Frequency-Domain Stochastic Modeling of Stationary Bivariate or Complex-Valued Signals

Adam M. Sykulski, *Member, IEEE*, Sofia Charlotta Olhede, *Member, IEEE*,
Jonathan M. Lilly, *Senior Member, IEEE*, and Jeffrey J. Early

Abstract—There are three equivalent ways of representing two jointly observed real-valued signals: as a bivariate vector signal, as a single complex-valued signal, or as two analytic signals known as the rotary components. Each representation has unique advantages depending on the system of interest and the application goals. In this paper, we provide a joint framework for all three representations in the context of frequency-domain stochastic modeling. This framework allows us to extend many established statistical procedures for bivariate vector time series to complex-valued and rotary representations. These include procedures for parametrically modeling signal coherence, estimating model parameters using the Whittle likelihood, performing semiparametric modeling, and choosing between classes of nested models using model choice. We also provide a new method of testing for impropriety in complex-valued signals, which tests for noncircular or anisotropic second-order statistical structure when the signal is represented in the complex plane. Finally, we demonstrate the usefulness of our methodology in capturing the anisotropic structure of signals observed from fluid dynamic simulations of turbulence.

Index Terms—Time series analysis, spectral analysis, parametric statistics, stochastic processes, parameter estimation, maximum likelihood estimation.

I. INTRODUCTION

IN MANY applications of signal processing, there is a need to jointly analyze two real-valued signals when they share a common dependence structure. Examples include radio frequency position and displacement measurements of blood-flow [1], and eastward and northward geophysical signals such as wind or ocean current velocities [2]. There are three distinct mathematical ways of representing two real-valued signals: as a bivariate vector signal, as a single complex-valued signal, and as two complex-valued analytic signals known as the rotary components. Each representation has unique advantages depending on the system of interest and the application goals.

To motivate the need for these different representations, in Fig. 1(a) we plot the satellite-tracked position trajectories of a large array of freely-drifting oceanographic instruments obtained from the Global Drifter Program [3]. In Figs. 1(b) and 1(c) we plot a 40-day position trajectory from a North-Atlantic drifter, together with the velocities corresponding to the eastward and northward displacements of this trajectory. In Fig. 1(d) we display a multi-taper spectral density estimate of a complex-valued velocity signal constructed from Fig. 1(c). The spectrum is supported over both negative and positive frequencies, distinguishing oscillatory behavior with a preferred direction of rotation, known commonly as the rotary components [2]. Fig. 1(d) reveals that the signal contains two counter-rotating oscillations at different frequencies. This is not as easily observed in the bivariate time-domain representation of Fig. 1(c), thus motivating the benefits of considering different representations of two jointly observed signals.

In many applications there is a need to specify a simple parametric model for the signal structure, and then to estimate these parameters from a set of observed signals. In this paper we describe a framework for parametrically modeling and estimating the parameters of two real-valued signals in each of the three representations as stationary Gaussian stochastic processes. This builds on ideas found in [4]–[8] for non-parametric or deterministic modeling of complex-valued signals, with further understanding developed in [2], [9], [10] for atmospheric and oceanographic processes.

Approaches to stochastic parametric modeling of complex-valued signals have been primarily focused on the class of autoregressive moving average (ARMA) models using widely-linear filters, see e.g. [11]–[14]. In this paper, we propose more general classes of stochastic parametric models for complex-valued and rotary representations; and we connect these with well-known bivariate modeling techniques such as [15]. The background for mathematically connecting the representations is given in Section II. Then in Section III we propose novel parametric models for specifying either the bivariate coherency or the rotary coherency—two important quantities which are shown to be distinct, thus providing new and easily interpreted model structures.
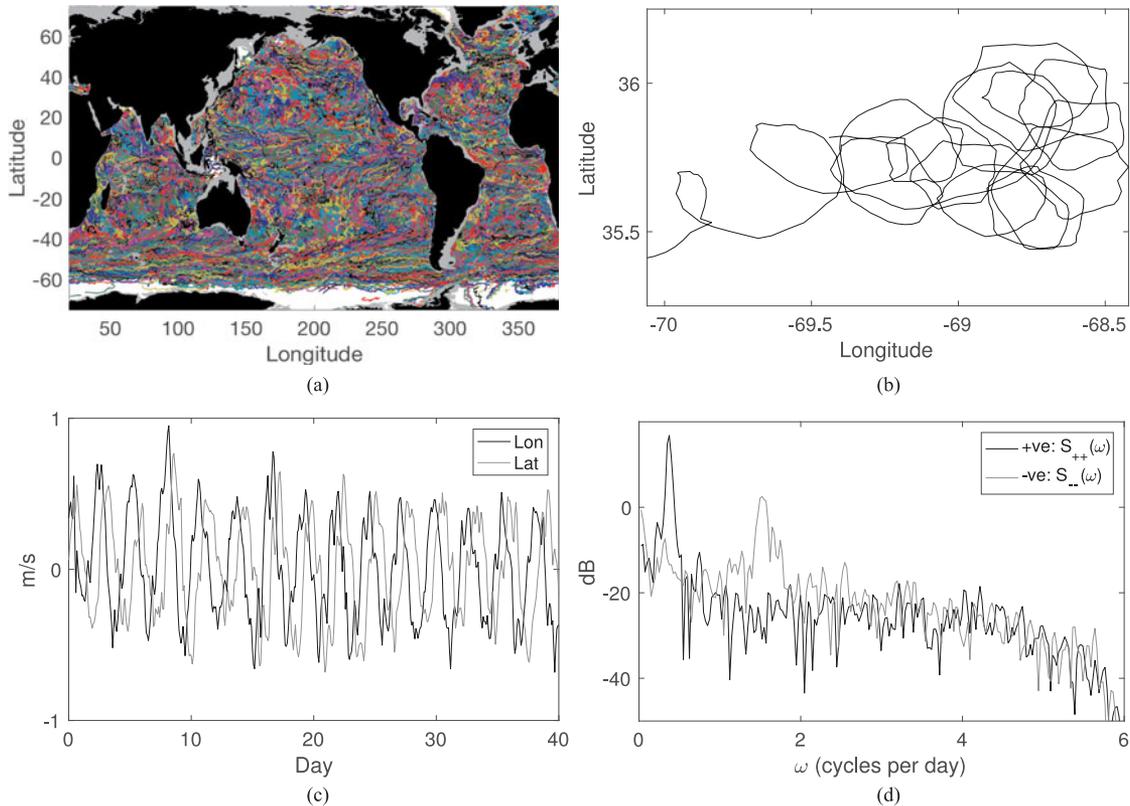
Fig. 1. (a) satellite-tracked position trajectories from multiple freely-drifting instruments from the Global Drifter Program (www.aoml.noaa.gov/phod/dac). (b) a 40-day position trajectory of North Atlantic drifter ID#44000. (c) velocities of this drifter over time in each Cartesian direction. (d) a multi-taper spectral density estimate of the rotary components, using the discrete prolate spheroidal sequence (dpss) tapers with bandwidth parameter 3, with negative and positive frequencies overlaid.

Our joint framework allows well-known procedures for bivariate modeling to be extended to complex-valued and rotary representations. First in Section IV we provide computationally efficient procedures for parameter estimation, by extending the Whittle likelihood from the bivariate to complex-valued and rotary representations. Then in Section V-A we demonstrate how the Whittle likelihood can naturally be used with semi-parametric modeling techniques (relating this to the seminal work of [16]). In Section V-B we construct a parametric test for impropriety, which also tests for anisotropy if the signals are tracking spatial positions like the ocean drifters. In Section V-C, we detail how model choice is correctly performed with complex-valued and rotary signals when selecting between nested models. Finally, in Section VI we demonstrate the usefulness of our methodology in developing key physical understanding, by applying our methods to signals obtained from fluid dynamic models of turbulence.

II. BACKGROUND

A bivariate pair of real-valued signals can be given two other equivalent representations: as a complex-valued signal, or as a pair of analytic signals which we call the rotary components. The information contained in each of these three representations—bivariate/complex/rotary—is equivalent, but the operations required to transform between them are in gen-

eral nontrivial. In this section we provide a brief background with necessary definitions and notation, culminating with Table I which links the various spectral representations. Much of the material in this section can be found in [4]–[8].

A. Bivariate Processes

Consider a zero-mean continuous-time bivariate Gaussian process, denoted at time t by $[X(t)\ Y(t)]^T$, where "T" denotes matrix transpose. Using the Cramér spectral representation theorem, we define this process in terms of the orthogonal increments processes $d\Psi_X(\omega)$ and $d\Psi_Y(\omega)$ such that

$$\begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} \equiv \frac{1}{2\pi} \int \begin{bmatrix} d\Psi_X(\omega) \\ d\Psi_Y(\omega) \end{bmatrix} e^{i\omega t}, \quad (1)$$

where $i \equiv \sqrt{-1}$. The statistics of the bivariate process can also be fully specified by the power spectral matrix $\boldsymbol{S}_B(\omega)$ defined by

$$\boldsymbol{S}_B(\omega)\delta(\omega - \nu)d\omega d\nu$$
$$\equiv \frac{1}{2\pi} \mathrm{E} \left\{ \begin{bmatrix} d\Psi_X(\omega) \\ d\Psi_Y(\omega) \end{bmatrix} \begin{bmatrix} d\Psi_X^*(\nu) & d\Psi_Y^*(\nu) \end{bmatrix} \right\}, \quad (2)$$

TABLE I
RELATIONSHIPS BETWEEN SPECTRA OF THREE DIFFERENT REPRESENTATIONS OF TWO JOINTLY OBSERVED REAL-VALUED SIGNALS: BIVARIATE (LEFT COLUMN), COMPLEX-VALUED (MIDDLE COLUMN), AND ROTARY (RIGHT COLUMN)

