# A Widely Linear Complex Autoregressive Process of Order One

Adam M. Sykulski, *Member, IEEE,* Sofia C. Olhede, *Member, IEEE,*
and Jonathan M. Lilly, *Senior Member, IEEE*

*Abstract*—We propose a simple stochastic process for modeling improper or noncircular complex-valued signals. The process is a natural extension of a complex-valued autoregressive process, extended to include a widely linear autoregressive term. This process can then capture elliptical, as opposed to circular, stochastic oscillations in a bivariate signal. The process is order one and is more parsimonious than alternative stochastic modeling approaches in the literature. We provide conditions for stationarity, and derive the form of the covariance and relation sequence of this model. We describe how parameter estimation can be efficiently performed both in the time and frequency domain. We demonstrate the practical utility of the process in capturing elliptical oscillations that are naturally present in seismic signals.

## I. Introduction

COMPLEX-valued stochastic processes are useful models for parameterizing bivariate signals. Such models are in widespread use in applications including oceanography [1] and functional Magnetic Resonance Imaging [2]. The theory for complex-valued representations has been developed both in the context of stochastic processes [3], as well as for signal processing [4], with notable recent developments in [5] and [6]. The complex-valued representation is sometimes preferred to the bivariate representation, due to its compactness and interpretability [7]. For example, complex-valued signals can be naturally decomposed into analytic and anti-analytic signals [8], and provide a practical framework for assessing impropriety or noncircularity in a complex signal [9]. On the other hand, the bivariate vector representation can provide better physical understanding of the generating mechanism of the modeled process, and its usage is commonplace in the time series community, see e.g. [10, Ch. 10–11].

The complex autoregressive process [11], [12] is a generalization of a real-valued autoregressive process, in which

A. M. Sykulski and S. C. Olhede are with the Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK (emails: a.sykulski@ucl.ac.uk, s.olhede@ucl.ac.uk).
S. C. Olhede is also with the Alan Turing Institute, UK.
J. M. Lilly is with NorthWest Research Associates, PO Box 3027, Bellevue, WA, USA (email: lilly@nwra.com)

the autoregressive coefficients and noise increments are both complex-valued. Typically, the real and imaginary components of the noise are assumed to be independent and identically distributed [12], thus creating statistically isotropic (i.e. circular) oscillations in the signal. The process is therefore said to be statistically *circular* or *proper*, defined subsequently, as opposed to one that is *noncircular* or *improper*.

In many real-world observations of complex-valued signals, noncircular or improper structure is expected to be present; examples include seismic traces [13], oceanographic velocity measurements [14], and wind measurements [5]. In Fig. 1, we display a bivariate signal from a seismic trace of the 1991 Solomon Islands Earthquake, previously studied in [15]–[17]. In Fig. 1(c), the elliptical orbital shape of the oscillations become apparent when viewed in the complex plane, and we will refer to such motion as "elliptical oscillations." In cases such as these, in which the signal appears improper, a proper process would be a poor choice of model and would fail to summarize important characteristics of the data.

Motivated by this, in this paper we generalize the complex autoregressive processes to a *widely linear* complex autoregressive process that is statistically improper or noncircular. The notion of wide linearity was introduced in [18], and we use this concept to relate the complex-valued process $Z_t$ to its previous value $Z_{t-1}$ *and* its complex conjugate $Z_{t-1}^*$. Our process is order one and hence Markovian, such that $Z_t$ is not dependent on $Z_{t-k}$ (given $Z_{t-1}$) for $k > 1$, and is a special case of the widely linear autoregressive moving average (ARMA) model of [19]. The widely linear ARMA is a general and flexible framework for improper processes that is well understood in the context of signal prediction [19], [20], and signal estimation [21], in settings where parameter values are assumed to be known. In practice, these parameters would need to be estimated when modeling real-world signals.

In this paper, we provide time- and frequency-domain techniques for estimating the parameters of our order one widely linear process, where we will also derive the form of the covariance and relation sequences. A key innovation will be to relate the process to a real-valued bivariate vector autoregressive process, which will provide intuitive understanding of the process generating mechanism, and will simplify the problem to the estimation of five real-valued parameters. We will demonstrate that despite its simplicity, the model we propose can effectively capture the elliptical oscillatory structure present in the seismic signal displayed in Fig. 1.

We contrast our widely linear process with the improper complex autoregressive process of [12], [22], where the im-
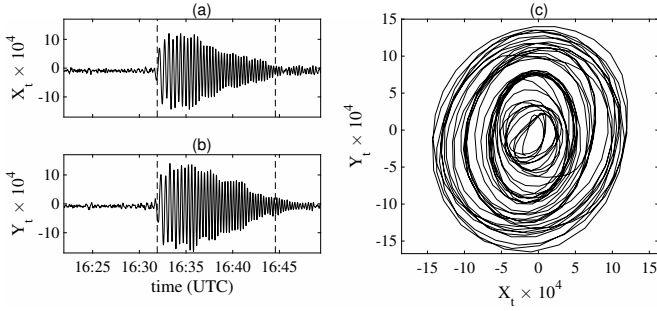
Fig. 1. Seismic traces from the Feb. 9th, 1991 Solomon Islands earthquake as measured from the Pasadena recording station in California. The radial component, $X_t$, is displayed in (a) and the vertical component, $Y_t$, is displayed in (b). In (c) we display the complex-valued signal, $Z_t = X_t + iY_t$, from 16:31:55 to 16:44:35 (UTC), as indicated by the vertical dashed boundaries in (a) and (b).

propriety is created using improper noise. We will demonstrate how the inclusion of a widely linear autoregressive component allows our process to reproduce stochastic elliptical oscillations despite being order one, whereas to generate elliptical oscillations in the framework of [12], an order two process is required, as investigated in [23]. We propose our order one process as a simpler and more intuitive model for generating elliptical oscillations in a complex-valued signal.

## II. BACKGROUND

Consider a zero-mean complex-valued stationary stochastic process $Z_t$ where $t \in \mathbb{Z}$. The covariance sequence, $s_\tau$, and the relation sequence, $r_\tau$, are defined at lag $\tau \in \mathbb{Z}$ by

$$s_\tau = \mathrm{E}\{Z_t Z_{t+\tau}^*\}, \quad r_\tau = \mathrm{E}\{Z_t Z_{t+\tau}\},$$

where $\mathrm{E}\{\cdot\}$ denotes expectation and $Z_t^*$ is the complex conjugate of $Z_t$. The process $Z_t$ is said to be proper if

$$r_\tau = 0, \quad \forall \tau \in \mathbb{Z},$$

and is improper otherwise. A proper process therefore has a relation sequence equal to zero, and is also commonly referred to as a *circular* process [4]. It can be shown that the second-order statistical properties of a proper process are isotropic, such that its distribution is invariant to rotation. Second-order proper processes can be treated much like second-order real-valued processes, in that the second-order properties are fully specified by the covariance sequence.

The improper complex autoregressive process of [12], [22] is given by

$$Z_t = \sum_{j=1}^{p} g_j Z_{t-j} + \nu_t, \quad g_j, \nu_t \in \mathbb{C}, \qquad (1)$$

where $p$ is the order of the process, and where $\nu_t$ is a noise process that is permitted to be noncircular or improper. The coefficients $g_j$ are in general assumed to be complex-valued. An important special case of this process is the proper complex autoregressive process of order one [11], which we denote by $Z_t'$ and has three real-valued parameters $\{a, \theta, \sigma_\epsilon^2\}$, and is given by

$$Z_t' = a e^{i\theta} Z_{t-1}' + \epsilon_t, \quad a \geq 0, \qquad (2)$$

where the complex autoregressive coefficient has been expressed in terms of an amplitude $a$ and phase $\theta$, both real-valued. Here $\{\epsilon_t\}$ is a sequence of i.i.d. complex-valued Gaussian noise where the real and imaginary parts are independent, and each has zero mean and variance given by $\sigma_\epsilon^2 > 0$. The process is stationary if and only if $a < 1$, with variance given by $2\sigma_\epsilon^2/(1-a^2)$. In such cases, $a$ is commonly referred to as the damping parameter. For identifiability $a$ is not permitted to be negative, as a negative autoregression is achieved instead when $a > 0$ and $\pi/2 + 2k\pi < \theta < 3\pi/2 + 2k\pi$, where $k \in \mathbb{Z}$. The parameter $\theta \in \mathbb{R}$ is the angle of a rotation of the process at each time step and is usually referred to as the spin parameter [24]. The process $Z_t'$ is an example of a proper process, and we will refer back to this process when we construct our widely linear improper process.

The improper complex autoregressive model (1) was generalized to a class of autoregressive moving average (ARMA) models in [19], which uses finite-length widely linear filters on both the autoregressive and noise components such that

$$Z_t = \sum_{j=1}^{p} g_j Z_{t-j} + \sum_{j=1}^{p} h_j Z_{t-j}^* + \sum_{j=0}^{q} k_j \epsilon_{t-j} + \sum_{j=0}^{q} l_j \epsilon_{t-j}^*. \quad (3)$$

This more general framework has larger flexibility in modeling improper signals. A challenge however is that as $\{g_j, h_j, k_j, l_j\}$ are complex-valued, then the number of parameters that need to be estimated is large, even for moderate values of $p$ and $q$. In this paper we shall focus on the special case of $p = 1$ and $q = 0$ in (3), thus creating a simple order-one widely linear improper process.

