\boldsymbol{Z}}(f)\boldsymbol{A}_j(f) &= l^2_j(f)\boldsymbol{A}_j(f) \\ \boldsymbol{S}^{-T}_{\boldsymbol{Z}}(-f)\boldsymbol{R}^H_{\boldsymbol{Z}}(f)\boldsymbol{S}^{-1}_{\boldsymbol{Z}}(f)\boldsymbol{R}_{\boldsymbol{Z}}(f)\boldsymbol{B}_j(f) &= l^2_j(f)\boldsymbol{B}_j(f). \end{aligned} \tag{21}$$

Moreover we have that as a result,

$$\mathrm{corr}\{\mathrm{d}Z_{\xi_j}(f), \mathrm{d}Z_{\eta_k}(f)\} = 0, \;\; \text{for } \; j, k = 1, \ldots, p; j \neq k. \tag{22}$$

Proof: With notational changes the result follows from [5, Theorem 10.3.2]. ■

Remark 4: From Lemma 1 the optimal $\boldsymbol{A}_j(f)$ and $\boldsymbol{B}_j(f)$ give rise to the jth pair of canonical series via (19) and (20).

IV. Generalized Likelihood Ratio Test (GLRT)

To develop a GLRT for testing $\boldsymbol{R_Z}(f) = \mathbf{0}$ at a specific frequency, we will make use of the structure of the covariance matrix (13).

A. Formulation

The GLRT statistic for the frequency-specific test

$$H_{0f} : \boldsymbol{R_Z}(f) = \mathbf{0} \quad \text{versus} \quad H_{1f} : \boldsymbol{R_Z}(f) \neq \mathbf{0}, \tag{23}$$

is given by ratio of the likelihood function (14) with $\boldsymbol{S_U}(f)$ constrained to have zero off-diagonal blocks ($\boldsymbol{R_Z}(f) = \mathbf{0}$) to the likelihood function with $\boldsymbol{S_U}(f)$ unconstrained, i.e.,

$$\frac{\max\limits_{\boldsymbol{S_U}(f):\boldsymbol{R_Z}(f)=\mathbf{0}} g_{\boldsymbol{J}}}{\max\limits_{\boldsymbol{S_U}(f)} g_{\boldsymbol{J}}} \stackrel{\text{def}}{=} L_G(f). \tag{24}$$

The unconstrained maximum likelihood estimate of the covariance matrix $\boldsymbol{S_U}(f)$ is given by the corresponding sample covariance matrix $\hat{\boldsymbol{S}}_{\boldsymbol{U}}(f)$ in (13), thus maximum likelihood estimate of $\boldsymbol{S_U}(f)$ under the constraint $\boldsymbol{R_Z}(f) = \mathbf{0}$ is,

$$\breve{\boldsymbol{S}}_{\boldsymbol{U}}(f) = \begin{bmatrix} \hat{\boldsymbol{S}}_{\boldsymbol{Z}}(f) & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{S}}_{\boldsymbol{Z}}^{T}(-f) \end{bmatrix}. \tag{25}$$

Following (2) we will only calculate $T(f) \stackrel{\text{def}}{=} L_G^{1/K}(f)$ over the positive frequency range $W_N < f < f_{\mathcal{N}} - W_N$.

By analogy to [33, eqn. (13)]

$$T(f) \;\; = \;\; \det\{\boldsymbol{I}_p - \hat{\boldsymbol{S}}_{\boldsymbol{Z}}^{-1}(f)\hat{\boldsymbol{R}}_{\boldsymbol{Z}}(f)\hat{\boldsymbol{S}}_{\boldsymbol{Z}}^{-T}(-f)\hat{\boldsymbol{R}}_{\boldsymbol{Z}}^{H}(f)\}. \tag{26}$$

Again, by analogy to [33, p. 434],

$$T(f) = \frac{\det\{\hat{\boldsymbol{S}}_{\boldsymbol{U}}(f)\}}{\det\{\breve{\boldsymbol{S}}_{\boldsymbol{U}}(f)\}} = \frac{\det\{\hat{\boldsymbol{S}}_{\boldsymbol{U}}(f)\}}{\det\{\hat{\boldsymbol{S}}_{\boldsymbol{Z}}(f)\}\det\{\hat{\boldsymbol{S}}_{\boldsymbol{Z}}(-f)\}}, \tag{27}$$

which is a convenient form for computation. Other forms are given in [7, Section 5.1].

By definition of the GLR test statistic (24), we shall reject the null hypothesis of $\boldsymbol{R_Z}(f) = \mathbf{0}$, for small values of $T(f)$. For a given size α, the rule is to reject H_{0f} iff

$$T(f; N, K, p) \leq c, \tag{28}$$

where $\Pr(T(f;N,K,p) \leq c|H_0) = \alpha$. Here we have used the more precise notation $T(f;N,K,p)$ which emphasizes the dependence of the GLR test on (i) the sample size N, (ii) the number of tapers K (also the number of complex degrees of freedom), and (iii) dimension p of the complex time series.

B. Invariance

We now consider transformations under which the test statistic $T(f)$ is invariant. These follow from [33, p. 434] with some modification. Now $\boldsymbol{R}_{\boldsymbol{Z}}(f)\mathrm{d}f \stackrel{\mathrm{def}}{=} E\{\mathrm{d}\boldsymbol{Z}(f)\mathrm{d}\boldsymbol{Z}^T(-f)\}$. Apply $\boldsymbol{L}(f) \in \mathbb{C}^{p\times p}$ to $\mathrm{d}\boldsymbol{Z}(f)$ so that $\mathrm{d}\boldsymbol{Z}(f) \rightarrow \boldsymbol{L}(f)\mathrm{d}\boldsymbol{Z}(f)$, and therefore $\mathrm{d}\boldsymbol{Z}^T(-f) \rightarrow \boldsymbol{L}^*(-f)\mathrm{d}\boldsymbol{Z}^T(-f)$. Then

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{Z}}(f) &= \boldsymbol{0} \implies E\{\boldsymbol{L}(f)\mathrm{d}\boldsymbol{Z}(f)[\boldsymbol{L}^*(-f)\mathrm{d}\boldsymbol{Z}^T(-f)]^H\} \\ &= \boldsymbol{L}(f)\boldsymbol{R}_{\boldsymbol{Z}}(f)\mathrm{d}f\boldsymbol{L}^T(-f) = \boldsymbol{0},\end{aligned}$$

i.e., $\boldsymbol{R}_{\boldsymbol{Z}}(f) = \boldsymbol{0}$ is invariant to the linear transformation $\mathrm{d}\boldsymbol{Z}(f) \rightarrow \boldsymbol{L}(f)\mathrm{d}\boldsymbol{Z}(f)$. So the decision rule for our GLR test must be likewise invariant.

Under this transformation,

$$\boldsymbol{U}(f) \rightarrow \begin{bmatrix}\boldsymbol{L}(f) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{L}^*(-f)\end{bmatrix}\boldsymbol{U}(f) \stackrel{\mathrm{def}}{=} \boldsymbol{Q}(f)\boldsymbol{U}(f),$$

so that we require invariance under the group action $\boldsymbol{S}_{\boldsymbol{U}}(f) \rightarrow \boldsymbol{Q}(f)\boldsymbol{S}_{\boldsymbol{U}}(f)\boldsymbol{Q}^H(f)$.

Under the null hypothesis the choice $\boldsymbol{L}(f) = \boldsymbol{S}_{\boldsymbol{Z}}^{-1/2}(f)$ (which exists for $\boldsymbol{S}_{\boldsymbol{Z}}(f)$ positive definite) renders the matrix $\boldsymbol{S}_{\boldsymbol{U}}(f)$ equal to $\boldsymbol{I}_{2p}$ and so under the null hypothesis we can always replace $\boldsymbol{S}_{\boldsymbol{U}}(f)$ by $\boldsymbol{I}_{2p}$ without loss of generality.