		Bivariate (X/Y)	Complex (Z/Z^*)	Rotary (Z^+/Z^-) $(\omega \neq 0)$
Bivariate	$S_{XX}(\omega)$		$\frac{1}{4}[S_{ZZ}(\omega) + S_{ZZ}(-\omega)]$ $+\frac{1}{2}\Re\{R_{ZZ}(\omega)\}$	$\frac{1}{4}[S_{++}(|\omega|) + S_{--}(|\omega|)]$ $+\frac{1}{2}\Re\{S_{+-}(|\omega|)\}$
	$S_{YY}(\omega)$		$\frac{1}{4}[S_{ZZ}(\omega) + S_{ZZ}(-\omega)]$ $-\frac{1}{2}\Re\{R_{ZZ}(\omega)\}$	$\frac{1}{4}[S_{++}(|\omega|) + S_{--}(|\omega|)]$ $-\frac{1}{2}\Re\{S_{+-}(|\omega|)\}$
	$S_{XY}(\omega)$		$\frac{1}{2}\Im\{R_{ZZ}(\omega)\}$ $+\frac{1}{4}\mathrm{i}[S_{ZZ}(\omega) - S_{ZZ}(-\omega)]$	$\frac{1}{2}\Im\{S_{+-}(|\omega|)\}$ $+\frac{\mathrm{i}}{4}\mathrm{sgn}(\omega)[S_{++}(|\omega|) - S_{--}(|\omega|)]$
Complex		$S_{XX}(\omega) + S_{YY}(\omega)$ $+2\Im\{S_{XY}(\omega)\}$	$S_{ZZ}(\omega)$	$S_{++}(\omega) + S_{--}(-\omega)$
		$S_{XX}(\omega) - S_{YY}(\omega)$ $+2\mathrm{i}\Re\{S_{XY}(\omega)\}$	$R_{ZZ}(\omega)$	$S_{+-}(\omega) + S_{+-}(-\omega)$
Rotary $(\omega > 0)$		$S_{XX}(\omega) + S_{YY}(\omega)$ $+2\Im\{S_{XY}(\omega)\}$	$S_{ZZ}(\omega)$	$S_{++}(\omega)$
		$S_{XX}(\omega) + S_{YY}(\omega)$ $-2\Im\{S_{XY}(\omega)\}$	$S_{ZZ}(-\omega)$	$S_{--}(\omega)$
		$S_{XX}(\omega) - S_{YY}(\omega)$ $+2\mathrm{i}\Re\{S_{XY}(\omega)\}$	$R_{ZZ}(\omega)$	$S_{+-}(\omega)$

The three elements in a given row are equal for the specified range of ω, for example the last row reads: $S_{XX}(\omega) - S_{YY}(\omega)$ $+2\mathrm{i}\Re\{S_{XY}(\omega)\} = R_{ZZ}(\omega) = S_{+-}(\omega)$ for $\omega > 0$, where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ refer to the real and imaginary parts respectively.

where $\delta(\cdot)$ is the Dirac-delta function and $*$ denotes the complex conjugate. We denote the elements of $\boldsymbol{S}_B(\omega)$ by

$$\boldsymbol{S}_B(\omega) = \begin{bmatrix} S_{XX}(\omega) & S_{XY}(\omega) \\ S_{XY}^*(\omega) & S_{YY}(\omega) \end{bmatrix}. \tag{3}$$

The power spectral densities $S_{XX}(\omega)$ and $S_{YY}(\omega)$ are real-valued, nonnegative, and symmetric in ω, whereas the cross-spectral density $S_{XY}(\omega)$ is complex-valued and Hermitian symmetric. Finally, the statistics of the bivariate process can be fully specified by the covariance matrix $\boldsymbol{R}_B(\tau)$ defined by

$$\boldsymbol{R}_B(\tau) \equiv \mathrm{E}\left\{ \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} \begin{bmatrix} X(t-\tau) & Y(t-\tau) \end{bmatrix} \right\}, \tag{4}$$

where the elements of the matrix $\boldsymbol{R}_B(\tau)$ are denoted by

$$\boldsymbol{R}_B(\tau) = \begin{bmatrix} s_{XX}(\tau) & s_{XY}(\tau) \\ s_{XY}^*(\tau) & s_{YY}(\tau) \end{bmatrix}. \tag{5}$$

From these definitions it follows that

$$\boldsymbol{R}_B(\tau) =$$

$$\frac{1}{4\pi^2}\iint e^{\mathrm{i}\omega t} e^{-\mathrm{i}\nu(t-\tau)} \mathrm{E}\left\{ \begin{bmatrix} d\Psi_X(\omega) \\ d\Psi_Y(\omega) \end{bmatrix} \begin{bmatrix} d\Psi_X^*(\nu) & d\Psi_Y^*(\nu) \end{bmatrix} \right\}$$

$$= \frac{1}{2\pi}\int \boldsymbol{S}_B(\omega) e^{\mathrm{i}\omega\tau} d\omega, \tag{6}$$

such that the covariance matrix $\boldsymbol{R}_B(\tau)$ forms a Fourier pair with the power spectral matrix $\boldsymbol{S}_B(\omega)$.

B. Complex-Valued Processes

The bivariate real-valued process $[X(t)\ Y(t)]^T$ can alternatively be expressed as a single complex-valued signal defined by

$$Z(t) \equiv X(t) + \mathrm{i}Y(t). \tag{7}$$

In this case the second-order statistics of $[X(t)\ Y(t)]^T$ are equivalent to the statistics of $Z(t)$ together with its complex conjugate $Z^*(t)$, as we now show; see also [5], [6].

The process $Z(t)$ can be expressed in terms of the orthogonal increments process $d\Psi_Z(\omega) \equiv d\Psi_X(\omega) + \mathrm{i}d\Psi_Y(\omega)$ such that

$$Z(t) = \frac{1}{2\pi}\int d\Psi_Z(\omega)e^{\mathrm{i}\omega t}. \tag{8}$$

The statistics of the complex-valued process $Z(t)$ can also be fully specified from the power spectral matrix $\boldsymbol{S}_C(\omega)$ defined by

$$\boldsymbol{S}_C(\omega)\delta(\omega - \nu)d\omega d\nu$$

$$\equiv \frac{1}{2\pi}\mathrm{E}\left\{ \begin{bmatrix} d\Psi_Z(\omega) \\ d\Psi_Z^*(-\omega) \end{bmatrix} \begin{bmatrix} d\Psi_Z^*(\nu) & d\Psi_Z(-\nu) \end{bmatrix} \right\}, \tag{9}$$

where the elements of the matrix $\boldsymbol{S}_C(\omega)$ are denoted by

$$\boldsymbol{S}_C(\omega) = \begin{bmatrix} S_{ZZ}(\omega) & R_{ZZ}(\omega) \\ R_{ZZ}^*(\omega) & S_{ZZ}(-\omega) \end{bmatrix}. \tag{10}$$

The power spectral density $S_{ZZ}(\omega)$ is real-valued and nonnegative, but not necessarily symmetric in ω, whereas the complementary spectrum[1] $R_{ZZ}(\omega)$ is in general complex-valued and is always symmetric in ω, i.e. $R_{ZZ}(\omega) = R_{ZZ}(-\omega)$.

The statistics of the complex-valued process $Z(t)$ can also be fully specified by the covariance matrix $\boldsymbol{R}_C(\tau)$ defined by

$$\boldsymbol{R}_C(\tau) \equiv \mathrm{E}\left\{ \begin{bmatrix} Z(t) \\ Z^*(t) \end{bmatrix} \begin{bmatrix} Z^*(t-\tau) & Z(t-\tau) \end{bmatrix} \right\}, \tag{11}$$

---

[1]We clarify a distinction in our notation whereby $R(\omega)$ refers to a complementary spectrum and $\boldsymbol{R}(\tau)$ refers to a covariance matrix. This is chosen as such to be consistent with the conventional notation in the literature.

where

$$\boldsymbol{R}_C(\tau) = \begin{bmatrix} s_{ZZ}(\tau) & r_{ZZ}(\tau) \\ r^*_{ZZ}(\tau) & s_{ZZ}(-\tau) \end{bmatrix}. \tag{12}$$

Similarly to (6) it can be shown that $\boldsymbol{R}_C(\tau)$ and $\boldsymbol{S}_C(\omega)$ form a Fourier pair. If the off-diagonal term of $\boldsymbol{R}_C(\tau)$ vanishes (that is $r_{ZZ}(\tau) = 0$ for all τ), or equivalently if the off-diagonal term of $\boldsymbol{S}_C(\omega)$ vanishes (that is $R_{ZZ}(\omega) = 0$ for all ω), then the process is said to be *proper* or *circular*, and otherwise it is *improper* or *noncircular* [6].

To link the bivariate and complex representations, we express $\boldsymbol{S}_C(\omega)$ in terms of $\boldsymbol{S}_B(\omega)$ in (3). To do this we first define the unitary matrix

$$\boldsymbol{T} \equiv \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & \mathrm{i} \\ 1 & -\mathrm{i} \end{bmatrix}, \tag{13}$$

which has the property that $\boldsymbol{T}\boldsymbol{T}^{\mathrm{H}} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the 2×2 identity matrix and H denotes the Hermitian transpose. It follows that

$$\begin{bmatrix} Z(t) \\ Z^*(t) \end{bmatrix} = \sqrt{2}\boldsymbol{T} \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix}. \tag{14}$$

Then from (9), and using the property $d\Psi_{Z^*}(\omega) = d\Psi_Z^*(-\omega)$, where $d\Psi_{Z^*}(\omega)$ is defined from (8) as the orthogonal increments of $Z^*(t)$, we have that

$$\boldsymbol{S}_C(\omega)\delta(\omega-\nu)d\omega d\nu$$
$$= \frac{1}{2\pi} \mathrm{E} \left\{ \begin{bmatrix} d\Psi_Z(\omega) \\ d\Psi_{Z^*}(\omega) \end{bmatrix} \begin{bmatrix} d\Psi_Z^*(\nu) & d\Psi_{Z^*}^*(\nu) \end{bmatrix} \right\}$$
$$= \frac{1}{2\pi} \mathrm{E} \left\{ \sqrt{2}\boldsymbol{T} \begin{bmatrix} d\Psi_X(\omega) \\ d\Psi_Y(\omega) \end{bmatrix} \begin{bmatrix} d\Psi_X^*(\nu) & d\Psi_Y^*(\nu) \end{bmatrix} \sqrt{2}\boldsymbol{T}^{\mathrm{H}} \right\}, \tag{15}$$

such that from (2) we can observe that

$$\boldsymbol{S}_C(\omega) = 2\boldsymbol{T}\boldsymbol{S}_B(\omega)\boldsymbol{T}^{\mathrm{H}}. \tag{16}$$

If we expand the matrices in (16) then we arrive at direct transformations between $\{S_{XX}(\omega), S_{YY}(\omega), S_{XY}(\omega)\}$ and $\{S_{ZZ}(\omega), R_{ZZ}(\omega)\}$, which are displayed as part of Table I—a table which provides transformations between the power spectra for the bivariate, complex, and rotary representations.