Finally, [12] more generally show that any second-order stationary process can be expressed as a widely linear filter of a complex-proper white noise process $\epsilon_t$ such that

$$Z_t = \sum_{j=-\infty}^{\infty} k_j \epsilon_{t-j} + \sum_{j=-\infty}^{\infty} l_j \epsilon_{t-j}^*, \quad k_j, l_j, \epsilon_t \in \mathbb{C}. \quad (4)$$

This is in essence the complex-analogue to the well-known Wold decomposition for real-valued processes [25], and in general this will be an infinite-order process.

## III. THE WIDELY LINEAR COMPLEX AUTOREGRESSIVE ORDER ONE PROCESS

In this section we introduce the widely linear complex autoregressive process of order one. We do this by extending (2) to a widely linear form for the autoregressive and noise components of the process, such that it is parameterized by seven real-valued parameters. We will subsequently reduce this model to five free parameters, by constraining two parameters, for practical reasons that we shall discuss shortly.

We call our process the widely linear complex autoregressive process of order one, denoted by $Z_t$, which has parameters $\{\lambda, \alpha, \gamma, \phi, \sigma_\nu^2, c_\nu\}$, and is given by

$$Z_t = \lambda e^{i\alpha} Z_{t-1} + \gamma e^{i\phi} Z_{t-1}^* + \nu_t, \quad \lambda, \gamma \geq 0, \qquad (5)$$

where $\{\nu_t\}$ is a sequence of i.i.d. complex-valued Gaussian noise with variance $\sigma_\nu^2 = \mathrm{E}\{\nu_t \nu_t^*\} > 0$ and relation at lag zero specified by $c_\nu = \mathrm{E}\{\nu_t \nu_t\} \in \mathbb{C}$. The noise process $\nu_t$ is specified by three real-valued parameters as $c_\nu$ is in

general complex-valued, therefore our process has seven real-valued parameters. The remaining four parameters $\{\lambda, \alpha, \gamma, \phi\}$ define an iterative relationship between the process $Z_t$ and the previous value $Z_{t-1}$ as well as its conjugate $Z_{t-1}^*$, where $\gamma$ and $\phi$ are respective analogues of the damping parameter, $\lambda$, and the spin parameter, $\alpha$, in this widely linear setting. Further intuition for each of the parameters is gained when we derive covariance sequences in Sections IV.

The widely linear complex autoregressive process of order one is recovered from the ARMA model (3) of [19] by taking $p = 1$ and $q = 0$. Furthermore, we combine the widely linear noise term $k_0 \epsilon_t + l_0 \epsilon_t^*$ into the noise process $\nu_t$, which has variance $\sigma_\nu^2 = |k_0|^2 + |l_0|^2$ and relation at lag zero $c_\nu = 2 k_0 l_0$. For this reason $\nu_t$ is commonly referred to as *doubly white noise* [12], as it is the pointwise superposition of two complex-proper white noise processes. As we have implicitly chosen the initial phase angle of the noise, we have for parsimony reduced the number of real-valued parameters from the eight used in (3) (with $p = 1$ and $q = 0$), to the seven we have defined in (5), without any loss of generality.

The model (5) introduces wide linearity in the autoregressive component of an order one process, and allows the process to map out stochastic elliptical oscillations, as we now demonstrate. In Fig. 2, we contrast realizations from our process with an improper order one process (1) from the framework of [12] using doubly white noise. The model (1) only has impropriety in the noise component—equivalent to setting $\gamma = 0$ in (5). It is clear from the figure that when the autoregressive coefficient $\lambda$ is close to unity (in panels (a) and (b)), then the widely linear complex autoregressive process (5) has a tendency to generate elliptical oscillations (panel (a)), whereas an improper order one process (1) generates oscillations that appear to be circular (panel (b)). This is because the motion (1) is largely determined by $Z_t = \lambda e^{i\alpha} Z_{t-1}$, a deterministic component which specifies a circular oscillation. When $\lambda$ is low (panels (c) and (d)), then the noise term $\nu_t$ dominates both signals. As $\nu_t$ is doubly white noise, then both processes are improper in their distribution, but neither generate elliptical oscillations that resemble the seismic traces seen in Fig 1(c)—this can only be generated using (5) with a larger autoregressive coefficient, and with $\gamma > 0$.

In general, the process (5) has seven real-valued parameters, whereas (2) has three. The additional four parameters in (5) are present because each component in (2)—the deterministic autoregressive component and the noise component—has been given its own improper elliptical structure with two parameters respectively, one to *stretch* ($\gamma$ and $|c_\nu|$) and one to *rotate* ($\phi$ and $\arg\{c_\nu\}$). Here we have defined $|c_\nu|$ as the complex modulus of $c_\nu$, and $\arg\{c_\nu\}$ as the complex argument or phase. In practical problems the four parameters $\{\gamma, \phi, |c_\nu|, \arg\{c_\nu\}\}$ will be difficult to identify simultaneously from observed signals. For this reason we reduce the widely linear process (5) to five free parameters, effectively making two parameters redundant, which in our case will be $|c_\nu|$ and $\arg\{c_\nu\}$. This is achieved by "aligning" the elliptical structure of the deterministic autoregression and the stochastic noise. The appeal of reducing to five parameters is that the process will then have the same number of free parameters as an improper order one
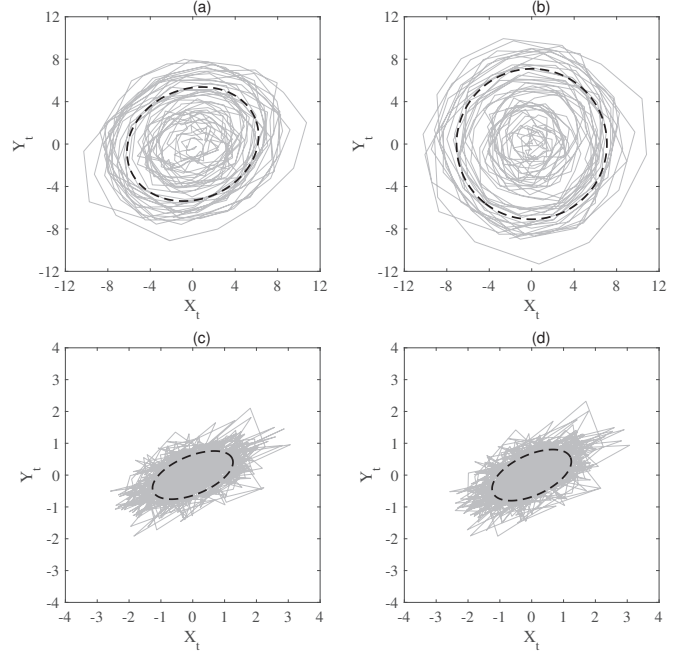


Fig. 2. Simulated signals of length 512 (——), from the widely linear complex autoregressive process (5) in (a) and (c), and from an order one improper process (1) in (b) and (d). In (a) and (b), the parameters are $\lambda = 0.99$, $\alpha = \pi/6$, $\sigma_\nu^2 = 1$, and $c_\nu = \exp(3\pi i/4)/5$. In (c) and (d), the parameters are $\lambda = 0.297$, $\alpha = \pi/6$, $\sigma_\nu^2 = 1$, and $c_\nu = 2\exp(3\pi i/4)/3$. For the widely linear complex autoregressive process (5) in (a) and (c), we additionally set $\gamma = 0.099$ and $\phi = -\pi/4$, which are set to zero in (b) and (d) for the model (1). Expected second moment ellipses for the standard deviation are also given (- - -), which are calculated from Section IV.

linear model (1), but will produce elliptical oscillations using widely linear forcing (*cf.* Fig. 2). The order two process for elliptical oscillations of [23] in general requires seven real-valued parameters.