From Lemma 1 we know that the eigenvalues $l_j^2(f)$ of $\boldsymbol{S}_{\boldsymbol{Z}}^{-1}(f)\boldsymbol{R}_{\boldsymbol{Z}}(f)\boldsymbol{S}_{\boldsymbol{Z}}^{-T}(-f)\boldsymbol{R}_{\boldsymbol{Z}}^H(f)$ are canonical coherencies which are invariant under the group action specified above; moreover, the corresponding empirical or sample canonical coherencies are maximal invariant and the GLR statistic — which requires this invariance — must be a function of them.

Let $\ell_j^2(f), j = 1,\ldots,p$, be the sample versions of the canonical coherencies $l_j^2(f)$ between $\mathrm{d}\boldsymbol{Z}(f)$ and $\mathrm{d}\boldsymbol{Z}^*(-f)$. From (21) they are the sample eigenvalues of $\hat{\boldsymbol{S}}_{\boldsymbol{Z}}^{-1}(f)\hat{\boldsymbol{R}}_{\boldsymbol{Z}}(f)\hat{\boldsymbol{S}}_{\boldsymbol{Z}}^{-T}(-f)\hat{\boldsymbol{R}}_{\boldsymbol{Z}}^H(f)$. From (26) it follows that for $W_N < f < f_{\mathcal{N}} - W_N$, (compare with [33, eqn. (21)]),

$$T(f) = \prod_{j=1}^{p}(1-\ell_j^2(f)). \tag{29}$$

V. Research Context

In view of Remark 2, the GLR test based on (27) falls in the class of multiple independence tests in multivariate statistics theory. Some distributional results for the complex case were given in [19]

but did not include the case of interest here, namely two p-vectors. A later paper [13] gave the exact distribution of a power of $T(f)$ but this involves an infinite sum with very complicated components; small K approximations were not discussed. Other relevant results can be found in [16] and [20], and these are discussed in detail in Section VII-A.

The statistic $T(f)$ is the frequency-domain time series analogue to those used in [28], [33] and [39] to examine independence between a Gaussian random vector and its complex conjugate. In [28], [33] a complex formulation was maintained but only an asymptotic approach to testing was considered. In [39] a real-valued representation of the problem was used and Box's scaled chi-square method was used to improve on the asymptotic critical values. In the rest of this paper we adopt the complex formulation, derive Box's refinement, but also improve on it for $p > 1$ by *exactly* matching the first three cumulants to a scaled F-distribution. (We point out that Box's refinement is exact for $p = 1$.) This latter F-method is very simple to implement practically, involving only the first three polygamma functions.

We emphasize that our efforts are directed at practical and accurate methodology for small K. This is important in a time series setting where as K increases so does resolution bandwidth which potentially causes spectral blurring. In many analyses K must necessarily be kept small. In the remainder of this paper we will always assume any frequency under consideration to lie in the interval $W_N < f < f_N - W_N$ unless stated otherwise.

VI. Basic Properties of Test Statistic

We now derive some statistical properties for simple functions of the test statistic under the null hypothesis H_{0f}. We start with the asymptotic $K \to \infty$ case, and then look at results suitable for the practically useful case of small K.

A. Asymptotic Behaviour

The application of Wilks' theorem [42, p. 132] gives that under H_{0f}, as $K \to \infty$,

$$M(f) \stackrel{\text{def}}{=} -2 \log L_G(f) = -2K \log T(f) \stackrel{\text{d}}{\to} \chi^2_\nu \tag{30}$$

where $\stackrel{\text{d}}{\to}$ denotes convergence in distribution and χ^2_ν denotes the chi-square distribution with ν degrees of freedom. Here ν is the difference between the number of free real parameters under H_{0f} and H_{1f}. Comparing $\breve{\boldsymbol{S}}_{\boldsymbol{U}}(f)$ in (25) (for H_{0f}) and $\boldsymbol{S}_{\boldsymbol{U}}(f)$ in (7) (for H_{1f}) we note that $\boldsymbol{R}^H_{\boldsymbol{Z}}(f)$ follows directly from $\boldsymbol{R}_{\boldsymbol{Z}}(f)$ so that there is only an additional $2p^2$ degrees of freedom, i.e., those contributed by $\boldsymbol{R}_{\boldsymbol{Z}}(f)$. Hence we have $\nu = 2p^2$.

While (30) is a very useful and convenient result when the exact distribution of the GLR test statistic is analytically intractable, K here denotes the number of tapers used for multitaper spectral estimation and

not the sample size N. For a given value of N, K could be around 10 or less. Since (30) is an asymptotic result, K must be sufficiently large to expect a reasonable χ^2_ν approximation to $-2K \log T(f)$. Since K may not be large in a time series setting, a small-K approximation to the distribution of the test statistic under the null hypothesis is imperative.

B. Moments

Since $\boldsymbol{J}_{\boldsymbol{U},k}(f), k = 0, \ldots, K-1$, are Gaussian distributed random vectors, from (13) it follows that

$$\boldsymbol{A}(f) \stackrel{\text{def}}{=} K\hat{\boldsymbol{S}}_{\boldsymbol{U}}(f) \stackrel{\text{d}}{=} \mathcal{W}^C_{2p}(K, \boldsymbol{S}_{\boldsymbol{U}}(f)), \tag{31}$$

i.e., $\boldsymbol{A}(f)$ is distributed as a $2p$-dimensional complex Wishart distribution with K complex degrees of freedom and mean $K\boldsymbol{S}_{\boldsymbol{U}}(f)$. Given the form of $\hat{\boldsymbol{S}}_{\boldsymbol{U}}(f)$, we partition $\boldsymbol{A}(f)$ analogously in terms of sub-matrices as

$$\boldsymbol{A}(f) = \begin{bmatrix} \boldsymbol{A}_{11}(f) & \boldsymbol{A}_{12}(f) \\ \boldsymbol{A}_{21}(f) & \boldsymbol{A}_{22}(f) \end{bmatrix}. \tag{32}$$

Then the GLR test statistic in (27) can be expressed as

$$T(f) = L_G^{1/K}(f) = \frac{\det\{\boldsymbol{A}(f)\}}{\det\{\boldsymbol{A}_{11}(f)\}\det\{\boldsymbol{A}_{22}(f)\}}. \tag{33}$$

Lemma 2: The rth moment of $L_G(f)$, namely $E\{L_G^r(f)\}$, is given by

$$\frac{\prod_{j=1}^p \Gamma(K-j+1)}{\prod_{j=1}^p \Gamma(K-j-p+1)} \frac{\prod_{j=1}^p \Gamma(K[1+r]-j-p+1)}{\prod_{j=1}^p \Gamma(K[1+r]-j+1)}. \tag{34}$$

Proof: This is given in Appendix A. ■

A random variable $0 \le W \le 1$ is said to be of Box-type [4, eqn. (70)] if for all $r \in \mathbb{N}$,

$$E\{W^r\} = C_0 \left[\frac{\prod_{j=1}^l b_j^{b_j}}{\prod_{i=1}^m a_i^{a_i}}\right]^r \frac{\prod_{i=1}^m \Gamma(a_i[1+r]+\vartheta_i)}{\prod_{j=1}^l \Gamma(b_j[1+r]+\zeta_j)}, \tag{35}$$

where $\sum_{i=1}^m a_i = \sum_{j=1}^l b_j$, and the constant term C_0 is

$$C_0 = \frac{\prod_{j=1}^l \Gamma(b_j+\zeta_j)}{\prod_{i=1}^m \Gamma(a_i+\vartheta_i)},$$

so that its zero'th moment is unity.