C. Rotary Components

A third representation we consider is to split the complex-valued signal into analytic and anti-analytic components, as is commonly performed in signal processing [7], [17], [18]. These components are also known as the rotary components [2], [8], and we define these by

$$\begin{bmatrix} Z^+(t) \\ Z^-(t) \end{bmatrix} \equiv \frac{1}{2\pi} \int \begin{bmatrix} d\Psi_{Z^+}(\omega) \\ d\Psi_{Z^-}(\omega) \end{bmatrix} e^{\mathrm{i}\omega t}, \tag{17}$$

where the orthogonal increments are given by

$$\begin{bmatrix} d\Psi_{Z^+}(\omega) \\ d\Psi_{Z^-}(\omega) \end{bmatrix} \equiv U(\omega) \begin{bmatrix} d\Psi_Z(\omega) \\ d\Psi_Z^*(-\omega) \end{bmatrix}. \tag{18}$$

Here $U(\omega)$ is the unit (or Heaviside) step function defined by

$$U(\omega) \equiv \begin{cases} 1 & \omega > 0 \\ 1/2 & \omega = 0 \\ 0 & \omega < 0. \end{cases} \tag{19}$$

It follows from these definitions that

$$Z(t) = [Z^-(t)]^* + Z^+(t). \tag{20}$$

As in [7], we construct the processes $Z^+(t)$ and $Z^-(t)$ such that they are both *analytic*, that is, supported only on positive frequencies. The anti-analytic component of $Z(t)$ is recovered by taking the conjugate of $Z^-(t)$. Therefore $Z^+(t)$ and $[Z^-(t)]^*$ are the analytic and anti-analytic signals associated with $Z(t)$, while $Z^-(t)$ is referred to as the conjugate analytic signal. The statistics of the process can also be specified in terms of the power spectral matrix $\boldsymbol{S}_\pm(\omega)$ where

$$\boldsymbol{S}_\pm(\omega)\delta(\omega-\nu)d\omega d\nu$$
$$\equiv \frac{1}{2\pi} \mathrm{E} \left\{ \begin{bmatrix} d\Psi_{Z^+}(\omega) \\ d\Psi_{Z^-}(\omega) \end{bmatrix} \begin{bmatrix} d\Psi_{Z^+}^*(\nu) & d\Psi_{Z^-}^*(\nu) \end{bmatrix} \right\}, \tag{21}$$

where the elements of the matrix $\boldsymbol{S}_\pm(\omega)$ are denoted by

$$\boldsymbol{S}_\pm(\omega) = \begin{bmatrix} S_{++}(\omega) & S_{+-}(\omega) \\ S_{+-}^*(\omega) & S_{--}(\omega) \end{bmatrix}. \tag{22}$$

It then follows by substituting (18) into (21) that

$$\boldsymbol{S}_\pm(\omega) = U^2(\omega)\boldsymbol{S}_C(\omega). \tag{23}$$

This then provides a direct mapping between $\{S_{++}(\omega), S_{--}(\omega), S_{+-}(\omega)\}$ and $\{S_{ZZ}(\omega), R_{ZZ}(\omega)\}$. Furthermore, by combining (16) and (23) we also arrive at transformations between $\{S_{XX}(\omega), S_{YY}(\omega), S_{XY}(\omega)\}$ and $\{S_{++}(\omega), S_{--}(\omega), S_{+-}(\omega)\}$. These are provided in Table I for $\omega > 0$. Our choice to make $Z^-(t)$ analytic simplifies the representations, as the rotary components are defined only for nonnegative frequencies, such that all the spectral densities in $\boldsymbol{S}_\pm(\omega)$ are zero for negative frequencies. The relationships between the bivariate and complex/rotary representations are nontrivial, and highlight some of the differences between working with bivariate and complex-valued data. We will refer to Table I in future sections to provide more intuition and simplify calculations.

III. STOCHASTIC MODELING OF COHERENCY

We now introduce a framework for the specification of stochastic models for pairs of real-valued signals. The approach we shall adopt is that first a representation of bivariate/complex/rotary is selected, and then the main diagonal of the corresponding spectral matrix $\boldsymbol{S}(\omega)$ in either (3), (10), or (22) is specified. Then the respective off-diagonal term must also be specified, which becomes equivalent to modeling the *coherency* of the signal. First, in Section III-A, we propose models for the *bivariate coherency* in the bivariate representation; then, in Section III-B, we propose models for the *rotary coherency* in the rotary representation. Note that we do not have a subsection for modeling in the complex-valued representation, as equivalent

models can be easily constructed from the rotary representation models and using the transformations of Table I.

A. Bivariate Coherency

For bivariate processes, we refer to the off-diagonal entry $S_{XY}(\omega)$ of $\boldsymbol{S}_B(\omega)$ in (3) as the bivariate cross-spectrum, and represent this object as a product of terms

$$S_{XY}(\omega) = \rho_{XY}(\omega) \left[S_{XX}(\omega) S_{YY}(\omega) \right]^{1/2}$$

$$\rho_{XY}(\omega) = \sigma_{XY}(\omega) e^{-i\theta_{XY}(\omega)}, \qquad (24)$$

where $|\rho_{XY}(\omega)| \leq 1$ is the *coherency* of $X(t)$ and $Y(t)$, $\sigma_{XY}(\omega)$ is the *coherence*, and $\theta_{XY}(\omega)$ is the *group delay*, quantifying whether $X(t)$ or $Y(t)$ are leading or lagging in time at frequency-cycle ω. We refer to $\rho_{XY}(\omega)$ as the *bivariate coherency*. Note that $\rho_{XY}(\omega)$ is in general complex-valued, whereas the coherence $\sigma_{XY}(\omega)$ and group delay $\theta_{XY}(\omega)$ are real-valued. The coherence and group delay must satisfy

$$\sigma_{XY}(\omega) = \sigma_{XY}(-\omega), \quad |\sigma_{XY}(\omega)| \leq 1$$

$$\theta_{XY}(-\omega) = -\theta_{XY}(\omega), \qquad (25)$$

on account of the required Hermitian symmetry of $S_{XY}(\omega)$. This follows because while its Fourier pair $s_{XY}(\tau)$ does not satisfy the evenness of $s_{XX}(\tau)$ or $s_{YY}(\tau)$, it is still real-valued.

From a modeling standpoint, see for example [15], we may also regard (24) as *specifying* the bivariate cross-spectrum for a given choice of the individual spectra $S_{XX}(\omega)$ and $S_{YY}(\omega)$. The proposed parametric model for $X(t)$ and $Y(t)$ is a valid Gaussian process if:

1) $S_{XX}(\omega) \geq 0$ and $S_{YY}(\omega) \geq 0$ for all ω,
2) $|\sigma_{XY}(\omega)| \leq 1$ for all ω, such that 1) and 2) together ensure the determinant of the spectral matrix $\boldsymbol{S}_B(\omega)$ defined in (3) is nonnegative for all ω (such that the spectral matrix itself is nonnegative definite),
3) $\sigma_{XY}(\omega)$ is an even function in ω and $\theta_{XY}(\omega)$ is an odd function in ω (such that $S_{XY}(\omega)$ is Hermitian symmetric), and
4) the spectral matrix $\boldsymbol{S}_B(\omega)$ is integrable.

Given the large flexibility for specifying $\sigma_{XY}(\omega)$ and $\theta_{XY}(\omega)$, we propose some practical forms. In essence the parameter $\sigma_{XY}(\omega)$ is providing the magnitude of correlation, and $\theta_{XY}(\omega)$ is strongly linked to time shifts or misalignments between $X(t)$ and $Y(t)$. If the system is dispersive [15] we would expect $\theta_{XY}(\omega)$ to change non-linearly in form across frequencies; if the system is non-dispersive, $\theta_{XY}(\omega)$ would be linear and the coefficient of the linear term would quantify a time-delay.

Many signals are correlated at low frequencies, but become less so at higher frequencies where erratic variability is found. We therefore would commonly find $\sigma_{XY}(\omega)$ to be a decaying function bounded above by unity. Three types of decay might be expected:

1) compactly supported $\sigma_{XY}(\omega)$ showing no correlation magnitude above frequency ω_0,
2) exponentially decaying $\sigma_{XY}(\omega)$ showing a rapid decay,
3) $\sigma_{XY}(\omega)$ exhibiting slow polynomial decay.

A realistic parametric model for the first case, that is compactly supported coherency, would correspond to

$$\sigma_{XY}(\omega) = \begin{cases} \left| \sum_{j=0}^{J} a_j \omega^{2j} \right|, & 0 \leq |\omega| \leq \omega_0, \\ 0, & \omega_0 < |\omega|. \end{cases} \qquad (26)$$

This form is straightforward and interpretable; although we must ensure that the a_j coefficients are selected such that $\sigma_{XY}(\omega)$ is bounded above by unity for all $|\omega| \leq \omega_0$, and that $\sigma_{XY}(\omega_0) = 0$ to achieve continuity in frequency. This model proposes no coherence for frequencies higher than ω_0, and so this parameter emerges as specifying a natural timescale of the bivariate process.