To reduce our model to five free parameters, we equate the process to a bivariate process with elliptical covariance structure. We start from the complex-proper autoregressive process of order one (2). We rewrite this process as a bivariate process in terms of $(X_t' \ Y_t')^T$, where $Z_t' = X_t' + iY_t'$,

$$\begin{pmatrix} X_t' \\ Y_t' \end{pmatrix} = aR \begin{pmatrix} X_{t-1}' \\ Y_{t-1}' \end{pmatrix} + \sigma_\epsilon \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}, \qquad (6)$$

where,

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

It follows that (6) is an isotropic real-valued bivariate vector process. Here $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are i.i.d. Gaussian random variables with mean 0 and variance 1. The matrix $R$ accomplishes a rotation by the angle $-\pi < \theta \leq \pi$ in the Cartesian plane. For identifiability we set $a \geq 0$ and for stationarity we require $a < 1$, as we show in Appendix A. We then construct a new process $(X_t \ Y_t)^T$, which we call the *elliptical bivariate autoregressive process (of order one)*, using an elliptical transformation

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = QP \begin{pmatrix} X_t' \\ Y_t' \end{pmatrix}, \qquad (7)$$

where

$$Q = \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix}, \quad P = \begin{pmatrix} \frac{1}{\rho} & 0 \\ 0 & \rho \end{pmatrix},$$

in which $0 \leq \psi < \pi$ defines the orientation of an ellipse, and $0 < \rho \leq 1$ defines the eccentricity, $0 \leq \varepsilon < 1$, where $\varepsilon = \sqrt{1 - \rho^4}$. We call this transformation "elliptical" as it first stretches the $X$-axis by a factor of $1/\rho$, and compresses the $Y$-axis by a factor of $\rho$, before then rotating these axes through angle $\psi$. This transformation leads to elliptical statistical properties, and the process remains stationary for $a < 1$. This can be seen by examining the $2 \times 2$ covariance matrices associated with (6) and (7) which are respectively given by

$$\mathrm{E}\left\{\begin{matrix} X_t'^2 & X_t'Y_t' \\ X_t'Y_t' & Y_t'^2 \end{matrix}\right\} = \left(\frac{\sigma_\epsilon^2}{1-a^2}\right) I,$$

and

$$\mathrm{E}\left\{\begin{matrix} X_t^2 & X_tY_t \\ X_tY_t & Y_t^2 \end{matrix}\right\} = \left(\frac{\sigma_\epsilon^2}{1-a^2}\right) QP^2Q^T,$$

where $I$ is the identity matrix. The covariance matrix of $(X_t' \ Y_t')^T$ is isotropic or circular, whereas in general the covariance matrix of $(X_t \ Y_t)^T$ is elliptical with orientation $\psi$ and ratio of semi-minor to semi-major axis $\rho^2$.

The elliptical bivariate autoregressive process of order one is defined by five parameters, namely $\{a, \theta, \rho, \psi, \sigma_\epsilon^2\}$. We now express the widely linear complex autoregressive process of order one as $Z_t = X_t + iY_t$, and in Proposition 1 we relate the parameters of this process to those of the elliptical bivariate autoregressive process to form a five-parameter process.

*Proposition 1:* Suppose the process $(X_t \ Y_t)^T$ is an elliptical bivariate autoregressive process of order one, as defined in (6) and (7) by the parameters $\{a, \theta, \rho, \psi, \sigma_\epsilon^2\}$. This process is equivalent to a widely linear complex autoregressive process of order one (5), specified by $Z_t = X_t + iY_t$ where

$$Z_t = a\left\{\cos\theta + \frac{i\sin\theta}{2}\left(\frac{1}{\rho^2} + \rho^2\right)\right\}Z_{t-1}$$
$$+ \frac{a\sin\theta}{2}\left(\frac{1}{\rho^2} - \rho^2\right)e^{i\left(2\psi - \frac{\pi}{2}\right)}Z_{t-1}^*$$
$$+ \sigma_\epsilon\left\{\frac{e^{i\psi}}{\rho}\epsilon_{1,t} + \rho e^{i\left(\psi + \frac{\pi}{2}\right)}\epsilon_{2,t}\right\}.$$

The relationship with the parameters $\{\lambda, \alpha, \gamma, \phi, \sigma_\nu^2\}$ in the specification of (5) is given in Table I, with the final redundant parameter, the relation at lag zero $c_\nu$, specified by

$$c_\nu = \sigma_\epsilon^2\left(\frac{1}{\rho^2} - \rho^2\right)e^{i2\psi} = \sigma_\nu^2\left(\frac{\gamma}{\lambda\sin\alpha}\right)e^{i\left(\phi + \frac{\pi}{2}\right)}. \quad (8)$$

The proof of Proposition 1 is given in Appendix B. Therefore in order to represent the elliptical bivariate process (7) in terms of the widely linear complex process (5), we have required to use only five parameters. We can therefore simply fix $c_\nu$ using either $\{a, \theta, \rho, \psi, \sigma_\epsilon^2\}$ or $\{\lambda, \alpha, \gamma, \phi, \sigma_\nu^2\}$, as given in (8). This equivalence can be verified using the transformations in Table I. We note that this is how the value of $c_\nu$ was set earlier in Fig 2.

The eccentricity parameter, $\varepsilon$, can also be found in terms of the widely linear complex process parameters using Table I

$$\varepsilon = \sqrt{1 - \rho^4} = \sqrt{\frac{2\gamma}{\lambda|\sin\alpha| + \gamma}}. \quad (9)$$

Therefore for the widely linear process to return valid values for the eccentricity, $0 \leq \varepsilon < 1$, we require

$$\gamma \leq \lambda|\sin\alpha|. \quad (10)$$

Additionally, for the widely linear process to be stationary we require $a = \sqrt{\lambda^2 - \gamma^2} < 1$. Therefore from (10) it follows that $a > 0$. Then we see that stationarity is guaranteed when $\lambda < 1$, and otherwise for stationarity we require that

$$\gamma > \sqrt{\lambda^2 - 1}, \quad \text{when } \lambda \geq 1. \quad (11)$$

These inequalities also ensure that $\theta$ and $\sigma_\epsilon^2$ return valid values when mapping parameters from the complex to bivariate specifications. Increasing $\gamma$ increases the eccentricity, until eventually larger values of $\gamma$ are not valid. This means that there is a range of values for $\gamma$, which depends on both $\lambda$ and $\alpha$, for our five-parameter process to be valid and stationary. Combining the inequalities in (10) and (11) we see that we require $\lambda^2 < 1/\cos^2\alpha$ for our process to be valid and stationary. When $\alpha = 0$, the case of no spin, then $\gamma = 0$ from (10) and we require $\lambda < 1$ for stationarity. However as $\alpha$ increases then interestingly, there are parameter values for which $\lambda > 1$ and the process can still be stationary, unlike the proper case, although this then requires a non-zero $\gamma$ as can be seen from (11).

We gain further insight by analyzing the relationships in Table I. The first three parameters on each side of the table, $\{\lambda, \alpha, \gamma\}$ and $\{a, \theta, \rho\}$, have a direct one-to-one mapping, where $\{\lambda, \alpha\}$ and $\{a, \theta\}$ become identical as $\rho \to 1$ or $\gamma \to 0$. The parameters $a$ and $\lambda$ are monotonic functions of each other, as are $\alpha$ and $\theta$, which have the same sign in the range $(-\pi, \pi]$. The bivariate ellipse orientation $\psi$ and the complex-conjugate spin parameter $\phi$ are directly related, but are adjusted depending on the sign of $\theta$ and $\alpha$. The ratio of the variance parameters, $\sigma_\nu^2$ and $\sigma_\epsilon^2$, is determined by the eccentricity. The effect of each parameter can therefore be closely related row-by-row in Table I.

By connecting to a bivariate process, we have gained the advantages of both specifications: we benefit from the compactness and applicability of a complex representation, and we benefit from the interpretability and physical understanding gained from a bivariate representation. A particularly useful feature of complex signals is that we can perform hypothesis tests for impropriety [9], and we will demonstrate the insight gained from such an analysis in our seismic data example in Section VI.

## IV. COVARIANCE AND RELATION SEQUENCE

In this section we compute the covariance and relation sequences for the widely linear complex autoregressive process of order one. It follows directly from (5) that the process is Gaussian, as it is a linear combination of complex-valued Gaussian random variables. Therefore the covariance and relations sequences fully specify the process. These sequences would have complicated expansions if expressed analytically, so instead we find recurrence relationships between the lags.

| Bivariate elliptical to widely linear complex | Widely linear complex to bivariate elliptical |
|---|---|
| $\lambda = a\sqrt{\cos^2\theta + \frac{\sin^2\theta}{4}\left(\frac{1}{\rho^2} + \rho^2\right)^2}$ | $a = \sqrt{\lambda^2 - \gamma^2}$ |
| $\alpha = \texttt{atan2}\left\{\frac{\sin\theta}{2}\left(\frac{1}{\rho^2} + \rho^2\right), \cos\theta\right\}$ | $\theta = \text{sgn}(\alpha)\texttt{acos}\left(\sqrt{\frac{\lambda^2}{\lambda^2 - \gamma^2}}\cos\alpha\right)$ |
| $\gamma = \frac{a}{2}|\sin\theta|\left(\frac{1}{\rho^2} - \rho^2\right)$ | $\rho = \left(\frac{\lambda|\sin\alpha| - \gamma}{\lambda|\sin\alpha| + \gamma}\right)^{1/4}$ |
| $\phi = 2\psi - \text{sgn}(\theta)\frac{\pi}{2}$ | $\psi = \frac{\phi}{2} + \text{sgn}(\alpha)\frac{\pi}{4}$ |
| $\sigma_\nu^2 = \sigma_\epsilon^2\left(\frac{1}{\rho^2} + \rho^2\right)$ | $\sigma_\epsilon^2 = \sigma_\nu^2\frac{\sqrt{\lambda^2\sin^2\alpha - \gamma^2}}{2\lambda|\sin\alpha|}$ |