We see that $L_G(f)$ is a random variable of Box-type with

$$m = l = p;\ a_i = K;\ b_j = K;\ \vartheta_i = 1-i-p,\ \zeta_j = 1-j,$$

and C_0 is

$$C_0 = \prod_{j=1}^p \frac{\Gamma(K-j+1)}{\Gamma(K-j-p+1)}.$$

C. Cumulants

The moment generating function for $M(f) = -2\log L_G(f)$ is given by (with f suppressed), $\phi_M(s) = E\{\mathrm{e}^{sM}\} = E\{L_G^{-2s}\}$ so using (34),

$$\phi_M(s) = C_0 \prod_{j=1}^{p} \frac{\Gamma(K[1-2s]-j-p+1)}{\Gamma(K[1-2s]-j+1)}.$$

The Gamma functions will be valid if $-2Ks + K - j - p + 1 > 0$ for all $j = 1, \ldots, p$, which requires $-2s > (2p-1-K)/K$.

The cumulants κ_i of M can be easily obtained from the cumulant generating function by successively differentiating $\log\phi_M(s)$ and setting $s = 0$. Notice that the requirement $-2s > (2p-1-K)/K$ corresponds to $K \geq 2p$ when $s = 0$. Then, for $i \geq 1$,

$$\kappa_i = \left.\frac{\mathrm{d}^i \log\phi_M(s)}{(\mathrm{d}s)^i}\right|_{s=0}$$

so that κ_i is

$$[-2K]^i \sum_{j=1}^{p}\Big[\psi^{(i-1)}(K-j-p+1) - \psi^{(i-1)}(K-j+1)\Big]. \tag{36}$$

Here for $i = 1$, $\psi(x) = [\mathrm{d}\log\Gamma(x)]/\mathrm{d}x$ is the digamma function, while for $i = 2$ and 3, $\psi^{(1)}(x)$ and $\psi^{(2)}(x)$ are the trigamma and tetragamma functions respectively; these are all 'polygamma functions.' κ_1 is the mean, κ_2 is the variance, $\kappa_3/\kappa_2^{3/2}$ is the skewness and κ_4/κ_2^2 is the excess kurtosis.

D. Scaled chi-square approximation

Box [4] provides a scaled chi-squared approximation for M of the form $M(f) \stackrel{\mathrm{d}}{=} c_B\chi_d^2$. The constant c_B is chosen so that the cumulants of $c_B\chi_d^2$ match those of $M(f)$ up to an error of order $O(K^{-2})$. The degrees of freedom d associated with the chi-square approximation for $M(f)$ is given by Box [4]

$$\begin{aligned} d &= -2\left[\sum_{i=1}^{p}\vartheta_i - \sum_{j=1}^{p}\zeta_j\right] \\ &= -2\left[\sum_{i=1}^{p}(1-i-p) - \sum_{j=1}^{p}(1-j)\right] \\ &= -2\left[-\sum_{i=1}^{p} i - \sum_{i=1}^{p} p + \sum_{j=1}^{p} j\right] = 2p^2 = \nu, \end{aligned}$$

as expected. The scaling factor c_B is a constant determined as follows [4, p. 338]. Define

$$\omega_n = \frac{(-1)^{n+1}}{n(n+1)}\left[\sum_{i=1}^{p}\frac{B_{n+1}(\vartheta_i)}{a_i^n} - \sum_{j=1}^{p}\frac{B_{n+1}(\zeta_j)}{b_j^n}\right] \tag{37}$$

where $B_n(x)$ is the Bernoulli polynomial of degree n and order unity, with

$$B_2(x) = x^2 - x + \frac{1}{6}; \quad B_3(x) = x^3 - \frac{3}{2}x^2 + \frac{1}{2}x.$$

Subsequently, let $W_1 = 2\omega_1/d$ and $W_2 = 4\omega_2/d$, then c_B is chosen according to the following rule:

$$c_B = \begin{cases} (1 - W_1)^{-1} & \text{if } W_2 \geq W_1^2 \\ 1 + W_1 & \text{otherwise.} \end{cases}$$

Using (37) we find that

$$W_1 = \frac{p}{K}; \quad W_2 = \frac{(7p^2 - 1)}{6K^2}.$$

It is straightforward to see that $W_2 \geq W_1^2$ for all (K, p) combinations, implying that $c_B = K/(K - p)$, giving Box's finite sample approximation as

$$M(f) \stackrel{\mathrm{d}}{=} \frac{K}{K - p}\chi^2_{2p^2}. \tag{38}$$

(This agrees with (30) asymptotically as $K \to \infty$ for a fixed dimension p.)

$T(f)$ in (29) is $\det(\boldsymbol{I}_p - \hat{\boldsymbol{S}}_{\boldsymbol{Z}}^{-1}(f)\hat{\boldsymbol{R}}_{\boldsymbol{Z}}(f)\hat{\boldsymbol{S}}_{\boldsymbol{Z}}^{-T}(-f)\hat{\boldsymbol{R}}_{\boldsymbol{Z}}^{H}(f))$, so for the case $p = 1$,

$$T(f) = 1 - \frac{|\hat{R}_Z(f)|^2}{\hat{S}_Z(f)\hat{S}_Z(-f)} = 1 - \hat{\gamma}_*^2(f)$$

where $\hat{\gamma}_*^2(f)$ is the 'conjugate coherence,' i.e., the ordinary coherence between $\{Z_t\}$ and $\{Z_t^*\}$ (e.g., [8]). Then $M(f) = -2K\log(1 - \hat{\gamma}_*^2(f))$. Under the null hypothesis it is known that

$$\hat{\gamma}_*^2(f) \stackrel{\mathrm{d}}{=} \mathrm{beta}(1, K - 1), \tag{39}$$

i.e., coherence has the $\mathrm{beta}(1, K - 1)$ distribution. It then follows readily that $M(f)$ has PDF

$$f_M(x) = \frac{K - 1}{2K}\mathrm{e}^{-x\left[\frac{K-1}{2K}\right]},$$

so that $M(f) \stackrel{\mathrm{d}}{=} \frac{K}{K-1}\chi_2^2$ and Box's approximation (38) is in fact *exact* for the case $p = 1$. When $p = 1$ we note that $W_2 = W_1^2$.

Remark 5: For small values of K, matching cumulants of $M(f)$ up to an error of order $O(K^{-2})$ could be problematic for $p > 1$ [4, p. 329]. This leads us to consider other approaches.

VII. Other Statistical Approaches

Using our results in Section VI-C, our aim here is to develop a cumulant matching approach which results in a scaled F approximation for $M(f)$. We contrast the simplicity of this approach with other existing methods and illustrate its accuracy.

A. Product of Independent Beta Random Variables

Lemma 3: Under the null hypothesis the distribution of $T(f)$ can be expressed as a product of independent beta random variables:

$$T(f) \stackrel{\mathrm{d}}{=} \prod_{j=1}^{p} B_j, \tag{40}$$

where $B_j \stackrel{\mathrm{d}}{=} \mathrm{beta}(K+1-j-p, p)$, independently.

Proof: This is given in Appendix B. ■

Remark 6: If $p = 1$, (40) gives $T(f) \stackrel{\mathrm{d}}{=} \mathrm{beta}(K-1, 1)$, as it should since $T(f) = 1 - \hat{\gamma}_*^2(f)$, and (39) holds.