In the second case, where we allow ω to go unbounded. A convenient form for a decaying $\sigma_{XY}(\omega)$ is the logistic function given by

$$\sigma_{XY}(\omega) = \frac{\sigma_0 \left[1 + e^{q(0)} \right]}{1 + e^{q(\omega)}},$$

$$q(\omega) = q_0 + q_1 \omega^2 + \dots q_r \omega^{2r}. \qquad (27)$$

This model guarantees that $\sigma_{XY}(\omega)$ is an even function as required in (25). We require that the highest-order polynomial coefficient is positive, i.e. $q_r > 0$, to ensure $\lim_{\omega \to \infty} \sigma_{XY}(\pm\omega) = 0$, and the q_i coefficients must be selected such that $|\sigma_{XY}(\omega)| \leq 1$ for all ω.

Finally, to yield the third case where the bivariate coherence decays polynomially we propose

$$\sigma_{XY}(\omega) = \frac{\sigma_0}{\left[(\omega - \omega_0)^2 + b^2 \right]^{\alpha/2}}, \quad 0 < \sigma_0 < \left(\omega_0^2 + b^2 \right)^{\alpha/2}. \qquad (28)$$

This form allows the smoothness of the cross-spectrum to be directly manipulated by varying α, and is related to spatial modeling using the Matérn covariance kernel [19].

Hermitian symmetry dictates that the group delay $\theta_{XY}(\omega)$ is required to be an odd function, and we propose the parametric form

$$\theta_{XY}(\omega) = \theta_1 \omega + \theta_2 \omega^3 + \dots + \theta_l \omega^{2l-1}. \qquad (29)$$

Note that while a constant phase is not valid across all frequencies, we may have a phase that is locally constant within particular frequency bands (see [15, p.137]), and where the polynomial form of (29) may simply be considered a non-parametric approximation to the group delay. A function that depends on higher terms than linear are exhibiting a non-linear frequency-dependent time delay, a characteristic of a wave passing through a dispersive medium [15].

Next we ask what constraint is imposed on the bivariate coherence and group delay if we require the signal to be proper, such that $R_{ZZ}(\omega) = 0$ in (10). It follows from Table I that we require $S_{XX}(\omega) = S_{YY}(\omega)$ and $\Re\{S_{XY}(\omega)\} = 0$ for the signal to be proper. For the second condition, by inspecting (24) and (25), we see that we require that *at each* positive frequency $|\omega|$ where $S_{XX}(|\omega|) = S_{YY}(|\omega|) > 0$, either that $\sigma_{XY}(|\omega|) = 0$ or $\theta_{XY}(|\omega|) = \pm\pi/2$ and $\theta_{XY}(-|\omega|) = \mp\pi/2$, such that X and Y are strictly "out of phase." The latter possibility follows

from the fact that a zero-mean Gaussian signal is proper if and only if it is *circular* (i.e. a signal whose properties are invariant under rotation), see [6]. An example of such would be a deterministic signal which follows an exact circle, which corresponds to $\sigma_{XY}(\omega) = 1$ and $\theta_{XY}(\omega) = \pm \text{sgn}(\omega)\pi/2$.

B. Rotary Coherency

We can alternatively specify a stochastic model for the coherency between rotary components $Z^+(t)$ and $Z^-(t)$. As we have chosen the two processes to be analytic, and thus supported only on positive frequencies, we only need to model the covariance between $d\Psi_{Z^+}(\omega)$ and $d\Psi_{Z^-}(\omega)$ for $\omega > 0$. This is equivalent to the covariance between the increment processes of $d\Psi_Z(\omega)$ and $d\Psi_Z^*(-\omega)$, as can be seen from contrasting (9) with (21). Therefore in addition to specifying the variance of $d\Psi_{Z^+}(\omega)$ and $d\Psi_{Z^-}(\omega)$, corresponding to specifying the rotary spectra $S_{++}(\omega)$ and $S_{--}(\omega)$, we only need to model their covariance at the same frequency, which corresponds to the rotary cross-spectrum given by

$$S_{+-}(\omega) = \rho_{\pm}(\omega)\left[S_{++}(\omega)S_{--}(\omega)\right]^{1/2},$$

$$\rho_{\pm}(\omega) = \sigma_{\pm}(\omega)e^{-i\theta_{\pm}(\omega)}, \qquad (30)$$

where $|\rho_{\pm}(\omega)| \leq 1$ is the *rotary coherency* of Z^+ and Z^-, $\sigma_{\pm}(\omega)$ is the rotary coherence, and $\theta_{\pm}(\omega)$ is the rotary group delay.

As in the bivariate case, one may construct models for the rotary coherence $\sigma_{\pm}(\omega)$ and rotary group delay $\theta_{\pm}(\omega)$ analogous to (26)–(29), with the difference that these functions now only need to be supported over positive frequencies. This conveniently means that we are not constrained to require $\sigma_{\pm}(\omega)$ and $\theta_{\pm}(\omega)$ to be even and odd functions respectively. For the proposed model for $Z^+(t)$ and $Z^-(t)$ to be a valid Gaussian process, we now only require the three conditions:

1) $S_{++}(\omega) \geq 0$ and $S_{--}(\omega) \geq 0$ for all ω,
2) $|\sigma_{\pm}(\omega)| \leq 1$ for all ω, such that 1) and 2) together ensure the determinant of the spectral matrix $\boldsymbol{S}_{\pm}(\omega)$ in (22) is nonnegative for all ω (such that the spectral matrix itself is nonnegative definite), and
3) the spectral matrix $\boldsymbol{S}_{\pm}(\omega)$ is integrable.

By inspecting Table I, we observe a useful property. If $X(t)$ and $Y(t)$ are known to be uncorrelated with each other, we see that $S_{+-}(\omega)$ is real-valued for all ω. This means that in such cases we only need to specify a real-valued model for $\rho_{\pm}(\omega)$. In such instances, the simplest positive and compactly supported frequency-dependent model for the rotary coherency (related to the model of (26)) is

$$\rho_{\pm}(\omega) = \begin{cases} |a_0 - a_1\omega|, & 0 < \omega \leq \omega_0, \quad \frac{a_0}{a_1} = \omega_0, \\ 0, & \omega_0 < \omega. \end{cases} \qquad (31)$$

We use this simple rotary coherency model in our oceanographic example in Section VI. If more complex forms are required for the rotary coherency, we can replace (26) with an expansion using both even and odd terms. Similarly (29) can also be replaced by an expansion using both even and odd terms, already advocated as an approximation by [15].

The magnitude of the coherency specified by (31) will have implications for representing the complex-valued signal as an

ellipse [8], [20] and therefore has a geometric interpretation. The ellipse defined between the axis of $dZ(\omega)$ and $dZ^*(-\omega)$ will represent a line if $\rho_{\pm}(\omega) = 1$, and a circle if $\rho_{\pm}(\omega) = 0$ (yielding propriety of Z_t). The intrinsic geometry of the complex-valued signal (e.g. the ellipse, circle or rectilinear motion mapped out by the trajectory) can therefore be seen as specifying $\sigma_{\pm}(\omega)$, while temporal shifts and misalignments are encoded by $\theta_{\pm}(\omega)$.

By inspecting Table I we see that a proper complex-valued signal has no rotary coherency, i.e. $S_{+-}(\omega) = 0$, which is a useful definition of propriety. This can be exploited to formulate a test for impropriety, as is detailed later in Section V-B. Furthermore, for spatially driven processes such as oceanographic signals obtained from drifters, propriety is found to be related to the condition of spatial isotropy, with impropriety in turn implying anisotropy. Thus by testing for impropriety in the time series signal, we can test for anisotropy in the spatial process that generates the sampled signal.

Modeling in bivariate or rotary components is of course equivalent, but as can be seen from Table I, a parametric model represented in one decomposition will not in general be interpretable and compactly represented in the other—hence the need to specify classes of models in both decompositions.

IV. PARAMETER ESTIMATION

In this section we turn to the problem of estimating the parameters of a chosen model from observations. Maximum likelihood is a standard approach but is computationally slow requiring calculation of the determinant and inverse of the time-domain covariance matrix. Instead, [21] proposed using what later became known as the Whittle likelihood, which approximates the time-domain log-likelihood in the frequency domain in $\mathcal{O}(N \log N)$ operations, where N is the length of the observed signal. In this section, we derive the correct form of the Whittle likelihood in each of the bivariate/complex/rotary representations, such that it can be easily implemented for any of the model specifications discussed in Section III.

A. Maximum Likelihood for Bivariate Processes

Henceforth we let X and Y denote length N samples of the corresponding stochastic processes arranged as column vectors. To find the log-likelihood appropriate for the bivariate process consisting of X and Y we concatenate the two time domain samples into a single vector: $B^T = [X^T \ Y^T]$. The theoretical $2N \times 2N$ covariance matrix for this vector under the assumed model is denoted by $\boldsymbol{C}_B(\boldsymbol{\theta}) = \text{E}\left\{BB^T\right\}$. The log-likelihood of the vector B can then be written as

$$\ell(\boldsymbol{\theta}) \stackrel{C}{=} -\frac{1}{2}\log|\boldsymbol{C}_B(\boldsymbol{\theta})| - \frac{1}{2}B^T\boldsymbol{C}_B^{-1}(\boldsymbol{\theta})B. \qquad (32)$$

The superscript "-1" is the matrix inverse, $|\boldsymbol{X}|$ denotes the determinant of matrix $\boldsymbol{X}$, and $\stackrel{C}{=}$ denotes equality up to an additive constant which can be ignored as we are optimizing the objective function. The best choice of parameter-vector $\boldsymbol{\theta}$ for our chosen model to characterize the observed data is found by

maximizing the log-likelihood function

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}),$$

where Θ denotes the parameter space of $\boldsymbol{\theta}$. Maximum likelihood estimation (for Gaussian processes) is generally asymptotically efficient, and a well-behaved procedure in the time series setting [22, p. 27].