First we find that the variance ($\sigma_Z^2$) and relation at lag zero ($c_Z$) are given by

$$
\begin{aligned}
\sigma_Z^2 &= \text{E}\{Z_t Z_t^*\} \\
&= \text{E}\left\{\left(\lambda e^{i\alpha}Z_{t-1} + \gamma e^{i\phi}Z_{t-1}^* + \nu_t\right) \times \right. \\
&\qquad\qquad \left. \left(\lambda e^{-i\alpha}Z_{t-1}^* + \gamma e^{-i\phi}Z_{t-1} + \nu_t^*\right)\right\} \\
&= (\lambda^2 + \gamma^2)\sigma_Z^2 + \lambda\gamma e^{i(\alpha-\phi)}c_Z + \lambda\gamma e^{i(\phi-\alpha)}c_Z^* + \sigma_\nu^2,
\end{aligned} \tag{12}
$$

$$
\begin{aligned}
c_Z &= \text{E}\{Z_t Z_t\} \\
&= \text{E}\left\{\left(\lambda e^{i\alpha}Z_{t-1} + \gamma e^{i\phi}Z_{t-1}^* + \nu_t\right) \times \right. \\
&\qquad\qquad \left. \left(\lambda e^{i\alpha}Z_{t-1} + \gamma e^{i\phi}Z_{t-1}^* + \nu_t\right)\right\} \\
&= 2\lambda\gamma e^{i(\alpha+\phi)}\sigma_Z^2 + \lambda^2 e^{2i\alpha}c_Z + \gamma^2 e^{2i\phi}c_Z^* + c_\nu.
\end{aligned} \tag{13}
$$

We now combine (12) and (13) to solve for $\sigma_Z^2$ and $c_Z$, yielding

$$
\begin{pmatrix} \sigma_Z^2 \\ c_Z \\ c_Z^* \end{pmatrix} = \begin{pmatrix} \lambda^2 + \gamma^2 & \lambda\gamma e^{i(\alpha-\phi)} & \lambda\gamma e^{i(\phi-\alpha)} \\ 2\lambda\gamma e^{i(\alpha+\phi)} & \lambda^2 e^{2i\alpha} & \gamma^2 e^{2i\phi} \\ 2\lambda\gamma e^{-i(\alpha+\phi)} & \gamma^2 e^{-2i\phi} & \lambda^2 e^{-2i\alpha} \end{pmatrix} \begin{pmatrix} \sigma_Z^2 \\ c_Z \\ c_Z^* \end{pmatrix} + \begin{pmatrix} \sigma_\nu^2 \\ c_\nu \\ c_\nu^* \end{pmatrix} \tag{14}
$$

and hence

$$
\begin{pmatrix} \sigma_Z^2 \\ c_Z \\ c_Z^* \end{pmatrix} = M^{-1} \begin{pmatrix} \sigma_\nu^2 \\ c_\nu \\ c_\nu^* \end{pmatrix}, \tag{15}
$$

where

$$
M = \begin{pmatrix} 1 - \lambda^2 - \gamma^2 & -\lambda\gamma e^{i(\alpha-\phi)} & -\lambda\gamma e^{i(\phi-\alpha)} \\ -2\lambda\gamma e^{i(\alpha+\phi)} & 1 - \lambda^2 e^{2i\alpha} & -\gamma^2 e^{2i\phi} \\ -2\lambda\gamma e^{-i(\alpha+\phi)} & -\gamma^2 e^{-2i\phi} & 1 - \lambda^2 e^{-2i\alpha} \end{pmatrix}.
$$

The analytic form for the inverse matrix $M^{-1}$ in (15) is provided as part of the online software available at http://ucl.ac.uk/statistics/research/spg/software and is not included here for space considerations. When simulating signals from the process, in addition to satisfying the inequalities specified in (10), then for the process to be

stationary we require that the first observation is generated from the complex-valued normal distribution with mean zero, variance $\sigma_Z^2$, and relation at lag zero $c_Z$. See Section V-A for more detail on the complex-valued normal distribution. This is how the signals in Fig. 2 have been simulated, and more details on this can be found in the supporting online code.

After computing $\sigma_Z^2$ and $c_Z$ from (15), the covariance sequence, $s_\tau$, and the relation sequence, $r_\tau$, can be found for any $\tau > 0$ using the following recurrence relationships

$$
\begin{aligned}
s_\tau &= \text{E}\left\{Z_t Z_{t+\tau}^*\right\} \\
&= \text{E}\left\{Z_t\left(\lambda e^{-i\phi}Z_{t+\tau-1}^* + \gamma e^{-i\phi}Z_{t+\tau-1} + \nu_{t+\tau-1}^*\right)\right\} \\
&= \lambda e^{-i\alpha}s_{\tau-1} + \gamma e^{-i\phi}r_{\tau-1}, \tag{16} \\
r_\tau &= \text{E}\left\{Z_t Z_{t+\tau}\right\} \\
&= \text{E}\left\{Z_t\left(\lambda e^{i\phi}Z_{t+\tau-1} + \gamma e^{i\phi}Z_{t+\tau-1}^* + \nu_{t+\tau-1}\right)\right\} \\
&= \lambda e^{i\alpha}r_{\tau-1} + \gamma e^{i\phi}s_{\tau-1}. \tag{17}
\end{aligned}
$$

Therefore, after solving for $s_0 = \sigma_Z^2$ and $r_0 = c_Z$, we iterate to find $s_\tau$ and $r_\tau$ using (16) and (17). To find the sequences for negative lags we use the simple relationship $s_{-\tau} = s_\tau^*$ and $r_{-\tau} = r_\tau$.

From (16) and (17) we see that $\lambda$ and $\alpha$ contribute to the exponential decay of the autocovariance—this is expected as autoregressive processes are short memory. Conversely, $\gamma$ and $\phi$ have a "flipping" effect on $s_\tau$ and $r_\tau$, where the covariance is dependent on the relation at previous lags and vice-versa. This is a consequence of the widely linear form in (5), where the conjugate action creates an iterative 'flip' of the process about the real axis in the complex plane.

## V. METHODS FOR ESTIMATING PARAMETERS

In this section we detail how the parameters of the widely linear complex autoregressive process of order one can be estimated from an observed signal using maximum likelihood. We first resolve the exact form of the likelihood in Section V-A, and then provide an approximate method in the frequency domain in Section V-B. The latter method has practical advantages in real-world applications, as we shall discuss.

## A. Maximum Likelihood

Suppose that $Z_t$ follows a stationary widely linear complex autoregressive process or order one, as specified in (5) with parameters satisfying (10) and (11), then the probability distribution of $Z_t$ follows a complex-valued normal distribution. For a general complex-valued normally distributed random variable $z$, we denote its distribution by $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2, c)$, with mean $\mu$, variance $\sigma^2$, and relation at lag zero $c$. The probability density function of $z$ is then given by

$$p(z) = \frac{1}{\pi \sqrt{(\sigma^2)^2 + |c|^2}} \times$$
$$\exp \left\{ -\frac{1}{2} \begin{pmatrix} z^* - \mu^* & z - \mu \end{pmatrix} \begin{pmatrix} \sigma^2 & c \\ c^* & \sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} z - \mu \\ z^* - \mu^* \end{pmatrix} \right\}.$$

It then follows that the probability distribution of the widely linear complex autoregressive process or order one, denoted $Z_t$, is given by

$$Z_t \sim \mathcal{N}_{\mathbb{C}} \left( 0, \sigma_Z^2, c_Z \right),$$

where $\sigma_Z^2$ and $c_Z$ are found using (15). For a given observed signal $z_0, \ldots, z_{N-1}$ from the process $Z_t$, the probability of observing the first value $z_0$ directly follows from the probability distribution of $Z_t$,

$$p(z_0; \boldsymbol{\theta}) = \frac{1}{\pi \sqrt{(\sigma_Z^2)^2 + |c_Z|^2}} \times$$
$$\exp \left\{ -\frac{1}{2} \begin{pmatrix} z_0^* & z_0 \end{pmatrix} \begin{pmatrix} \sigma_Z^2 & c_Z \\ c_Z^* & \sigma_Z^2 \end{pmatrix}^{-1} \begin{pmatrix} z_0 \\ z_0^* \end{pmatrix} \right\},$$

where $\boldsymbol{\theta} = \{\sigma_Z^2, c_Z\}$. Next we make use of the Markovian property of the process to find the conditional distribution of $Z_t$ given $Z_{t-1} = z_{t-1}$, for $1 \leq t \leq N-1$

$$(Z_t | Z_{t-1} = z_{t-1}) \sim \mathcal{N}_{\mathbb{C}} \left( \lambda e^{i\alpha} z_{t-1} + \gamma e^{i\phi} z_{t-1}^*, \sigma_\nu^2, c_\nu \right),$$

such that the condition probability of observing $z_t$ given $z_{t-1}$ is

$$p(z_t | z_{t-1}; \boldsymbol{\theta}) = \frac{1}{\pi \sqrt{(\sigma_\nu^2)^2 + |c_\nu|^2}} \times$$
$$\exp \left\{ -\frac{1}{2} \begin{pmatrix} z_t^* - \mu_{z_t}^* & z_t - \mu_{z_t} \end{pmatrix} \begin{pmatrix} \sigma_\nu^2 & c_\nu \\ c_\nu^* & \sigma_\nu^2 \end{pmatrix}^{-1} \begin{pmatrix} z_t - \mu_{z_t} \\ z_t^* - \mu_{z_t}^* \end{pmatrix} \right\},$$

where $\boldsymbol{\theta} = \{\mu_{z_t}, \sigma_\nu^2, c_\nu\}$ and

$$\mu_{z_t} = \lambda e^{i\alpha} z_{t-1} + \gamma e^{i\phi} z_{t-1}^*. \tag{18}$$