In a different context Gupta [16] developed the distribution of the product of p independent beta distributions: a likelihood ratio criterion for testing a hypothesis about regression coefficients in a multivariate normal setting takes the form $\Lambda = \det\{\boldsymbol{V}_1\}/\det\{\boldsymbol{V}_1 + \boldsymbol{V}_2\}$ under the corresponding null hypothesis, with $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ independently distributed as

$$\boldsymbol{V}_1 \stackrel{\mathrm{d}}{=} \mathcal{W}_p^C(f_1, \boldsymbol{\Sigma}),\ \boldsymbol{V}_2 \stackrel{\mathrm{d}}{=} \mathcal{W}_p^C(f_2, \boldsymbol{\Sigma}),$$

for integer parameters f_1, f_2 and covariance matrix $\boldsymbol{\Sigma}$. Then Λ has the three-parameter complex U distribution $U(p, f_2, f_1)$ which is distributed as a product of p beta variables with $B_j \stackrel{\mathrm{d}}{=} \mathrm{beta}(f_1 - j + 1, f_2)$. So setting Gupta's parameters f_1 and f_2 to $K-p$ and p, respectively, shows that $T(f)$ has the three-parameter complex U distribution $U(p, p, K-p)$. This helps only a little because there are no simple expressions for this distribution's PDF or quantiles etc. However, by using convolution techniques Gupta did obtain some exact results for the case $p = 2$. In fact it turns out that for $p = 2$ the right-side of (38) can be improved to

$$\frac{K}{K-2} G(1-\alpha) \chi_8^2(1-\alpha) \tag{41}$$

where $G(1-\alpha)$ is an exact (tabulated) correction factor and $\chi_8^2(1-\alpha)$ is the $100(1-\alpha)\%$ point of the chi-square distribution with 8 degrees of freedom. For example for $p = 2, K = 6$ and $\alpha = (0.05, 0.01)$ the factors are $(1.043, 1.051)$ [16, Table 1]. The work of Gupta was extended as part of [20, p. 5] who produced tables of approximate correction factors for the right-side of (38) for $p \geq 3$ so that $M(f)$ is compared to

$$\frac{K}{K-p} G(1-\alpha) \chi_{2p^2}^2(1-\alpha). \tag{42}$$

Setting their parameters n and q to $K-p$ and p respectively, shows that for example for $p = 3, K = 8$ and $\alpha = (0.05, 0.01)$ the factors are $(1.076, 1.087)$ [20, Table 7]. The effect of these correction factors will be discussed shortly.

Remark 7: The result (40) is very nice, and quantiles of $T(f)$ could be found through, say, successive convolution techniques, but this is very complicated — see [6], [17] who develop this approach for a related statistic.

B. Matching the first three cumulants exactly

The look-up tables of [16] and [20] are not convenient and so we now develop a simple and fast method for approximating the percentage points of the distribution of $M(f)$. Box [4] considered using the very flexible Pearson system for approximating the distribution of likelihood ratios. Box [4, p. 330] introduced a discriminant $D = (\kappa_1\kappa_3)/(2\kappa_2^2)$, such that if $D > 1$ a Pearson type VI should be fitted; this corresponds to $W_2 > W_1^2$. For $p = 2 : 20, K = 1 : 100$, with $K \geq 2p$ we always found $D > 1$ using (36). (Note $p = 1$ is excluded since $W_2 = W_1^2$ in that case.)

Box [4] considered distributions of the form bF_{ν_1,ν_2}, i.e., a scaled F distribution (Pearson type VI) with parameters ν_1, ν_2, and suggested matching cumulants *approximately*.

We have chosen to match the first three cumulants of the form (36) *exactly*; the parameters of bF_{ν_1,ν_2} are related to the cumulants via [14]

$$\begin{aligned} b &= \frac{2\kappa_1\left(\kappa_1^2\kappa_2 - \kappa_2^2 + \kappa_1\kappa_3\right)}{2\kappa_1^2\kappa_2 - 4\kappa_2^2 + 3\kappa_1\kappa_3}, \\ \nu_1 &= \frac{4\kappa_1\left(\kappa_1^2\kappa_2 - \kappa_2^2 + \kappa_1\kappa_3\right)}{4\kappa_1\kappa_2^2 - \kappa_1^2\kappa_3 + \kappa_2\kappa_3}, \\ \nu_2 &= \frac{4\kappa_1^2\kappa_2 - 8\kappa_2^2 + 6\kappa_1\kappa_3}{\kappa_1\kappa_3 - 2\kappa_2^2}. \end{aligned} \tag{43}$$

Then to carry out the test $M(f)$ would be compared to

$$bF_{\nu_1,\nu_2}(1-\alpha), \tag{44}$$

where $F_{\nu_1,\nu_2}(1-\alpha)$ is the $100(1-\alpha)\%$ point of the F distribution with parameters b, ν_1, ν_2 given by (43).

C. Comparison of Approximations

For some combinations of (p, K) the asymptotic result (30) is compared to Box's basic approximation (38), the adjusted Box method (41), (42) and the scaled F method (44) in Table I which gives the 95% and 99% points of the distribution of $M(f)$ according to the four approaches. There is very good agreement between the adjusted Box method and the scaled F method, the latter being quick and simple to compute. Box's basic approximation is a massive improvement on the asymptotic result. For $p = 2$ the adjusted Box approximation due to [16] is exact and we see that the scaled F approximation is therefore

(p, K)	Method	$\alpha = 0.05$	$\alpha = 0.01$
$(2, 6)$	Asymptotic	15.51	20.09
	Box	23.26	30.14
	Adjusted Box	24.26	31.67
	scaled F	24.26	31.68
$(3, 8)$	Asymptotic	28.87	34.81
	Box	46.19	55.69
	Adjusted Box	49.70	60.53
	scaled F	49.71	60.54
$(4, 10)$	Asymptotic	46.19	53.49
	Box	76.99	89.14
	Adjusted Box	84.84	99.31
	scaled F	84.85	99.30
$(5, 12)$	Asymptotic	67.50	76.15
	Box	115.72	130.55
	Adjusted Box	129.96	148.17
	scaled F	129.94	148.18

TABLE I

COMPARISON OF PERCENTAGE POINTS OF $M(f)$ ACCORDING TO THE ASYMPTOTIC RESULT (30), BOX'S APPROXIMATION (38), ADJUSTED BOX METHOD (41), (42) AND THE SCALED F METHOD (44).

very accurate. Other combinations of p and small K lead to similar results. The agreement of the scaled F approximation with the previous historically tabulated results (adjusted Box approximation) leads us to the following recommendation.

D. Recommended testing approach

In view of the discusssions and results above, the following is recommended for a given choice of α :

- If $p = 1$, reject H_{0f} in (23) if

$$M(f) > \frac{K}{K-1}\chi_2^2(1-\alpha). \tag{45}$$

  This test is distributionally exact.

- If $p \geq 2$, reject H_{0f} in (23) if

$$M(f) > bF_{\nu_1,\nu_2}(1-\alpha). \tag{46}$$

  The accuracy of the scaled F approximation for our time series test (23) is now confirmed by simulation.

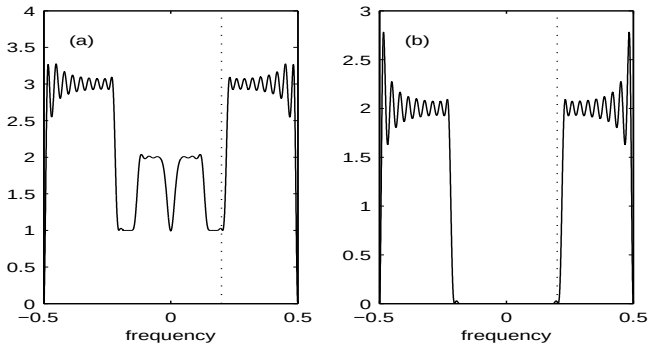

Fig. 1. (a) $S_Z(f)$ and (b) $R_Z(f)$. The vertical dotted line marks the frequency $f = 0.2$.

VIII. Simulation Results

For $p \geq 2$ we will show that using the scaled F approximation test where we reject H_{0f} in (23) if (46) holds brings about a worthwhile accuracy improvement over Box's approximation test where we reject H_{0f} if

$$M(f) > \frac{K}{K-p}\chi^2_{2p^2}(1-\alpha). \tag{47}$$

To be able to do this we need to simulate from a model such that $\boldsymbol{S_U}(f)$ in (7) has $\boldsymbol{R_Z}(f) = \mathbf{0}$ for some frequency range. We can proceed as follows.