B. The Whittle Likelihood for Bivariate Processes

Computation of the matrix inverse and determinant in (32) is computationally expensive, with complexity typically scaling as $\mathcal{O}(N^2)$ for stationary regularly-sampled processes (and $\mathcal{O}(N^3)$ more generally), where N is the length of the signal. A standard technique to approximating (32) is the Whittle likelihood [23]. This estimation technique approximates the time-domain log-likelihood function in the frequency domain, which results in improved $\mathcal{O}(N\log N)$ computational complexity. For bivariate signals, we first define the bivariate Discrete Fourier Transform (DFT) vector as

$$J_B(\omega) = \begin{bmatrix} J_X(\omega) \\ J_Y(\omega) \end{bmatrix} = \sqrt{\frac{\Delta}{N}} \sum_{t=1}^{N} \begin{bmatrix} X_t \\ Y_t \end{bmatrix} e^{-i\omega t\Delta},$$

where $\Delta > 0$ denotes the length of the sampling interval. The standard bivariate Whittle likelihood, once discretized, is given in [21] by

$$\ell_W(\boldsymbol{\theta})$$
$$\overset{C}{=} -\frac{1}{2} \sum_{\omega\in\Omega} \left[\log|\boldsymbol{S}_B(\omega;\boldsymbol{\theta})| + J_B^{\mathrm{H}}(\omega)\boldsymbol{S}_B^{-1}(\omega;\boldsymbol{\theta})J_B(\omega) \right],$$
(33)

where the subscript "W" stands for "Whittle," $\boldsymbol{S}_B(\omega)$ is as defined in (3), and Ω is the set of discrete Fourier frequencies: $\frac{2\pi}{N\Delta}(-\lceil N/2 \rceil + 1, \ldots, -1, 0, 1, \ldots, \lfloor N/2 \rfloor)$. The optimal parameter choice for the Whittle likelihood is made by maximizing

$$\widehat{\boldsymbol{\theta}}^{(W)} = \arg\max_{\boldsymbol{\theta}\in\Theta} \ell_W(\boldsymbol{\theta}).$$

This procedure is $O(N\log N)$ because $J_B(\omega)$ can be computed using a Fast Fourier Transform (FFT), the summation in (33) is over $\mathcal{O}(N)$ frequencies, and the matrix inverse of $S_B(\omega;\boldsymbol{\theta})$ now involves a 2×2 matrix, instead of the $2N\times 2N$ covariance matrix in (32).

It is known that the log-likelihood function $\ell_W(\boldsymbol{\theta})$ given in (33) approximates the bivariate log-likelihood function $\ell(\boldsymbol{\theta})$ given in (32), but in general $\widehat{\boldsymbol{\theta}}^{(W)} \neq \widehat{\boldsymbol{\theta}}$. The quality of the approximation (see [22]) depends on properties of the spectrum (as is clear from the results of [24]). The difference between the two objective functions specified by (32) and (33) can be bounded, both in terms of mean and variance [22, Theorem 5.2], and the rates achieved asymptotically in estimation using the Whittle likelihood is established by [22, Theorem 5.5]. Asymptotically, the same rates of convergence ($\sqrt{N}$) are achieved for the Whittle and standard log-likelihood.

C. The Whittle Likelihood for Complex-Valued Processes

The Whittle likelihood for complex-valued processes has an identical form to the Whittle likelihood for bivariate processes, as is now shown. To find its form, we express (33) in terms of $Z \equiv X + iY$. We start by defining the DFT for complex-valued processes given by

$$J_C(\omega) = \begin{bmatrix} J_Z(\omega) \\ J_{Z^*}(\omega) \end{bmatrix} = \sqrt{\frac{\Delta}{N}} \sum_{t=1}^{N} \begin{bmatrix} Z_t \\ Z_t^* \end{bmatrix} e^{-i\omega t\Delta}.$$

There is a simple linear relationship between the DFTs $J_C(\omega)$ and $J_B(\omega)$

$$J_C(\omega) = \sqrt{2}\boldsymbol{T} J_B(\omega), \tag{34}$$

where $\boldsymbol{T}$ is the unitary matrix defined in (13). We then substitute (34) and (16) into (33) such that

$$\ell_W(\boldsymbol{\theta}) \overset{C}{=} -\sum_{\omega\in\Omega} \left[\log\left| \frac{1}{2}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{S}_C(\omega;\boldsymbol{\theta})\boldsymbol{T} \right| \right.$$
$$+ J_C^{\mathrm{H}}(\omega)\frac{\boldsymbol{T}}{\sqrt{2}}\left\{ \frac{1}{2}\boldsymbol{T}^{\mathrm{H}}\boldsymbol{S}_C(\omega;\boldsymbol{\theta})\boldsymbol{T} \right\}^{-1} \frac{\boldsymbol{T}^{\mathrm{H}}}{\sqrt{2}}J_C(\omega) \right]$$
$$\overset{C}{=} -\frac{1}{2} \sum_{\omega\in\Omega} \left[\log|\boldsymbol{S}_C(\omega;\boldsymbol{\theta})| + J_C^{\mathrm{H}}(\omega)\boldsymbol{S}_C^{-1}(\omega;\boldsymbol{\theta})J_C(\omega) \right],$$
(35)

and this objective function is thus identical in appearance to the one provided in (33). This result holds because $[Z\ Z^*]^T$ is a scaled unitary transformation of $[X\ Y]^T$, and would still hold if any scaled unitary matrix were substituted for $\boldsymbol{T}$.

We note that for proper signals, when $R_{ZZ}(\omega) = 0$, this log-likelihood function takes the simplified form

$$\ell_W(\boldsymbol{\theta}) \overset{C}{=} -\sum_{\omega\in\Omega} \left[\log\{S_{ZZ}(\omega;\boldsymbol{\theta})\} + \frac{\hat{S}_{ZZ}(\omega)}{S_{ZZ}(\omega;\boldsymbol{\theta})} \right], \tag{36}$$

where $\hat{S}_{ZZ}(\omega) = |J_Z(\omega)|^2$ is the periodogram of the vector Z. This form occurs because the off-diagonal elements of $\boldsymbol{S}_C$ vanish, while the two main diagonal terms contribute identically, thus removing the factor of 1/2 from (35). The log-likelihood function (36) has a recognizable form which is similar to the Whittle likelihood for real-valued signals. In general however, the complementary spectrum must be considered and the form of (35) must be adopted.

D. The Whittle Likelihood for Rotary Components

We find the Whittle likelihood for rotary components Z^+ and Z^- takes a slightly different form. We start by defining the rotary DFT vector by

$$J_\pm(\omega) = \begin{bmatrix} J_+(\omega) \\ J_-(\omega) \end{bmatrix} = U(\omega) \begin{bmatrix} J_Z(\omega) \\ J_Z(-\omega) \end{bmatrix}, \tag{37}$$

where the scalar function $U(\omega)$ is given in (19). It then follows from (23) that (35) can be expressed in terms of rotary

components by

$$\ell_W(\boldsymbol{\theta}) \stackrel{C}{=} - \sum_{\omega \in \Omega^+} \left[\log |\boldsymbol{S}_\pm(\omega; \boldsymbol{\theta})| + J_\pm^{\text{H}}(\omega) \boldsymbol{S}_\pm^{-1}(\omega; \boldsymbol{\theta}) J_\pm(\omega) \right],$$
(38)

where $\boldsymbol{S}_\pm(\omega; \boldsymbol{\theta})$ is defined as in (22). The summation only includes the frequencies $\omega \in \Omega^+$ where Ω^+ is the set of positive discrete Fourier frequencies: $\frac{2\pi}{N\Delta}(1, \ldots, \lfloor N/2 \rfloor)$. This is because $\boldsymbol{S}_\pm$ vanishes for negative frequencies, which removes the need for the factor of 1/2 that accounts for the double-counting of contributions to the other two Whittle likelihood functions ((33) and (35)) at positive and negative frequencies. These multiplicative factors do not affect the optimization of the log-likelihood function, but are important when performing generalized likelihood ratio tests, as we shall demonstrate in Section V-B. This form of the log-likelihood function allows the direct implementation and estimation of rotary coherence models specified in Section III-B, such as (31); and we therefore use (38) in our practical example in Section VI.

The Whittle likelihood for rotary components does not require the observed signal Z to be explicitly split into rotary components; instead this split is performed implicitly from the definition of $J_\pm(\omega)$ in (37). If one does wish to obtain the analytic and conjugate analytic time series from Z then this can be done through discrete Hilbert transforms (see also [25])

$$\begin{bmatrix} Z^+ \\ (Z^-)^* \end{bmatrix} \equiv \frac{1}{\sqrt{2}} \boldsymbol{T} \begin{bmatrix} Z \\ \mathcal{H}Z \end{bmatrix} = \frac{1}{2} \begin{bmatrix} (X - \mathcal{H}Y) + i(Y + \mathcal{H}X) \\ (X + \mathcal{H}Y) + i(Y - \mathcal{H}X) \end{bmatrix},$$

with $\mathcal{H}Z$ denoting the discrete Hilbert transform of Z, and similarly for X and Y. The vectors Z^+ and Z^- are not direct samples from the theoretical processes $Z^+(t)$ and $Z^-(t)$, but are instead discrete approximations to the continuous-time Hilbert transforms, rather like how the discrete Fourier transform is related to the Fourier transform.

Note that the frequency $\omega = 0$ is not included in Ω^+. This occurs because the rotary DFT (37) is split evenly between the components at $\omega = 0$, therefore the rotary coherency cannot be estimated at this frequency. This is unlikely to be a substantial drawback in practice, as the zero-frequency content of the signal is lost whenever the mean of the signal is removed prior to analysis, such that there is no contribution to the log-likelihood function at this frequency in any case. If it is important to include the zero-frequency component in the model, and the amplitude of the signal at this frequency is appreciably greater than zero, then the contribution to the log-likelihood function at this frequency can be added into (38) by taking the zero frequency contribution from (33) or (35).