The likelihood of observing the signal $z_0, \ldots, z_{N-1}$ is found by evaluating

$$p(z_0, \ldots, z_{N-1}; \boldsymbol{\theta}) = p(z_0; \boldsymbol{\theta}) \prod_{t=1}^{N-1} p(z_t | z_{t-1}; \boldsymbol{\theta}).$$

The log-likelihood (denoted $\ell_t(\boldsymbol{\theta})$) is therefore

$$\ell_t(\boldsymbol{\theta}) = \log \left( p(z_0; \boldsymbol{\theta}) \right) + \sum_{t=1}^{N-1} \log \left( p(z_t | z_{t-1}; \boldsymbol{\theta}) \right),$$

which for a widely linear complex autoregressive process of order one is found to be

$$\ell_t(\boldsymbol{\theta}) = -N \log \pi - \frac{N-1}{2} \log \left( (\sigma_\nu^2)^2 + |c_\nu|^2 \right)$$
$$- \sum_{t=1}^{N-1} \frac{1}{2} \begin{pmatrix} z_t^* - \mu_{z_t}^* & z_t - \mu_{z_t} \end{pmatrix} \begin{pmatrix} \sigma_\nu^2 & c_\nu \\ c_\nu^* & \sigma_\nu^2 \end{pmatrix}^{-1} \begin{pmatrix} z_t - \mu_{z_t} \\ z_t^* - \mu_{z_t}^* \end{pmatrix}$$
$$- \frac{1}{2} \log \left( (\sigma_Z^2)^2 + |c_Z|^2 \right) - \frac{1}{2} \begin{pmatrix} z_0^* & z_0 \end{pmatrix} \begin{pmatrix} \sigma_Z^2 & c_Z \\ c_Z^* & \sigma_Z^2 \end{pmatrix}^{-1} \begin{pmatrix} z_0 \\ z_0^* \end{pmatrix}, \tag{19}$$

with $\mu_{z_t}$ given in (18). The optimal parameter choice $\hat{\boldsymbol{\theta}}$ is then found by maximizing the log-likelihood (19)

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell_t(\boldsymbol{\theta}), \tag{20}$$

where $\boldsymbol{\Theta}$ is the permitted parameter range for $\boldsymbol{\theta}$, recalling that for the five-parameter process, $c_\nu$ is specified by (8), and that the inequalities (10) and (11) should also be satisfied.

## B. Frequency Domain "Whittle" Likelihood

The parameters of the widely linear complex autoregressive process of order one can also be computed in the frequency domain using Whittle's approximation to maximum likelihood [26], known as the 'Whittle likelihood.' This approximation of the time-domain likelihood is in the frequency domain, and relies solely on applying Fourier Transforms which can be computed in $\mathcal{O}(N \log N)$ operations. We use a bias-corrected form of the Whittle likelihood, which was extended to complex-valued signals in [27] and is given by

$$\ell(\boldsymbol{\theta}) = - \sum_{\omega \in \Omega} \left\{ \log \left( |f(\omega; \boldsymbol{\theta})| \right) + J^H(\omega) f^{-1}(\omega; \boldsymbol{\theta}) J(\omega) \right\}, \tag{21}$$

where $\Omega$ is the set of Fourier frequencies used in the estimation, $\boldsymbol{\theta}$ is the unknown parameter vector, and $J(\omega)$ and $f(\omega; \boldsymbol{\theta})$ are given by

$$J(\omega) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \begin{pmatrix} Z_t \\ Z_t^* \end{pmatrix} e^{-i\omega t}, \tag{22}$$

$$f(\omega; \boldsymbol{\theta}) = \begin{pmatrix} \bar{S}(\omega; \boldsymbol{\theta}) & \bar{R}(\omega; \boldsymbol{\theta}) \\ \bar{R}^*(\omega; \boldsymbol{\theta}) & \bar{S}(-\omega; \boldsymbol{\theta}) \end{pmatrix}. \tag{23}$$

The vector $J(\omega)$ is the Discrete Fourier Transform for a signal $Z_0, \ldots, Z_{N-1}$, evaluated at the Fourier frequencies. The matrix $f(\omega; \boldsymbol{\theta})$ contains the expected periodogram, $\bar{S}(\omega; \boldsymbol{\theta}) = \mathrm{E}\{|J(\omega)|^2\}$, and the expected *complementary* periodogram, $\bar{R}(\omega; \boldsymbol{\theta}) = \mathrm{E}\{J(\omega)J(\omega)\}$, and is dependent on both the signal length $N$, and the parameter vector $\boldsymbol{\theta}$. Expected periodograms are used in (21), rather than theoretical spectral densities, as this removes the known bias effects from using Discrete Fourier Transforms in (21) for finite $N$. We then compute $\bar{S}(\omega; \boldsymbol{\theta})$ and $\bar{R}(\omega; \boldsymbol{\theta})$ directly from the covariance and relation sequences using the relationships

$$\bar{S}(\omega; \boldsymbol{\theta}) = \sum_{\tau = -(N-1)}^{N-1} s_\tau(\boldsymbol{\theta}) \left( 1 - \frac{|\tau|}{N} \right) e^{-i\omega\tau}, \tag{24}$$

$$\bar{R}(\omega; \boldsymbol{\theta}) = \sum_{\tau = -(N-1)}^{N-1} r_\tau(\boldsymbol{\theta}) \left( 1 - \frac{|\tau|}{N} \right) e^{-i\omega\tau}. \tag{25}$$

The expected complementary periodogram $\bar{R}(\omega; \boldsymbol{\theta})$ will in general be a complex-valued quantity. The usefulness of (24) and (25) is that they can be computed in $\mathcal{O}(N \log N)$ time as they are Discrete Fourier Transforms, ensuring the Whittle likelihood remains an $\mathcal{O}(N \log N)$ procedure in this bias corrected version. This method is proven to be statistically consistent in [28], and is shown to significantly remove bias effects for sample sizes as large as 1,000 data points.

For the widely linear complex autoregressive process of order one, $s_\tau$ and $r_\tau$ in (24) and (25) are computed using the recurrence relationships given in Section IV, which is an $\mathcal{O}(N)$ computation. Both $f(\omega; \boldsymbol{\theta})$ and $J(\omega)$ are then computed using Fast Fourier Transforms, thus giving $\mathcal{O}(N \log N)$ efficiency overall. The optimal parameter choice $\hat{\boldsymbol{\theta}}$ is then found by maximizing the Whittle likelihood in the same way as (20).

The advantage of performing maximum likelihood in the frequency domain is that we can restrict the range of frequencies used in the summation in (21). This allows the parameter estimation to be performed *semi-parametrically*, by ignoring frequencies that are known to be contaminated or not specified well by the model. For example, this was used in [29] to remove the effect of eddies when estimating the parameters of a turbulent flow model for the ocean surface. We will also employ such semi-parametric procedures in our data example in Section VI.

We note that the Whittle likelihood can be alternatively used with tapered spectral estimates in (22), where the triangle kernel $1 - |\tau|/N$ in (24) and (25) is then replaced with a modified kernel of smaller width, as documented in [28]. Tapering the likelihood helps reduce mean square error in parameter estimates when the process is long memory or has steep spectral slopes. As the widely linear complex autoregressive process of order one is short memory, then tapering is unnecessary and we use the periodogram approach defined in (21)–(25).

## VI. APPLICATION TO SEISMIC DATA

In this section we investigate using the widely linear complex autoregressive process of order one as a model for seismic data. We analyze the seismic trace from the Feb 9th 1991 Solomon Islands earthquake, as presented in Fig. 1 in Section I. The data is sampled every second and is freely available from `http://ds.iris.edu/wilber3`. All the results in this section (and all figures in this paper) are exactly reproducible with MATLAB code available from `http://ucl.ac.uk/statistics/research/spg/software`.