We know [30] that any complex second-order stationary scalar process (assumed zero mean here), whether proper or improper, can be written as the output of a widely linear filter driven by proper white noise, i.e.,

$$Z_t = \sum_{l=-\infty}^{\infty} g_l \epsilon_{t-l} + \sum_{l=-\infty}^{\infty} h_l \epsilon^*_{t-l}, \tag{48}$$

where $\{g_l\}$ and $\{h_l\}$ are sequence of complex constants, and $\{\epsilon_t\}$ is proper white noise for which $\text{cov}\{\epsilon_{t+\tau}, \epsilon_t\} = \sigma^2_\epsilon \delta_{\tau,0}$ and $\text{cov}\{\epsilon_{t+\tau}, \epsilon^*_t\} = 0$, for $\tau \in \mathbb{Z}$, where $\delta_{j,k}$ is the Kronecker delta. For simulation purposes it is convenient to set $\sigma^2_\epsilon = 1$. Then [30]

$$S_Z(f) = |G(f)|^2 + |H(f)|^2 \tag{49}$$

$$R_Z(f) = G(f)H(-f) + G(-f)H(f), \tag{50}$$

where $G(f)$ is the frequency response function of $\{g_l\}$ given by $G(f) = \sum_{l=-\infty}^{\infty} g_l \mathrm{e}^{-\mathrm{i}2\pi f l}$ and $H(f)$ is the frequency response function of $\{h_l\}$.

(p,K)	$100\alpha\%$	f 0.06	0.12	0.18	0.24	0.42
$(2,6)$	1%	1.5	1.5	1.4	31.1	35.4
		1.1	1.1	0.9	25.8	30.0
	5%	6.1	6.2	6.3	60.0	63.6
		5.0	5.1	5.2	55.3	59.3
$(3,8)$	1%	2.0	2.1	2.2	55.7	58.9
		0.9	1.1	1.1	42.3	45.5
	5%	8.2	8.3	8.3	81.0	82.9
		4.9	5.1	5.2	72.6	75.2

TABLE II

REJECTION PERCENTAGES OVER 10 000 REPETITIONS. THE TOP LINE OF EACH ENTRY IS FOR BOX'S χ^2 APPROXIMATION (38) AND THE LOWER LINE IS FOR THE F APPROXIMATION OF (46).

For $p \geq 2$ we generate processes $\{Z_{j,t}\}, j = 1, \ldots, p$, such that

$$Z_{j,t} = \sum_{l=-\infty}^{\infty} g_l \epsilon_{j,t-l} + \sum_{l=-\infty}^{\infty} h_l \epsilon^*_{j,t-l} + \sum_{l=-\infty}^{\infty} a_l \bar{\epsilon}_{j,t-l} + \sum_{l=-\infty}^{\infty} a_l \bar{\epsilon}^*_{j,t-l}, \tag{51}$$

where the $2p$ processes $\{\{\epsilon_{j,t}\}, \{\bar{\epsilon}_{j,t}\}, j = 1, \ldots, p\}$ are all independent of each other. The filter $\{g_l\}$ was chosen to be low-pass with a frequency transition zone $[0.125, 0.15]$. The filter $\{h_l\}$ was of 'Hilbert-type' or all-pass in the frequency zone $[0.05, 0.45]$. Thus $G(f)$ is real and symmetric while $H(f)$ is imaginary and skew-symmetric. According to (50), if using just these two filters, the resulting $R_Z(f)$ is zero for $f \in [-0.5, 0.5]$. However, the filter $\{a_l\}$ was chosen to be high-pass above $f = 0.2$ and therefore generates non-zero $R_Z(f)$ values at these high frequencies. The resulting $S_Z(f)$ and $R_Z(f)$ are shown in Fig. 1.

The matrix $\boldsymbol{S_Z}(f)$ is thus of the form $\boldsymbol{S_Z}(f) = S_Z(f)\boldsymbol{I}_p$ with frequency dependence as shown in Fig. 1(a) while $\boldsymbol{R_Z}(f)$ is of the form $\boldsymbol{R_Z}(f) = R_Z(f)\boldsymbol{I}_p$ with frequency dependence as shown in Fig. 1(b). We can thus simulate from this model to evaluate our hypothesis tests, knowing that for frequencies where $\boldsymbol{R_Z}(f) = \boldsymbol{0}$ in fact $\boldsymbol{S_Z}(f) \neq \boldsymbol{0}$ and thus (27) is well-defined.

Sample results are shown in Table II for $(p, K) = (2, 6)$ and $(3, 8)$. So here $K = 6$ and 8 are indeed small. Here $N = 512$ but smaller time series lengths such as 128 produced very similar results. Shown are rejection percentages for H_{0f} over 10 000 independent repetitions. The nominal rates are shown in the second column. The first three columns of rejection percentages are for frequencies where $\boldsymbol{R_Z}(f) = \boldsymbol{0}$,

(H_{0f} in (23) is true) and the latter two are for frequencies where $\boldsymbol{R}_{\boldsymbol{Z}}(f) \neq \mathbf{0}$ (H_{0f} is false) — see Fig. 1(b). The top line of each entry is for Box's χ^2 approximation (47) and the lower line is for the F approximation of (46). We see that, proportionately, the latter has a much more accurate rejection rate than Box's approximation when H_{0f} is true, but is slightly less accurate when H_{0f} is false.

IX. Overall Test

A. Background

Plotting $M(f)$ against f and identifying frequencies where the critical value is exceeded is potentially quite informative. Of course, by the definition of propriety, for an *overall* test we need to test H_0 in (2). Here the test domain is finite, while for the alternative $\boldsymbol{r}_{\boldsymbol{Z},\tau} = \mathbf{0}$ for all $\tau \in \mathbb{Z}$, the test domain is infinite. However, $f \in [0, f_{\mathcal{N}}]$ is still a continuum.

The approach taken in [36] is to construct a single test statistic from GLRTs conducted at frequencies where the spectral estimators are approximately independent, (spaced apart by the frequency bandwidth of the spectral estimator). However if, say, $\boldsymbol{R}_{\boldsymbol{Z}}(f) \neq \mathbf{0}$ for a narrow band of frequencies midway between two of the testing frequencies, i.e., on the edge of the bandwidths for two of the independent statistics, the test would be expected to be quite problematic.

Rather than requiring independent statistics we prefer the flexibility of a multiple hypothesis testing approach. As a proxy for the formal overall test we consider testing the set of L null hypotheses

$$H_l : \boldsymbol{R}_{\boldsymbol{Z}}(f_l) = \mathbf{0} \text{ for } f_1, \ldots, f_L, \tag{52}$$

where $\{f_l\}$ is a set of frequencies *densely* sampling the positive Nyquist range.

B. Controlling the FWER

The familywise error rate (FWER) is defined as $\Pr(V \geq 1)$ where V is the number of false rejections. Following [18], let $P_{(1)} < \cdots < P_{(L)}$ denote the ordered p-values corresponding to the L tests defined by (52). Let $H_{(1)}, \ldots, H_{(L)}$ be the associated null hypotheses. Let J be the minimal index such that $P_{(j)} > \alpha/[L + 1 - j]$. Reject only the null hypotheses $H_{(1)}, \ldots, H_{(J-1)}$. If $J = 1$ then do not reject any hypotheses; if no such J exists, reject all hypotheses. Then $\mathrm{FWER} \leq \alpha$. The tests need not be independent.