Finally, (33), (35), or (38) can be combined with procedures to reduce known bias effects of the Whittle likelihood with moderate sample sizes [24]. These procedures include tapering the signal, for example with a cosine or Slepian taper [26]; or by using the *de-biased* Whittle likelihood, proposed in [27] for real-valued signals. It has been observed that the bias reductions in these methods are often still significant for sample sizes of the order of 1,000 observed points, particularly with processes whose spectral densities are hard to approximate because of leakage and aliasing effects [28].

V. STATISTICAL PROCEDURES FOR COMPLEX-VALUED AND ROTARY SIGNALS

In this section we discuss several useful practical procedures when stochastically modeling pairs of signals using the complex-valued or rotary representation.

A. Semi-Parametric Modeling and Estimation

Commonly signals are acquired to be both low-pass filtered and notch filtered [29] before digitized, thus missing a range of frequencies. In other application areas, the proposed parametric model only fits over a range of frequencies [15], [30]–[34]. In these instances

$$S_\pm(\omega) = \begin{cases} S_\pm(\omega; \boldsymbol{\theta}) & \omega \in \Omega_S \\ S_\pm(\omega) & \omega \notin \Omega_S, \end{cases}$$
(39)

where Ω_S is the range of frequencies in which $S_\pm(\omega)$ can be modeled parametrically from the data.

For all such instances it is not suitable to infer the parameters of the generating mechanism of the continuous time process using (35) or (38) when any of the Fourier frequencies $\frac{2\pi}{N\Delta}(1, \ldots, \lfloor N/2 \rfloor) \notin \Omega_S$. The log-likelihood function for $\boldsymbol{\theta}$ in (38) should instead be restricted to the Fourier frequencies $\omega \in \Omega_S$ such that

$$\ell_W(\boldsymbol{\theta}) \stackrel{C}{=} - \sum_{\omega \in \Omega_S} \left[\log |S_\pm(\omega; \boldsymbol{\theta})| + J_\pm^{\text{H}}(\omega) S_\pm^{-1}(\omega; \boldsymbol{\theta}) J_\pm(\omega) \right],$$
(40)

and equivalently with (35). In fact, for the case of [35] the optimal number of frequencies to use in estimation can be determined and is not $\sim N$ but rather scales as $\sim N^{4/5}$. In general we would expect the precision of parameter estimates to scale as $|\Omega_S|^{-1/2}$, where $|\Omega_S|$ denotes the number of Fourier frequencies contained in Ω_S, thus loosing in precision for the reduced bias.

This means that only the frequencies used in the log-likelihood summation determine the parameter estimates, and all other sampled frequencies are excluded. In such instances the resulting parameter estimation procedure is *semi-parametric* [16]. For set problems, such as that treated in [32], [35], the subset can be determined analytically prior to analysis. Sometimes high frequencies will be omitted to account for sampling errors such as measurement noise. This effect is well-documented in econometrics, finance and geostatistics [36], [37], and while noise often affects all frequencies equally, the signal is usually concentrated at low-frequencies. Additionally, effects due to sampling are sometimes stronger in their contamination of higher frequencies [38].

We can also use the semi-parametric approach to model only one of the rotary components, $Z^+(t)$ or $Z^-(t)$, akin to modeling only one side of the spectrum in the complex representation. This can be performed, for instance, when the physical process of interest is known to only spin in one direction for a given signal. This can simplify analysis when one side of the spectrum is known to be contaminated by nuisance effects [34], and while excluding these frequencies will increase variance, the bias will

be decreased by a more significant amount, thus reducing overall estimation error.

B. Hypothesis Testing for Impropriety

In Section III-B we specified models for the coherency between the rotary components $Z^+(t)$ and $Z^-(t)$. We may wish to test a proposed model for coherency against the simpler scenario where $\rho_\pm(\omega) = 0$ for all ω in (30). This hypothesis corresponds to there being no relationship between the positive rotating and negative rotating phasors. We note from Table I that this hypothesis is *not* equivalent to $S_{XY}(\omega) = 0$; rather, a process with zero rotary cross-spectra, $S_{+-}(\omega) = 0$, is equivalent to a proper process, $R_{ZZ}(\omega) = 0$ for all $\omega \neq 0$. This is a deliberate and convenient consequence of the way we have constructed $Z^+(t)$ and $Z^-(t)$ to *both* be analytic. Testing for impropriety in the time domain is then equivalent to testing for a non-zero cross-spectrum between rotary components in the frequency domain.

There are many tests for impropriety using data sets of multiple replicates of complex-valued vectors, see for example [6], [39]–[42]. We note that these are *non-parametric tests* in general requiring *replicated* complex-valued vectors. Here, by contrast, we have a simple *parametric* time series model suitable for a *single time series*, and will derive a test statistic for this scenario.

Given a chosen model for $S_{++}(\omega)$ and $S_{--}(\omega)$, we estimate the coherency structure by specifying a parametric model for $\rho_\pm(\omega)$ in (30). We may however prefer to use the simpler model where $\rho_\pm(\omega) = 0$, such that $S_{+-}(\omega) = 0$. We can make use of the parametric model—in combination with the Whittle likelihood for complex-valued signals—to perform a generalized likelihood ratio test to check for evidence if $\rho_\pm(\omega) \neq 0$ such that $S_{+-}(\omega) \neq 0$. Specifically, this is done by computing the generalized likelihood ratio test statistic

$$W = 2\left[\ell_W(\widehat{\boldsymbol{\theta}}_1) - \ell_W(\widehat{\boldsymbol{\theta}}_0)\right], \tag{41}$$

where $\widehat{\boldsymbol{\theta}}_1$ includes non-zero estimates for the coherency $\rho_\pm(\omega)$, while for $\widehat{\boldsymbol{\theta}}_0$, $\widehat{\rho}_{+-}(\omega) = 0$ for all ω. The estimators $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_0$ are obtained by maximizing (38). We assume that the full model with non-zero coherence has p extra parameters that are not linearly dependent. For example, in the case of the coherency structure proposed in (27) and (29), then $p = r + l + 1$. Whereas in (31) if we set $a_1 = 0$ for example, such that we have a fixed constant for $\rho_\pm(\omega)$ when $\omega < \omega_0$, then we have $p = 1$.

In more generality, if we use the time domain bivariate log-likelihood function of (32) in (41), then from standard likelihood theory, the test statistic W is asymptotically distributed according to a χ_p^2 distribution with degrees of freedom equal to the number of extra parameters p in the alternative hypothesis versus the null. In Appendix A, we prove that the Whittle likelihood for complex signals also asymptotically yields a test statistic with a χ_p^2 distribution under the null, for special cases where the rotary spectra are equal in magnitude across frequency, but possibly correlated with each other such that the process is improper.

C. Model Choice

In many practical scenarios the most appropriate choice for a model, and the corresponding number of parameters, is unknown *a priori*. When it is possible to define a model such that possible candidate models are *nested*—that is, with simpler models being recovered from a more complete model by setting certain parameters to constants—then it is possible to apply the method of *model choice* [43]. In this section, we provide the correct form of various model choice procedures for complex-valued signals.

The Akaike Information Criterion (AIC) is a model choice procedure that can be used, for example, to find the most appropriate choice of q_1 and q_2 for a range of $\mathrm{ARMA}(q_1, q_2)$ processes [44]. Alternatively, the AIC can be used to select the order of the coherency models proposed in (26) or (27). For complex-valued signals, the AIC can be closely approximated in the frequency domain by using the Whittle likelihood to approximate the time-domain log-likelihood, and combining this with a model penalization parameter as in [44] it follows that it takes the same form as for real-valued signals

$$\mathrm{AIC}(\hat{\boldsymbol{\theta}}) = -2\ell(\hat{\boldsymbol{\theta}}) + 2q \approx -2\ell_W(\hat{\boldsymbol{\theta}}^{(W)}) + 2q,$$

where q is the number of parameters in the model. The model with the smallest AIC value is then selected.

For small sample sizes, it is often recommended to use a correction to the AIC known as the AICC [44], which for complex-valued signals is given by

$$\mathrm{AICC}(\hat{\boldsymbol{\theta}}) = -2\ell(\hat{\boldsymbol{\theta}}) + \frac{4qN}{2N - q - 1}$$

$$\approx -2\ell_W(\hat{\boldsymbol{\theta}}^{(W)}) + \frac{4qN}{2N - q - 1}, \tag{42}$$

which converges to the AIC for large sample sizes. The correction uses $2N$ (rather than N as would be done for real-valued signals), as there are $2N$ degrees of freedom in a length N complex-valued signal. In the case of semi-parametric estimation as discussed in Section V-A, (42) should not be used in this exact form, as the sample-size correction needs to take account of degrees of freedom in the data not used in the estimation. We instead replace N by $|\Omega_S|$ in (42), which is the number of Fourier frequencies in Ω_S in (40).