The seismic trace consists of three components: a radial, vertical, and transverse signal [30]. We model the radial and vertical components as a complex-valued signal, as they are strongly coupled due to the presence of a Rayleigh wave. We do not analyze the transverse component as the expression of the Rayleigh wave does not exist in the transverse signal [17]. The radial, $X_t$, and vertical, $Y_t$, signal are displayed in Figs. 1(a) and 1(b) respectively. Our analysis will first focus on the segment between the dashed vertical lines in the figure. We combine these signals within this partition to form a single complex-valued signal, $Z_t = X_t + iY_t$, as displayed on the
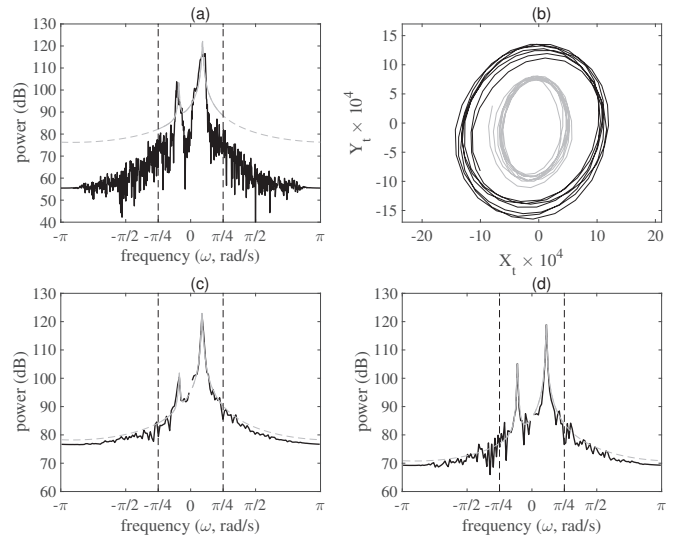


Fig. 3. Panel (a) displays the periodogram (——) and model fit (——) of the signal $Z_t$ displayed in Fig. 1(c). Panel (b) shows $Z_t$ in the interval (UTC) 16:33:46 to 16:36:26 (——) and 16:38:19 to 16:40:59 (——). The periodogram (——) and model fit (——) are displayed for $Z_t$ in the intervals (c) 16:33:46 to 16:36:26 and (d) 16:38:19 to 16:40:59. In (a), (c), and (d), parameter estimation is performed for $\omega \in [-\pi/4, \pi/4]$, as indicated by the vertical dashed boundaries, and we extend the fitted lines to all frequencies (- - -).

complex plane in Fig. 1(c). The signal has evident improper structure, as can be seen from the elliptical paths of the signal in the complex plane.

We first fit the widely linear complex autoregressive process of order one to the entire signal displayed in Fig. 1(c) using the Whittle likelihood, as detailed in Section V. The periodogram of the data, and the resulting model fit of the periodogram, are displayed in Fig. 3(a). For all parameter estimates in this section, we perform the Whittle likelihood estimation semi-parametrically over a reduced range of frequencies, $\omega \in [-\pi/4, \pi/4]$, as the signal energy is strongly concentrated within this frequency range. For complex-valued signals, the spectrum is defined at both negative and positive frequencies, and will in general be asymmetric. The two peaks of different magnitude on each side of the spectra, at approximately the same frequency, indicate elliptical oscillatory motion. Our fitted process has located the frequency of these peaks, but overall is a poor fit to the periodogram. This is due to the non-stationarity of the signal. Inspecting Fig. 1(c) in more detail we can see that the amplitude, eccentricity and orientation of the elliptical oscillations are changing in time. Our model, which is stationary, is not able to capture this variable structure.

To investigate these nonstationary effects we separately analyze two segments of the data, each 161 seconds (or data points) long, spanning the periods 16:33:46 to 16:36:26 and 16:38:19 to 16:40:59 (UTC) respectively. The complex-valued signals corresponding to these time periods are displayed in Fig. 3(b). The choice of window length is motivated by the example signal itself, and has been selected such that it is as short as possible (to capture as much time variability as possible), while still being able to robustly estimate all five free parameters. For automated windowing procedures we refer the reader to [31], which is outside the scope of this
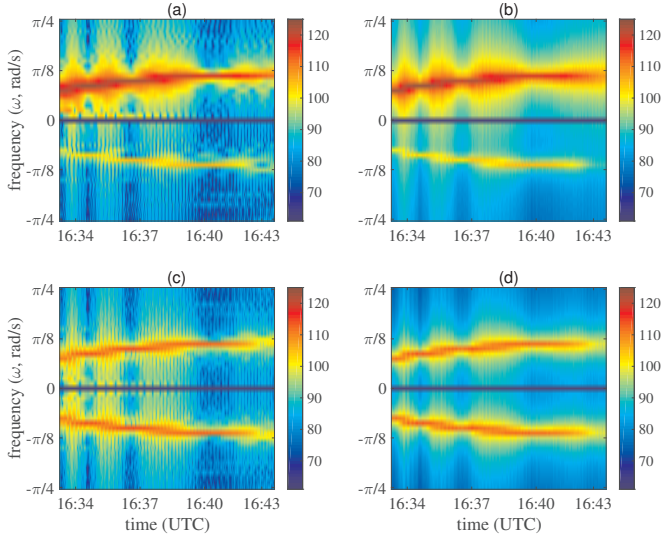
Fig. 4. Spectrograms using a 161-second sliding window of the seismic signal of Fig. 1. Panel (a) is the evolving periodogram of $Z_t$, (b) is the evolving model fit to the periodogram using our widely linear process, (c) is the magnitude of the evolving complementary periodogram of $Z_t$ and (d) is the magnitude of the evolving model fit to the complementary periodogram. The color scale is given in decibels.



Fig. 5. Estimates of (a) the eccentricity, $\varepsilon$, and (b) the orientation, $\psi$, in radians, of the seismic signal of Fig. 1. (—-) are the estimates from fitting a widely linear complex autoregressive process of order one across a 161-second sliding window, with the 95% confidence intervals given in gray. (- · -) are the estimates from the method of [33], and (- - -) are the estimates obtained from the Fourier Transform evaluated at the two frequency peaks. All estimates and confidence intervals, including the nonparametric techniques, have been smoothed in time using a moving average window of width 11.

paper. The periodograms and model fits of each segment are displayed in Figs. 3(c) and 3(d). The process is now seen to be a good fit to these shorter signals. The optimal parameters are significantly different for each segment. For example, the eccentricity estimate in the first segment is 0.29 whereas in the second segment it is 0.56. These differing eccentricity estimates are related to the different ratios of the amplitudes of the two peaks in each respective periodogram.

The appropriateness of our process to modeling shorter segments of this signal suggests that a *locally stationary* modeling assumption should be used, see e.g. [32]. To investigate this in more detail we perform the model fit to a rolling 161-second window over the entire signal.

In Fig. 4(a) and 4(b) we plot the spectrogram of the data—that is, a moving window of the periodogram—together with the spectrogram of the expected periodogram from our model fit. These spectrograms are only shown for the frequencies that have been used in the fit. Note that the zero frequency is not included in the fit as we have removed the sample mean for each segment. From the figure it can be seen that the widely linear complex autoregressive process of order one captures the overall shape of the spectrum at each time slice, as well as its evolution over time. In particular, the process has captured the gradually changing frequency of the oscillations.

We also display, in Fig. 4(c) and 4(d), the time-frequency plots for the magnitude of the complementary periodogram, and the resulting model fit (respectively). The complementary periodogram forms a Fourier pair with the sample relation sequence, and as a consequence complementary periodograms from observed improper processes are expected to exhibit noticeable structure. This structure has been captured well in the model fit, which is important, as the complementary periodogram is where information regarding impropriety—such as the expected orientation of elliptical motion—is contained.
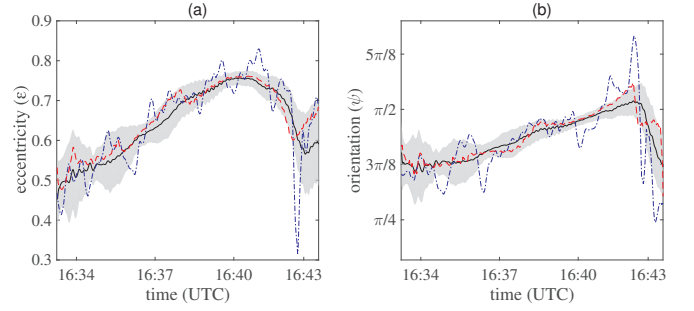
The complementary periodogram is complex-valued, but we have not included plots for its phase here, for space saving considerations.

In Fig. 5(a) and 5(b) we display the time-varying eccentricity and orientation parameters from the model fit, calculated using Table I and (9). Other informative time-varying summaries can also be found, such as of the noise variance or damping terms. These can be generated as part of the online software. In Fig. 5 we compare with two alternative methods. First we compare against results obtained from the nonparametric deterministic approach of [33], which models the signal as a time varying ellipse, thus providing a good comparison to our results despite being a complementary approach. Secondly, we compare against a simple nonparametric approach of comparing the Fourier transform at the positive and negative frequency peaks in the power spectral density, which we denote as $\pm\omega_{\max}$, where it can be readily shown that eccentricity and orientation estimates can be obtained from

$$\hat{\varepsilon} = \frac{2\sqrt{|J_Z(\omega_{\max})J_Z(-\omega_{\max})|}}{|J_Z(\omega_{\max})| + |J_Z(-\omega_{\max})|}$$

$$\hat{\psi} = \frac{1}{2}\left[\arg\{J_Z(\omega_{\max})\} + \arg\{J_Z(-\omega_{\max})\}\right],$$

with $J_Z(\omega)$ defined as the top row of (22). In Fig. 5, the estimated eccentricities and orientations broadly agree across the different methods. The values obtained from our parametric model are in general smoother, as the method smooths over information across frequencies when estimating parameters. The usefulness of our stochastic process is that it prescribes a generating mechanism providing physical insight and the ability to replicate signals, which the alternative purely diagnostic metrics do not provide.