C. Controlling the FDR

Controlling the FWER is equivalent to making it unlikely that even one false rejection is made. A less stringent approach is to control the false discovery rate (FDR); see [2], [3]. Let FDP denote the

$\alpha = 0.05$	frequency interval		
	0.005	0.01	0.02
FWER	4.7	5.0	4.8
FDRi	4.8	5.2	4.9
FDRd	1.0	1.2	1.2

TABLE III

RATES ACHIEVED (PERCENTAGES) OVER 10 000 REPETITIONS.

proportion of rejections that are incorrect. Then $\mathrm{FDR} \stackrel{\mathrm{def}}{=} E\{\mathrm{FDP}\}$. As before let $P_{(1)} < \cdots < P_{(L)}$ denote the ordered p-values. Define $l_i = i\alpha/[C_L L]$ and $R = \max\{i : P_{(i)} < l_i\}$. Here $C_L = 1$ for independent tests and $C_L = \sum_{i=1}^{L} 1/i$, for dependent tests. Let $\ell = P_{(R)}$, and

$$\text{reject all null hypotheses } H_i \text{ for which } P_i \leq \ell, \tag{53}$$

then $\mathrm{FDR} \leq \alpha$. The independent tests approach is also valid for some forms of positive correlation [3]. The non-unity adjustment C_L for dependent tests is general: the level of dependency does not matter; as a result, this procedure is rather conservative. Nevertheless [3], the dependent tests approach can still prove much more powerful than the comparable FWER. Note that C_L changes as more/less tests are made (L increases/decreases) or the sampling rate of the frequency range is increased/decreased.

D. Simulation of overall test for true null

Here we will look at an example of when the null hypothesis (2) is true. We use the simulation set-up of Section VIII for $p = 2$ but setting the coefficients $\{a_l\}$ to zero in (51). Hence $\boldsymbol{R_Z}(f) = \mathbf{0}$ for $f \in [-0.5, 0.5]$. With $N = 512$ and $K = 6$ the bandwidth of the spectral window (16) is about 0.014. Over the range $f \in [0.02, 0.48]$ we consider three frequency samplings: steps of 0.005, 0.01 and 0.02; the first two are within the spectral window bandwidth and the latter outside, so we would expect the tests to be dependent for the first two cases, and independent for the third. We carry out $10\,000$ independent repetitions with $\alpha = 0.05$ and from these calculate (a) the FWER, and (b) the FDR under assumptions of (i) independence (FDRi) and (ii) dependence (FDRd). p-values are calculated via the scaled F approximation. The results are given in Table III. We see that $\mathrm{FWER} \leq \alpha$, as required. (The test does not require independence.) For FDRi there is variation about 5% in line with the assumed independence being false for frequency intervals 0.01 and 0.005. FDRd can be seen to be always quite conservative.

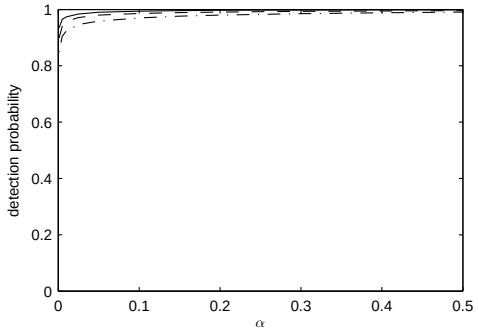

Fig. 2. The ROC curve for the model of Section IX-E. On the x-axis is α and we are using FDRd. The y-axis gives the probability of detection of impropriety. The curves are for $\sigma_\lambda^2 = 0.4$ (solid), 0.5 (dashed) and 0.6 (dash-dot).

E. Simulation of overall test for false null (ROC curve)

We now consider behaviour when the null hypothesis (2) is false. To illustrate the receiver operating characteristic (ROC) we use a signal structure used in [27] and [33]. Consider the scalar process $Z_t = X_t \mathrm{e}^{\mathrm{i}\lambda_t} + \xi_t$ where $\{X_t\}$ is a real-valued stationary zero mean Gaussian process with autocovariance $s_{X,\tau} = \mathrm{e}^{-\tau^2/5}$, λ_t is a random sequence drawn from a Gaussian distribution with mean zero and variance σ_λ^2, and $\{\xi_t\}$ is proper complex-valued Gaussian noise with variance unity. All three random sequences are independent. So the signal-to-noise ratio is unity and the signal is improper [27], [33] while the noise is proper, as in [36]. A $p = 2$ vector-valued process was created from two independent copies of $\{Z_t\}$. Fig. 2 shows the probability of detection of impropriety (constructed from 10 000 repetitions) as a function of α using FDRd. With $N = 1000$ and $K = 12$ the bandwidth of the spectral window (16) is about 0.013, and we sampled the frequencies at an interval of 0.005. The curves are for $\sigma_\lambda^2 = 0.4, 0.5$ and 0.6. As σ_λ^2 increases, $|R_Z(f)|$ decreases towards zero, and impropriety is increasingly hard to detect.

X. Data Analysis

A. Background

We apply our results to ocean current speed and direction time series recorded at a mooring in the Labrador Sea [8], [21], [22]. We associate the eastward (zonal) measurement of current speed with $\{X_t\}$ and the northward (meridional) measurement with $\{Y_t\}$ and thus obtain the complex-valued series from (3). We consider series recorded at the three depths 110, 760 and 1260m. The series are labelled 1 to 3 with increasing depth, giving $\boldsymbol{Z}_t = [Z_{1,t}, Z_{2,t}, Z_{3,t}]^T$. We used $N = 1600$ observations for the 3-vector-valued complex time series, with a sampling interval of $\Delta_\mathrm{t} = 1$hr. In the spectral analysis $K = 12$ sine tapers were applied. Since W_N in (16) is 0.004c/hr, the validity range $W_N \le |f| \le f_\mathcal{N} - W_N$ for our

statistical results for a finite-N sample is given by $0.004 \leq |f| \leq 0.496$c/hr. For this data there was no evidence to reject stationarity [7, p. 49], [8] or Gaussianity [9].

Of great interest to oceanographers are deep ocean motions well away from boundaries, especially in the internal wave frequency band. We pay special attention to low frequencies $f \in [0.02, 0.14]$c/hr within which interesting rotational effects have been observed in previous studies [8], [38]. The very dominant semi-diurnal tide at around $f = 0.08$c/hr was estimated and removed to avoid spectral leakage.

By way of example Fig. 3 shows the scalar quantities $S_Z(f)$ and $R_Z(f)$ for one of the series. We see that $S_Z(f)$ is asymmetric about $f = 0$ as expected. For frequencies around $f = 0.04$c/hr, we see that $|\mathrm{Im}\{R_Z(f)\}|$ is noticeably larger than $|\mathrm{Re}\{R_Z(f)\}|$; these frequencies are of interest later (see Fig. 4).

B. Testing for Propriety

Fig. 4 shows propriety-testing results for our region of interest $f \in [0.02, 0.14]$c/hr. The upper plot shows the test statistic $M(f)$ and the critical value for $\alpha = 0.05$ for the frequency-specific propriety test for the vector $\boldsymbol{Z}_t = [Z_{1,t}, Z_{2,t}, Z_{3,t}]^T$. On a frequency-by-frequency basis rejection occurs at all frequencies where $M(f)$ exceeds the critical value; these frequencies are shown by heavy dots. The lower plot shows the results from the overall FDRd approach. Frequencies causing rejection according to (53) are also marked by heavy dots. As might be expected, the frequencies leading to the overall rejection of propriety — those around $f = 0.04$c/hr — are a subset of those causing rejection on a frequency-by-frequency basis.

We also implemented the testing approach of [36] using — in order to get independent statistics — multitaper estimates spaced apart by the bandwidth, 0.008c/hr, of the spectral window. Again, with $\alpha = 0.05$, overall propriety was rejected, as would be expected given that our test identified the band of frequencies around 0.04c/hr as responsible for rejection, and this would be easily 'seen' with a frequency spacing of 0.008c/hr. Note, the information conveyed in Fig. 4 is clearly useful in specifying the frequencies causing rejection, whereas the approach in [36] provides only a decision on propriety. In situations where isolated narrow bands of frequencies might cause rejection the approach in [36] could be problematic, as explained in Section IX-A.