When a tapered spectral estimate is used in the Whittle likelihood, then the degrees of freedom are further reduced by the correlation induced in neighboring Fourier frequencies by the taper [28]. This loss of resolution is also known as narrowband blurring, in contrast to the broadband blurring which is attributed to the leakage of the Fejér kernel in the periodogram for example. Without tapering, nominally each pair of Fourier coefficients, spaced $1/(N\Delta)$ apart, are uncorrelated; however, with tapering, coefficients spaced $\gamma/(N\Delta)$ apart are approximately uncorrelated, where $\gamma > 1$ is some constant reflecting the implicit bandwidth of frequency-domain smoothing. We must then further reduce N (or $|\Omega_S|$ for semi-parametric estimation) by a factor of γ in the AICC, which can be computed by Fourier transforming the taper onto a fine grid
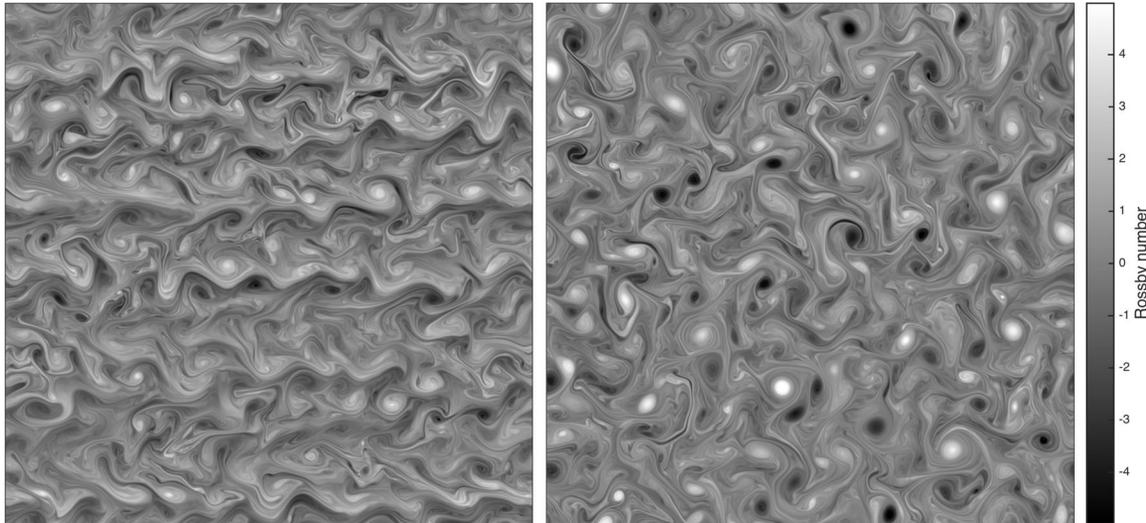
Fig. 2. Snapshots from numerical experiments for anisotropic (left) and isotropic (right) two-dimensional turbulence. The field presented is relative vorticity, the curl of the velocity vector field at each point. East-west bands are apparent in the anisotropic simulation that do not appear in the isotropic simulation. The color scale shows the Rossby number, a non-dimensional measure of the vorticity strength.

and then finding the frequency width where the transformed sequence is only correlated below some specified threshold.

VI. APPLICATION TO TURBULENT FLOW DATA

In this section we present an application of the modeling and estimation methods we have developed for bivariate and complex-valued signals. All results can be exactly reproduced using MATLAB code freely downloadable from http://ucl.ac.uk/statistics/research/spg/software, which also includes a technical description of how the turbulent flow data is numerically generated. The numerical simulation can be reproduced using software available at http://jeffreyearly.com/numerical-models/

We test our modeling and estimation procedures on data obtained from quasi-geostrophic turbulent flow simulations. Specifically we track the spatial trajectories of 256 particles over time from two sets of numerical experiments, snapshots of which are displayed in Fig. 2. Here trajectories of particles are computed from the time-varying velocity fields associated with quasi-steady state forced-dissipative two-dimensional turbulence simulations under both isotropic (f-plane) and anisotropic (β-plane) dynamics. Such simulations are standard in oceanography, see e.g. [45] for details.

We represent the velocities of the Lagrangian trajectories as complex-valued signals, which we denote as $z = u + iv$. The time between observations of the signal is set to 1 day. We model z as following a Matérn process, as motivated in [46], which in the isotropic case has rotary spectra defined by

$$S_{++}(\omega) = S_{--}(\omega) = \frac{\phi^2}{(\omega^2 + \alpha^2)^{\nu+1/2}}, \quad \omega > 0, \quad (43)$$

where ω is given in cycles per day. The smoothness parameter $\nu > 0$ defines the Hausdorff dimension of the graph—equal to $\max(1, 2 - \nu)$—as well as the degree of differentiability of the

process. The range parameter $\alpha > 0$ is a timescale parameter, where $1/\alpha$ can be referred to as the correlation timescale, and $\phi^2 > 0$ defines the magnitude of the variability of the process. The motivation for using the Matérn for turbulence velocities is based on its properties: both the Matérn and Lagrangian trajectories such as these exhibit close to power law behavior at mid-to-high frequencies (see also the review paper of [47]). Moreover, in contrast to fractional Brownian Motion, the power law behavior of a Matérn breaks at low frequency making the process stationary (and not self-similar)—this is again consistent with the discussions of [46], [47]. We choose to model the two rotary components as Matérn processes with the same parameters, as in this experiment there is no preferred or dominant direction of rotation of the particles, such that the rotary spectra are expected to be symmetric.

We account for potential anisotropy by modeling the rotary coherency. We employ the model of (31) setting $a_0 = 1$, such that the coherency approaches unity as ω approaches zero. The justification for this is physical, as at long timescales we expect the east-west bands in the left panel of Fig. 2 to be entirely dominant, which is consistent with a fully improper/anisotropic model with rotary coherency equal to unity. This leads to the following 1-parameter model

$$S_{+-}(\omega) = \rho_\pm(\omega) \left[S_{++}(\omega)S_{--}(\omega)\right]^{1/2}$$
$$\rho_\pm(\omega) = \max(0, 1 - c\omega), \quad \omega \geq 0, \quad (44)$$

where $c \in \mathbb{R}^+$. The choice of compactly supporting $\rho_\pm(\omega)$—as discussed in Section III-B—is also physical, where isotropic behavior is expected at high frequencies beyond some physical timescale. Finally, as also discussed in Section III-B, we ignore the group delay and model $\rho_\pm$ as real-valued, which is reasonable if the signal components, in this case u and v, are uncorrelated with each other. To check this, in Fig. 3 we display the normalized cross-covariance between u and v averaged
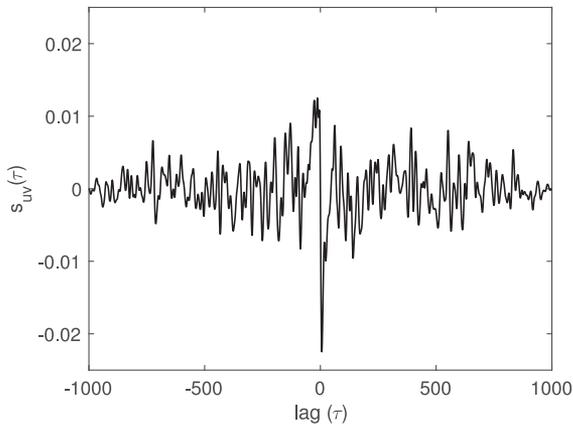
Fig. 3.   The average autocorrelation of u and v across all observable lags of the signals analyzed from the anisotropic experiment displayed in Fig. 2 (left). The estimated autocorrelation sequence has been averaged across the 256 signals analyzed in Section VI. The autocorrelations are estimated using the biased autocovariance estimator to reduce variance, where each lag is normalized by N rather than $N - \tau$.

across 256 trajectories (each of length 1,001) from the numerical model, where it can be seen that it is reasonable to make the assumption that u and v are uncorrelated, as the estimated correlation never exceeds 0.025 at any lag.

The full stochastic model therefore has four parameters: three for the Matérn process, and one to specify the rotary coherency. The proposed model is a valid Gaussian process as $S_{++}(\omega) \geq 0$, $S_{--}(\omega) \geq 0$, $|\rho_{\pm}(\omega)| \leq 1$ for all $\omega > 0$, and the spectral matrix $\boldsymbol{S}_{\pm}(\omega)$ is integrable, as discussed in Section III-B. The choice of the "triangle function" for the coherency in (44) naturally defines a timescale at which anisotropy begins, as the coherence is zero at $\omega = 1/c$, corresponding to a period of $2\pi c$, and then increases linearly as frequency decreases until at long timescales all energy is in the dominant Cartesian component, which is u. The model of (44) defines the rotary coherency and its parameter can be estimated, together with the parameters of the Matérn process (43), using the Whittle likelihood for rotary components as provided in Section IV-D. The parameters in the optimization are initialized using least squares, namely starting from (43) we first assume $\alpha = 0$ and rewrite

$$\log S_{++}(\omega) = \log \phi^2 - (2\nu + 1)\log(\omega), \qquad (45)$$

and regress the observed log-periodogram $\log \hat{S}_{++}(\omega)$ over mid-range frequencies to make a least squares fit of ν and ϕ. The parameter α is then set to a mid-range value of 0.1 cycles per day, and c is set by finding the lowest frequency at which the estimated rotary coherency from multi-taper spectral estimates is zero. More details can be found in the online code.

In Fig. 4, we display the Whittle likelihood fit to an individual signal (of length 1,001) obtained from the anisotropic numerical simulation shown in the left panel of Fig. 2. Because of the steep energy roll-off, we have used a semi-parametric fit by excluding 60% of the frequency range, and thus only modeling up to 0.2 cycles per day, where the Nyquist is 0.5 cycles per day. We have also excluded the zero frequency from the fit, as we

have removed the sample mean, and hence there is no spectral content at $\omega = 0$. As the spectral slopes are steep, we use the tapered version of the Whittle likelihood [24], as discussed in Section IV. Specifically we estimate the parameters of the modeled spectra using non-parametric multi-taper spectral estimates obtained from discrete prolate spheroidal sequences (dpss) [48], otherwise known as Slepian tapers, with bandwidth parameter set to 3.

Analyzing the fits in Fig. 4, it can be seen that the extra parameter in our model has succinctly captured the difference between the flow in each Cartesian component (u and v) at low frequencies. Thus the Matérn model of (43), plus rotary coherency as modeled in (44), appears to generally be a good fit for this complex-valued signal. We note that the estimated rotary coherency from multi-tapers is a noisy estimate, as compared with the estimate of the rotary power spectra contained in $S_{ZZ}(\omega)$. This is expected, particularly at higher frequencies, as the estimated coherency is measured from an individual signal, and at higher frequencies is often the ratio of two small but noisy quantities. Despite this, the parametric model for rotary coherency appears to have obtained a reasonable estimate, and captured the frequency-dependent structure of the coherency, particularly at low frequencies.