Another useful feature of a stochastic modeling approach is that we can calculate confidence intervals for parameter estimates, by numerically computing the Hessian of the Whittle likelihood, as detailed in [29]. In Fig. 5 we include the 95% confidence intervals for our parameter estimates, where care must be taken when assessing significance across time as these are not pointwise simultaneous confidence intervals.
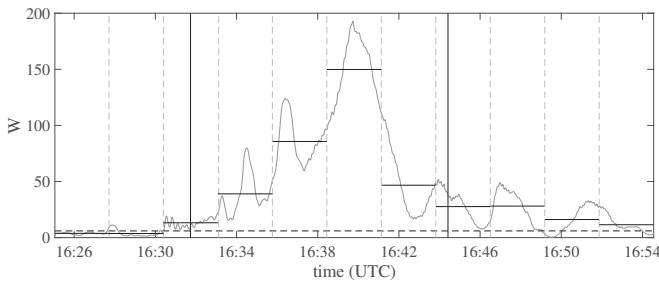
Fig. 6. The likelihood ratio statistic $W$ (——) over time, smoothed with a moving average window of width 11, from fitting the proper (2) and widely linear improper (5) complex autoregressive order one processes to $Z_t$ across a 161-second sliding window. The signal is then divided into 11 non-overlapping windows, as indicated by the vertical gray-dashed lines, where the vertical black solid lines indicate the analysis window of Figs. 4–6. The likelihood ratio statistic for each of these windows is then given by (——). We also display the 95th percentile of a $\chi_2^2$ distribution (- - -), and after applying a False Discover Rate (FDR) procedure to control the rate of false positives, we reject propriety in each segment except the first two.

Finally, another advantage of the stochastic modeling approach is that we can perform a parametric hypothesis test for impropriety, to test for statistical significance for when an improper model should be used. This approach is simpler than performing the test of [9], under the assumption that the parametric model is appropriate. We perform the parametric test by also fitting the *proper* complex autoregressive order one process (2) to rolling windows of observations, in exactly the same manner as performed for the widely linear improper process. We then perform a likelihood ratio test, as proposed in [27], to see if there is significant statistical evidence to suggest the null hypothesis of a proper process should be rejected in favor of an improper process. To do this we compute the likelihood ratio statistic, given by $W = 2\{\ell(\boldsymbol{\theta}_{\text{alt}}) - \ell(\boldsymbol{\theta}_{\text{null}})\}$, where "alt" and "null" denote the alternative and null models respectively. We compare the likelihood ratio statistic with the 95th percentile of a $\chi_2^2$ distribution. The $\chi_2^2$ distribution is used because there are two additional parameters in the alternate than in the null.

The results of the test are displayed in Fig. 6, where we have extended the analysis and computed $W$ over a longer period of time. Similarly to Fig. 5, care must be taken here when performing such an analysis over time, as a correction must be made for multiple testing, to control the rate of false positives. As a result, we have divided the analysis into 11 non-overlapping windows, as indicated, and reported the likelihood ratio test statistic within each window. Then to control the rate of false positives, rather than rejecting all segments with $p$-values less than .05 (found using the $\chi_2^2$ distribution), a False Discovery Rate (FDR) procedure is applied using the Benjamini-Hochberg procedure [34]. This procedure ranks the $p$-values in ascending order (denoted $p_1, \ldots, p_{11}$) and finds the largest $j$ such that $p_j \leq .05j/11$, and then rejects all segments corresponding to $p_1, \ldots, p_j$. This procedure formally requires data segments to be independent, and while mild correlations do exist, these can only result in positive correlations between the statistics (as $\chi_2^2$ distributions can only be positively correlated). Therefore we may still employ this procedure, but the rejection rate is conservative.

The FDR analysis, which is included in the online code, rejects all but the first two segments, which have the highest associated $p$-values and are before the arrival of the Rayleigh wave. Propriety is rejected within our main analysis window of Figs. 4–6, and also afterwards where there is still some seismic activity (as can be seen in Fig. 1). We can see that the rejection of propriety in favor of our model is most significant at time points where the signal is most eccentric, around 16:40 (*cf.* Fig. 5(a)), which is intuitive as here a circular/proper model is the least appropriate.

## VII. CONCLUSIONS

In this paper we have proposed a widely linear complex autoregressive process of order one. The key novelty of the stochastic process is that impropriety is constructed by relating the process to its conjugate at the previous timestep using a widely linear representation, building on ideas developed in [19] for higher order ARMA processes. Our approach is in contrast to alternative approaches to modeling improper complex autoregressive processes, where only the noise component is improper. Our stochastic process can generate improper structure in the form of elliptical oscillations, which is not possible using alternative order one processes in the literature.

We reduced the specification of our process from seven free parameters to five free parameters, by relating the process to a bivariate elliptical process with interchangeable parameters, and "aligning" the ellipticity of the signal and noise. Reducing to five free parameters has the advantage that parameters of our process are easier to identify and estimate in real-world problems. Furthermore, linking to a bivariate process provided the benefits of using both representations, where we were then able to find conditions for stationarity, and describe the structure of the elliptical oscillations. In general the parameter connections between the representations are non-trivial, but transforming between representations provided useful insight, shedding light on the behavior of the more compact complex widely linear representation.

A promising direction for gaining insight into the full unconstrained seven-parameter widely linear specification, is to relate the process to a bivariate process with two separate elliptical transforms—one each for the autoregressive and noise components in (6)—thus also now being specified by seven parameters. Such an analysis should be performed if the problem is known to have elliptical signal and noise structure that is not aligned.

Another important innovation of this paper is that we have provided time- and frequency-domain techniques to parameter estimation, and then applied them to demonstrate how our widely linear improper process can effectively capture elliptical oscillations in observed seismic traces. An advantage of the complex-valued approach is that frequency-domain understanding then becomes natural, as the asymmetry in the power spectra defines preferred direction of rotation. The process we propose has the potential to be applied to other improper signals in numerous applications, including for example modeling elliptical eddies as documented in [14], or modeling phase-shifted stochastic cycles in econometric time series [35].

## APPENDIX A
### STATIONARITY OF THE BIVARIATE ELLIPTICAL AUTOREGRESSIVE PROCESS

From [36, Ch. 11] we have that for (6) to be stationary we require that the eigenvalues of the matrix

$$aR = \begin{pmatrix} a\cos\theta & -a\sin\theta \\ a\sin\theta & a\cos\theta \end{pmatrix}$$

have modulus less than one. The eigenvalues of this matrix are found to be

$$\lambda_1 = a\cos\theta + ia\sin\theta, \quad \lambda_2 = a\cos\theta - ia\sin\theta.$$

Therefore as $|\lambda_1| = |\lambda_2| = a$, it follows that requiring both $|\lambda_1| < 1$ and $|\lambda_2| < 1$ for stationarity is equivalent to requiring that $a < 1$ (as we already have that $a \geq 0$).

## APPENDIX B
### PROOF OF PROPOSITION 1

Combining (6) and (7) we have the relationship

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = QP\left\{ aR\begin{pmatrix} X'_{t-1} \\ Y'_{t-1} \end{pmatrix} + \sigma_\epsilon \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \right\},$$

and then substituting $(X'_{t-1}\ Y'_{t-1})^T$ for $(X_{t-1}\ Y_{t-1})^T$, and using that $Q^{-1} = Q^T$, it follows that

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = QP\left\{ aRP^{-1}Q^T\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \sigma_\epsilon \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \right\}. \quad (26)$$

To simplify (26) we first define

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad K = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

It then follows that

$$PRP^{-1} = \cos\theta I + \frac{\sin\theta}{2}\left(\frac{1}{\rho^2} + \rho^2\right)J - \frac{\sin\theta}{2}\left(\frac{1}{\rho^2} + \rho^2\right)K.$$

We then use the properties that

$$QIQ^T = I, \quad QJQ^T = J, \quad QKQ^T = \begin{pmatrix} -\sin 2\psi & \cos 2\psi \\ \cos 2\psi & \sin 2\psi \end{pmatrix},$$

which allows for a simple expression for $L = QPRP^{-1}Q^T$ where

$$L = \cos\theta I + \frac{\sin\theta}{2}\left(\frac{1}{\rho^2} + \rho^2\right)J$$
$$- \frac{\sin\theta}{2}\left(\frac{1}{\rho^2} - \rho^2\right)\begin{pmatrix} -\sin 2\psi & \cos 2\psi \\ \cos 2\psi & \sin 2\psi \end{pmatrix}.$$

Therefore (26) simplifies to

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = aL\begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + QP\sigma_\epsilon \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}. \quad (27)$$

To reformulate this in terms of $Z_t = X_t + iY_t$ we use the relationship

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \frac{1}{2}T\begin{pmatrix} Z_t \\ Z_t^* \end{pmatrix}, \quad \text{where} \quad T = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}. \quad (28)$$

Substituting (28) into (27) and using that $T^HT = 2I$, where the subscript $H$ denotes the Hermitian transpose, after rearranging we see that

$$\begin{pmatrix} Z_t \\ Z_t^* \end{pmatrix} = \frac{a}{2}T^H LT\begin{pmatrix} Z_{t-1} \\ Z_{t-1}^* \end{pmatrix} + T^H QP\sigma_\epsilon\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}. \quad (29)$$

Expanding (29) and taking the top row we then get the relationship given in the proposition.