XI. Summary and Conclusion

We have developed a frequency domain approach to test for propriety of complex-valued vector time series. For the vector case ($p \geq 2$) we have justified use of the rule that the frequency-specific hypothesis H_{0f} in (23) is rejected if $M(f) = -2K \log T(f) > bF_{\nu_1,\nu_2}(1-\alpha)$. There is no assumption that K is large, and indeed this would rarely be expected in practice. We have shown in detail how the statistic $T(f)$ arises

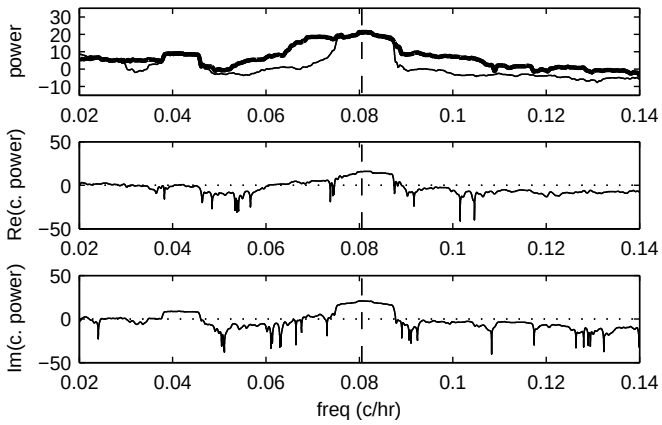

Fig. 3. Upper plot: the spectrum $S_Z(f)$, in dB, for one of the three current series. Thin line: the counterclockwise (positive) frequencies; thick line: clockwise (negative) frequencies. Middle plot: $|\mathrm{Re}\{R_Z(f)\}|$ on $10\log_{10}$ scale. Lower plot: $|\mathrm{Im}\{R_Z(f)\}|$ on $10\log_{10}$ scale. The dashed line marks the semi-diurnal tidal frequency.

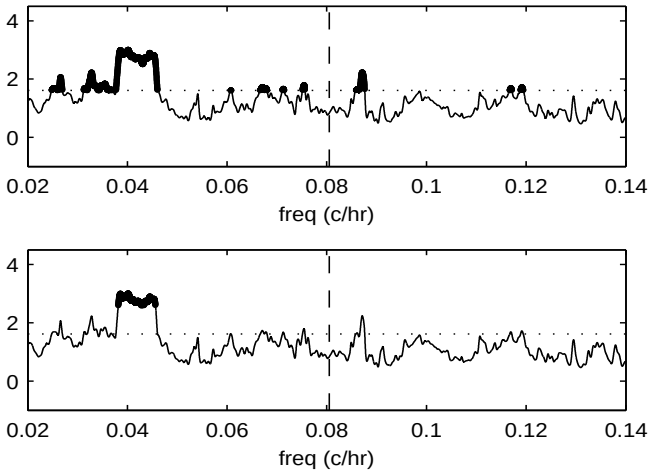

Fig. 4. The test statistic $M(f)$ (solid) and the critical value (dotted line) for $\alpha = 0.05$ for the frequency-specific propriety test for the vector $\boldsymbol{Z}_t = [Z_{1,t}, Z_{2,t}, Z_{3,t}]^T$. Upper: test at individual frequencies; rejection occurs at all frequencies where $M(f)$ exceeds the critical value, such frequencies being marked by heavy dots. Lower: overall test results using FDRd with $\alpha = 0.05$. Frequencies causing rejection are again marked by heavy dots. The dashed line marks the semi-diurnal tidal frequency.

by consideration of canonical coherencies for complex-valued vector time series. The frequency-specific tests can be combined using multiple hypothesis testing to give an overall test of H_0 in (2).

When propriety is invalid, the frequency domain approach has the scientific advantage of showing which frequency bands are causing rejection, possibly allowing linkage to known or hypothesized properties of the physical processes involved.

APPENDIX

A. Proof of Lemma 2

To simplify notation we drop explicit frequency dependence. Consider the distribution of $T = L_G^{1/K}$, given by (33), *under the null hypothesis*. From [19, eqn. (2.6)]

$$E\{T^r\} = \frac{\prod_{j=1}^{2p}\Gamma(K+r+1-j)[\prod_{j=1}^{p}\Gamma(K+1-j)]^2}{\prod_{j=1}^{2p}\Gamma(K+1-j)[\prod_{j=1}^{p}\Gamma(K+r+1-j)]^2}. \tag{54}$$

Now $T^r = L_G^{r/K}$ so if we let $r \to rK$, then $T^{rK} = L_G^r$. So

$$\begin{aligned}
E\{L_G^r\} &= \frac{\prod_{j=1}^{2p}\Gamma(K[1+r]+1-j)\prod_{j=1}^{p}\Gamma(K+1-j)}{\prod_{j=1}^{2p}\Gamma(K+1-j)\prod_{j=1}^{p}\Gamma(K[1+r]+1-j)} \\
&\times \left[\frac{\prod_{j=1}^{p}\Gamma(K+1-j)}{\prod_{j=1}^{p}\Gamma(K[1+r]+1-j)}\right] \\
&= \frac{\prod_{j=1}^{p}\Gamma(K[1+r]+1-j-p)}{\prod_{j=1}^{p}\Gamma(K+1-j-p)} \\
&\times \left[\frac{\prod_{j=1}^{p}\Gamma(K+1-j)}{\prod_{j=1}^{p}\Gamma(K[1+r]+1-j)}\right] \\
&= \left[\frac{\prod_{j=1}^{p}\Gamma(K+1-j)}{\prod_{j=1}^{p}\Gamma(K+1-j-p)}\right] \\
&\times \frac{\prod_{j=1}^{p}\Gamma(K[1+r]+1-j-p)}{\prod_{j=1}^{p}\Gamma(K[1+r]+1-j)},
\end{aligned}$$

which is (34).

B. Proof of Lemma 3

Under the null hypothesis the rth moment of $T(f)$ is

$$E\{T^r(f)\} = \prod_{j=1}^{p}\frac{\Gamma(K+r+1-j-p)\Gamma(K+1-j)}{\Gamma(K+r+1-j)\Gamma(K+1-j-p)}. \tag{55}$$

To see this start with (54) and proceeed in analogous vein to the proof of Lemma 2; since we are continuing to look at $E\{T^r\}$ the step $r \to rK$ is not made. Note that when $j = p$ the critical gamma function argument is still positive: $K+r+1-j-p = K+r+1-2p > 0$ since $K \geq 2p$ with $r \geq 0$.

A real scalar random variable X is said to have a (type-1) beta distribution, $X \stackrel{\mathrm{d}}{=} \mathrm{beta}(\alpha,\beta)$, if the PDF is

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1},\ 0<x<1, \alpha>0, \beta>0.$$

The rth moment for this distribution is

$$E\{X^r\} = \frac{\Gamma(\alpha+r)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+r)\Gamma(\alpha)},\ \alpha+r>0. \tag{56}$$

Comparing (55) and (56) we see for a fixed j that $\alpha = K+1-j-p$ and $\beta = p$ which gives the required result.

ACKNOWLEDGEMENT

The authors are grateful to Jon Lilly for the Labrador Sea data and to the three referees for their very helpful comments leading to a much improved paper.

REFERENCES

[1] T. Adalı, P. J. Schreier & L. L. Scharf, "Complex-valued signal processing: the proper way to deal with impropriety," *IEEE Transactions on Signal Processing*, vol. 59, pp. 5101–5125, 2011.

[2] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," J. Roy. Stat. Soc. B, vol. 57, pp. 289-300, 1995.

[3] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Annals of Statistics*, vol. 29, pp. 1165-1188, 2001.

[4] G. E. P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, 317–46, 1949.

[5] D. R. Brillinger, *Time Series: Data Analysis and Theory (Expanded Edition)*. New York: McGraw-Hill Inc., 1981.