## REFERENCES

[1] J. Gonella, "A rotary-component method for analysing meteorological and oceanographic vector time series," *Deep-Sea Res.*, vol. 19, pp. 833–846, 1972.

[2] D. B. Rowe, "Modeling both the magnitude and phase of complex-valued fMRI data," *NeuroImage*, vol. 25, no. 4, pp. 1310–1324, 2005.

[3] K. S. Miller, *Complex stochastic processes: an introduction to theory and application*. Addison-Wesley, 1974.

[4] P. J. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*. Cambridge University Press, 2010.

[5] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE T. Signal Proces.*, vol. 59, no. 11, pp. 5101–5125, 2011.

[6] A. T. Walden, "Rotary components, random ellipses and polarization: a statistical perspective," *Phil. Trans. R. Soc. A*, vol. 371:20110554, 2013.

[7] S. C. Olhede and A. T. Walden, "Local directional denoising," *IEEE T. Signal Proces.*, vol. 53, pp. 4725–4730, 2005.

[8] ——, "'Analytic' wavelet thresholding," *Biometrika*, vol. 91, no. 4, pp. 955–973, 2004.

[9] P. J. Schreier, L. L. Scharf, and A. Hanssen, "A generalized likelihood ratio test for impropriety of complex signals," *IEEE Signal Proc. Let.*, vol. 13, no. 7, pp. 433–436, 2006.

[10] J. D. Hamilton, *Time series analysis*. Princeton University Press, 1994.

[11] A. Le Breton, "A note on maximum likelihood estimation for the complex-valued first-order autoregressive process," *Stat. Probabil. Lett.*, vol. 7, no. 2, pp. 171–173, 1988.

[12] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE T. Signal Proces.*, vol. 45, no. 2, pp. 411–420, 1997.

[13] J. C. Samson, "Pure states, polarized waves, and principal components in the spectra of multiple, geophysical time-series," *Geophys. J. Int.*, vol. 72, no. 3, pp. 647–664, 1983.

[14] J. M. Lilly and J. C. Gascard, "Wavelet ridge diagnosis of time-varying elliptical signals with application to an oceanic eddy," *Nonlinear Proc. Geoph.*, vol. 13, no. 5, pp. 467–483, 2006.

[15] J. M. Lilly and J. Park, "Multiwavelet spectral and polarization analyses of seismic records," *Geophys. J. Int.*, vol. 122, no. 3, pp. 1001–1021, 1995.

[16] S. C. Olhede and A. T. Walden, "Polarization phase relationships via multiple Morse wavelets. II. Data analysis," *P. Roy. Soc. A-Math. Phy.*, vol. 459, no. 2031, pp. 641–657, 2003.

[17] J. M. Lilly, "Modulated oscillations in three dimensions," *IEEE T. Signal Proces.*, vol. 59, no. 12, pp. 5930–5943, 2011.

[18] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE T. Signal Proces.*, vol. 43, no. 8, pp. 2030–2033, 1995.

[19] J. Navarro-Moreno, "ARMA prediction of widely linear systems by using the innovations algorithm," *IEEE T. Signal Proces.*, vol. 56, no. 7-2, pp. 3061–3068, 2008.

[20] J. Navarro-Moreno, J. Moreno-Kaiser, R. M. Fernández-Alcalá, and J. C. Ruiz-Molina, "Widely linear prediction for transfer function models based on the infinite past," *Comput. Stat. Data An.*, vol. 58, pp. 139–146, 2013.

[21] J. Navarro-Moreno, M. D. Estudillo-Martínez, R. M. Fernández-Alcalá, and J. C. Ruiz-Molina, "Estimation of improper complex-valued random signals in colored noise by using the Hilbert space theory," *IEEE T. Inform. Theory*, vol. 55, no. 6, pp. 2859–2867, 2009.

[22] P. Rubin-Delanchy and A. T. Walden, "Simulation of improper complex-valued sequences," *IEEE T. Signal Proces.*, vol. 55, no. 11, pp. 5517–5521, 2007.

[23] ——, "Kinematics of complex-valued time series," *IEEE T. Signal Proces.*, vol. 56, no. 9, pp. 4189–4198, 2008.

[24] M. Arató, A. N. Kolmogorov, and Y. G. Sinai, "Estimation of the parameters of a complex stationary Gaussian Markov process," *Dokl. Akad. Nauk SSSR*, vol. 146, pp. 747–750, 1962.

[25] H. Wold, *A study in the analysis of stationary time series*. Almqvist & Wiksell, 1954.

[26] P. Whittle, "Estimation and information in stationary time series," *Ark. Mat.*, vol. 2, no. 5, pp. 423–434, 1953.

[27] A. M. Sykulski, S. C. Olhede, J. M. Lilly, and J. J. Early, "Stochastic modeling and estimation of stationary complex-valued signals," *arXiv preprint: 1306.5993*, 2016.

[28] A. M. Sykulski, S. C. Olhede, and J. M. Lilly, "The de-biased Whittle likelihood for second-order stationary stochastic processes," *arXiv preprint: 1605.06718*, 2016.

[29] A. M. Sykulski, S. C. Olhede, J. M. Lilly, and E. Danioux, "Lagrangian time series models for ocean surface drifter trajectories," *J. R. Statist. Soc. C*, vol. 65, no. 1, pp. 29–50, 2016.

[30] A. T. Walden and R. E. White, "Estimating the statistical bandwidth of a time series," *Biometrika*, vol. 77, no. 4, pp. 699–707, 1990.

[31] S. Adak, "Time-dependent spectral analysis of nonstationary time series," *J. Am. Stat. Assoc.*, vol. 93, no. 444, pp. 1488–1501, 1998.

[32] M. Fuentes, "Spectral methods for nonstationary spatial processes," *Biometrika*, vol. 89, no. 1, pp. 197–210, 2002.

[33] J. M. Lilly and S. C. Olhede, "Bivariate instantaneous frequency and bandwidth," *IEEE T. Signal Proces.*, vol. 58, no. 2, pp. 591–603, 2010.

[34] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. R. Statist. Soc. B*, pp. 289–300, 1995.

[35] G. Rünstler, "Modelling phase shifts among stochastic cycles," *Economet. J.*, vol. 7, no. 1, pp. 232–248, 2004.

[36] E. Zivot and J. Wang, *Modeling financial time series with S-Plus*. Springer Science & Business Media, 2007.

**Jonathan M. Lilly** (M'05) was born in Lansing, Michigan, in 1972. He received the B.S. degree in geology and geophysics from Yale University, New Haven, Connecticut, in 1994, and the M.S. and Ph.D. degrees in physical oceanography from the University of Washington (UW), Seattle, Washington, in 1997 and 2002, respectively.

He was a Postdoctoral Researcher with the UW Applied Physics Laboratory and School of Oceanography, from 2002 to 2003, and with the Laboratoire d'Océanographie Dynamique et de Climatologie, Université Pierre et Marie Curie, Paris, France, from 2003 to 2005. From 2005 until 2010, he was a Research Associate with Earth and Space Research in Seattle, Washington. In 2010 he joined NorthWest Research Associates, an employee-owned scientific research corporation in Redmond, Washington, as a Senior Research Scientist. His research interests are oceanic vortex structures, time/frequency analysis methods, Lagrangian observations, and high latitude oceanography.

Dr. Lilly is a member of the American Meteorological Society and of the American Geophysical Union.

**Adam M. Sykulski** (M'16) was born in Southampton, UK, in 1985. He received the M.Sci. and Ph.D. degrees in mathematics from Imperial College London in 2007 and 2011 respectively.

He was a corporate analyst at Lazard, London, from 2007–2008 specializing in corporate mergers and acquisitions. After completing his Ph.D. he was a Postdoctoral Research Associate at the Department of Statistical Science, University College London, from 2011 to 2014. Since 2014, he has been a Research Fellow with a joint position between University College London, and NorthWest Research Associates, Redmond, WA. He is currently funded by a Marie Curie Research Fellowship from the European Commission. His research interests include time series analysis, spatiotemporal data analysis, frequency/spectral domain analysis, and applications in physical applications, with a current focus on the analysis of large-scale oceanographic data.

Dr. Sykulski is a member of the American Geophysical Union and a Fellow of the Royal Statistical Society.

**Sofia C. Olhede** (M'06) was born in Spanga, Sweden, in 1977. She received the M.Sci. and Ph.D. degrees in mathematics from Imperial College London, London, U.K., in 2000 and 2003, respectively. She was a Lecturer from 2002 to 2006 and Senior Lecturer from 2006 to 2007 with the Mathematics Department, Imperial College London, and in 2007, she joined the Department of Statistical Science, University College London, where she is currently a Professor of Statistics, Honorary Professor of Computer Science, and Director of Research of the Department of Statistical Science. Her research interests include the analysis of network data, nonstationary time series, inhomogeneous random fields and multiscale methods.

Prof. Olhede is a member of the Programme Committee of the International Centre for Mathematical Sciences. She is currently co-organizing a six-month Isaac Newton Institute programme on networks. Prof. Olhede is a Fellow of the Royal Statistical Society and a member of the Institute of Mathematical Statistics, the London Mathematical Society, and the Society of Industrial and Applied Mathematics.