[6] E. M. Carter, C. G. Khatri & M. S. Srivastava, "Nonnull distribution of likelihood ratio criterion for reality of covariance matrix," *J. Multivariate Analysis*, vol. 6, pp. 176–184, 1976.

[7] S. Chandna, *Frequency domain analysis and simulation of multi-channel complex-valued time series.* PhD thesis, Imperial College London, 2014.

[8] S. Chandna and A. T. Walden, "Statistical properties of the estimator of the rotary coefficient," *IEEE Transactions on Signal Processing*, vol. 59, pp. 1298–1303, 2011.

[9] S. Chandna and A. T. Walden, "Simulation methodology for inference on physical parameters of complex vector-valued signals," *IEEE Transactions on Signal Processing*, vol. 61, pp. 5260–5269, 2013.

[10] T. Chonavel, *Statistical Signal Processing*. London UK: Springer-Verlag, 2002.

[11] S. Elipot and R. Lumpkin, "Spectral description of oceanic near-surface variability," *Geophys. Res. Lett.* **35**, L05606, 2008.

[12] W. J. Emery and R. E. Thomson, *Data Analysis Methods in Physical Oceanography*. New York: Pergamon, 1998.

[13] C. Fang, P. R. Krishnaiah and B. N. Nagarsenker, "Asymptotic distributions of the likelihood ratio test statistics for covariance structures of the complex multivariate normal distributions," *J. Multivariate Analysis*, vol. 12, pp. 597–611, 1982.

[14] P. Ginzberg, "Quaternion matrices: statistical properties and applications to signal processing and wavelets," Ph. D. dissertation, Dept. Mathematics, Imperial College London, 2013.

[15] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, pp. 152–77, 1963.

[16] A. K. Gupta, "Distribution of Wilks' likelihood-ratio criterion in the complex case," *Ann. Instit. Statist. Math.*, vol. 23, pp. 77–87, 1971.

[17] A. K. Gupta, "On a test for reality of the covariance matrix in a complex Gaussian distribution," *J. Statist. Computation and Simulation,* vol. 2, pp. 333-342, 1973.

[18] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Stat.*, vol. 6, pp. 65–70, 1979.

[19] P. R. Krishnaiah,, J. C. Lee and T. C. Chang, "The distributions of the likelihood ratio statistics for tests of certain covariance structures of complex multivariate normal populations," *Biometrika*, vol. 63, pp. 543–549, 1976.

[20] P. R. Krishnaiah,, J. C. Lee and T. C. Chang, "Likelihood ratio tests on covariance matrices and mean vectors of complex multivariate normal populations and their applications in time series," Technical report TR-83-03, Center for Multivariate Analysis, Univ. Pittsburgh, 1983.

[21] J. M. Lilly, P. B. Rhines, M. Visbeck, R. Davis, J. R. Lazier, F. Schott and D. Farmer, "Observing deep convection in the Labrador Sea during winter 1994/95," *J. Phys. Oceanogr.,* vol. 29, 2065–98, 1999.

[22] J. M. Lilly and P. B. Rhines, "Coherent eddies in the Labrador Sea observed from a mooring," *J. Phys. Oceanogr.,* vol. 32, 585–98, 2002.

[23] C. N. K. Mooers, "A technique for the cross spectrum analysis of pairs of complex-valued time series, with emphasis on properties of polarized components and rotational invariants," *Deep-Sea Research*, vol. 20, pp. 1129–1141, 1973.

[24] W. Min and R. S. Tsay, "On canonical analysis of multivariate time series," *Statistica Sinica*, vol. 15, pp. 303–323, 2005.

[25] M. Miyata, "Complex generalization of canonical correlation and its application to a sea-level study," *J. Marine Research*, vol. 28, pp. 202–214, 1970.

[26] J. Navarro-Moreno, M. D. Estudillo-Martínez, R. M. Fernández-Alcalá & J. C. Ruiz-Molina, "Estimation of improper complex-valued random signals in colored noise by using the Hilbert space theory," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2859–2867, 2009.

[27] J. Navarro-Moreno, J. Moreno-Kayser, R. M. Fernández-Alcalá & J. C. Ruiz-Molina, "Widely linear estimation algorithms for second-order stationary signals," *IEEE Transactions on Signal Processing*, vol. 57, pp. 4930–4935, 2009.

[28] E. Ollila and V. Koivunen, "Generalized complex elliptical distributions," in *Proc. Third Sensor Array and Multichannel Signal Processing Workshop*, Sitges, Spain, July, 2004, pp. 460–4.

[29] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Transactions on Signal Processing* vol. 44, pp. 2637–2640, 1996.

[30] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 411–420, 1997.

[31] G. Reinsel, *Elements of Multivariate Time Series Analysis (2nd Ed)*. New York: Springer, 1997.

[32] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*, Cambridge UK: Cambridge University Press, 2010.

[33] P. J. Schreier, L. L. Scharf and A. Hanssen, "A generalized likelihood ratio test for impropriety of complex signals," *IEEE Signal Process. Lett.*, vol. 13, pp. 433–6, 2006.

[34] A. M. Sykulski, S. C. Olhede, J. M. Lilly & J. J. Early, "Stochastic modeling and estimation of stationary complex-valued signals," ArXiv preprint arXiv:1306.5993v3, 2016. Available: http://arxiv.org/abs/1306.5993v3.

[35] A. M. Sykulski, S. C. Olhede & J. M. Lilly, "An improper complex autoregressive process of order one," *IEEE Transactions on Signal Processing*, vol. 64, pp. 6200-6210, 2016.

[36] J. K. Tugnait and S. A. Bhaskar, "Testing for impropriety of multivariate complex random processes," In Proc. 2016 IEEE Intern. Conf. Acoustics, Speech & Signal Processing (ICASSP 2016), pp. 4264-4268, Shanghai, China, March 20-25, 2016.

[37] H. van Haren and C. Millot, "Rectilinear and circular inertial motions in the Western Mediterranean Sea," *Deep-Sea Research Part I*, vol. 51, pp. 1441–55, 2004.

[38] A. T. Walden, "Rotary components, random ellipses and polarization: a statistical perspective," *Phil. Trans. R. Soc. A*, vol. 371, doi:10.1098/rsta.2011.0554, 2013.

[39] A. T. Walden and P. Rubin-Delanchy, "On testing for impropriety of complex-valued Gaussian vectors," *IEEE Transactions on Signal Processing* vol. 57, pp. 825–834, 2009.

[40] A. T. Walden, E. J. McCoy and D. B. Percival, "The effective bandwidth of a multitaper spectral estimator," *Biometrika*, vol. 82, 201–214, 1995.

[41] A. M. Yaglom, *Correlation Theory of Stationary and Related Random Functions, Volume I: Basic Results*. New York: Springer, 1987.

[42] G. A. Young and R. L.Smith, *Essentials of Statistical Inference*. Cambridge UK: Cambridge University Press, 2005.

Swati Chandna received the M.S. degree in Mathematics from the University of Houston, Texas, U.S.A. in 2009, and the Ph.D. degree in statistics from Imperial College London, U.K., in 2013. She worked as a Research Fellow at the University of Surrey, U.K., from 2013 to 2014, and then joined the Department of Statistical Science at University College London, U.K., where she is currently a Research Associate. Her research interests include modeling and inference for networks, frequency domain analysis of multichannel complex-valued time series and its applications in signal processing.

Andrew T. Walden (A'86-M'07-SM'11) received the B.Sc. degree in mathematics from the University of Wales, Bangor, U.K., in 1977, and the M.Sc. and Ph.D. degrees in statistics from the University of Southampton, Southampton, U.K., in 1979 and 1982, respectively. He was a Research Scientist at BP, London, U.K., from 1981 to 1990, and then joined the Department of Mathematics at Imperial College London, London, U.K., where he is currently a Professor of statistics.