To put our choice of model under further scrutiny, we fit this 4-parameter anisotropic Matérn model to all of the 256 trajectories from the anisotropic experiment and compare the value of the log-likelihood function versus the null hypothesis of a three-parameter isotropic Matérn model, in which the rotary coherency is identically zero ($\rho_{\pm}(\omega) = 0$). We can then use a generalized likelihood ratio test, as described in Section V-B, to test for evidence of anisotropy. For consistency, we also repeat this procedure with the same number of trajectories (of the same length) from the isotropic experiment in the right panel of Fig. 2, to see if we correctly do not reject the null in such cases. While the null model is nested within the alternate model, the null value of the parameter c is at the boundary of its range, requiring adjustments to calculate the critical value for the test statistic. Therefore, we compute 95% confidence intervals for our test statistic by bootstrapping a number of isotropic simulated Matérns (with parameters similar to those estimated in Fig. 4), rather than using a chi-squared statistic as derived in the Appendix.

The set of test statistics, calculated from (41), from both experiments is displayed in Fig. 5(a). The statistics are compared with the 95% one-sided bootstrapped confidence interval. The isotropic model is correctly always rejected for the anisotropic data and rejected only 16 times for the isotropic data (6.25% of the 256 observed signals). This is in broad agreement with a type I error level set to 5%. In experimental rather than controlled data we would account for multiple testing issues if performing this test over multiple trials, applying techniques such as False Discovery Rates (FDR).

The timescale associated with the observed anisotropy is of interest. As the signals were observed from a numerical model, we can assess which frequencies are associated with the anisotropic behavior from the settings. A spatial scale known as
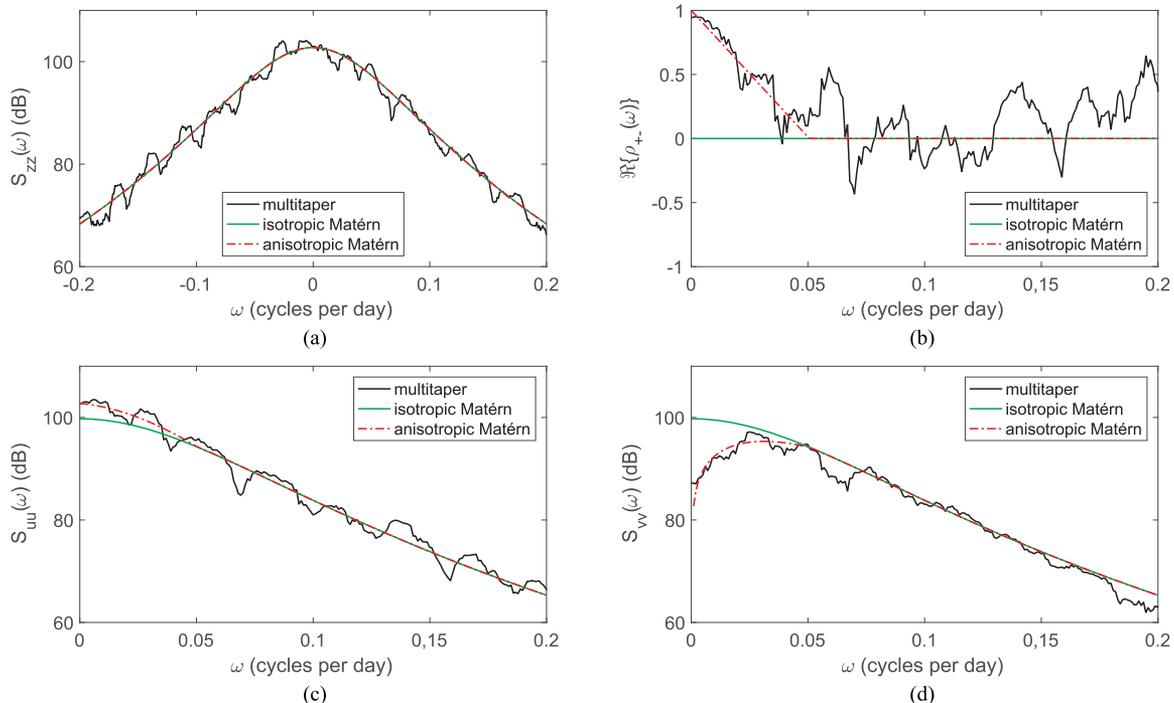
Fig. 4. Spectra from the isotropic and anisotropic Matérn model, with parameters estimated from the Whittle likelihood, plotted against non-parametric multi-taper spectral estimates from a complex-valued velocity signal observed from the anisotropic numerical simulation shown in the left panel of Fig. 2. Panel (a) is the Matérn fit to the power spectrum $S_{ZZ}(\omega)$, which is identical for the isotropic and anisotropic models. Panel (b) displays the fit to the anisotropic model of $\rho_{\pm}(\omega)$ defined in (44), where the isotropic model is zero at all frequencies by definition. Panels (c) and (d) display the anisotropic and isotropic model fits to the power spectrum of the velocities in the u and v direction only. The multi-tapers used are the discrete prolate spheroidal sequences (dpss) with bandwidth parameter set to 3.
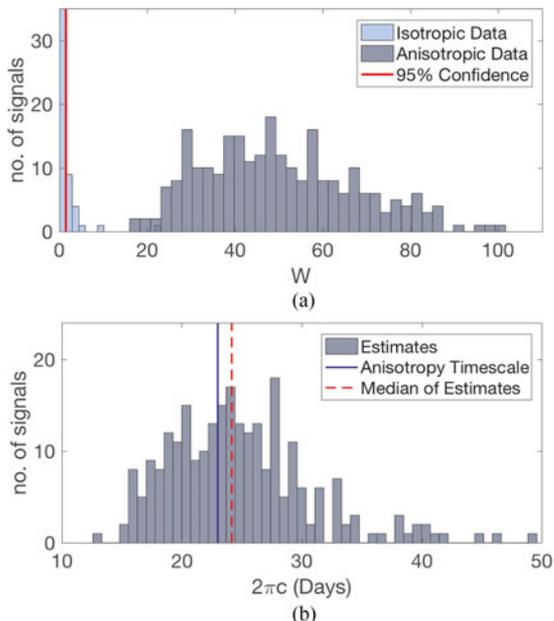

Fig. 5. (a) generalized likelihood ratio test statistics computed from 256 isotropic and anisotropic trajectories where the first column contains 240 values. The 95% confidence interval is obtained from bootstrapping. (b) anisotropy timescale estimates $2\pi\hat{c}$ from the anisotropic trajectories.

the Rhines scale [49], well-known in oceanography, determines the scale at which the transition to anisotropic large-scale behavior begins; this can be converted to a temporal scale through

a division by the root-mean-square velocity. This gives a time-scale of approximately 23 days for the anisotropic experiment. Fig. 5(b) displays estimates of this timescale for each signal from our parametric model, based on the estimate of the frequency at which the rotary coherence becomes zero, which is $2\pi c$. The median value is found to be 24.2 days, consistent with the 23 days computed from the Rhines scale. This apparent ability to infer a key *spatial* scale based solely on the *frequency* structure of time signals obtained from Lagrangian trajectories is an interesting result showing the power of this method. Translating the temporal content garnered from multiple signals using our methods, to local spatial summaries, is an important avenue for future work.

VII. CONCLUSIONS

In this paper we have proposed a framework for stochastic modeling and estimation of stationary bivariate or complex-valued signals. We have shown the power of separating out behavior in frequency using the rotary components, and modeling different ranges of frequencies separately. This permitted us to handle a plethora of different effects directly in the frequency domain, and introduce new signal characteristics. For example, we demonstrated how our techniques can be used to effectively capture anisotropy in oceanographic flow models. In addition, we have proposed appropriate computationally-efficient parameter estimation procedures by extending the Whittle likelihood objective function to complex-valued and rotary signals.

There remain significant challenges in the frequency domain analysis of bivariate and complex-valued signals. A key challenge is to extend our modeling framework to nonstationary and higher order processes, where advances in non-parametric modeling have been made in [6]. The main application challenge is to continue building models from our framework for use in a wide range of physical applications, as we have performed recently in [14] with seismic data signals.

APPENDIX A
HYPOTHESIS TESTING FOR IMPROPRIETY-PROOF OF χ^2_p-DISTRIBUTED TEST STATISTIC

Hypothesis Test Set-Up

We shall consider the special cases of

$$\mathbf{S}^{(0)}_\pm (\omega; \boldsymbol{\theta}_0) = S(\omega; \boldsymbol{\theta}_M) \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{vs}$$

$$\mathbf{S}^{(1)}_\pm (\omega; \boldsymbol{\theta}_1) = S(\omega; \boldsymbol{\theta}_M) \cdot \begin{bmatrix} 1 & \rho(\omega; \boldsymbol{\psi}) \\ \rho(\omega; \boldsymbol{\psi}) & 1 \end{bmatrix}.$$

Where the null and alternate parameters are denoted as

$$\boldsymbol{\theta}_0 = \begin{bmatrix} \boldsymbol{\theta}_M \\ \boldsymbol{\psi}_o \end{bmatrix}, \quad \text{vs} \quad \boldsymbol{\theta}_1 = \begin{bmatrix} \boldsymbol{\theta}_M \\ \boldsymbol{\psi} \end{bmatrix}.$$

We shall now develop a test for the null hypothesis of

$$H_0: \quad \mathbf{S}_\pm(\omega) = \mathbf{S}^{(0)}_\pm(\omega; \boldsymbol{\theta}_0), \quad \text{vs}$$

$$H_1: \quad \mathbf{S}_\pm(\omega) = \mathbf{S}^{(1)}_\pm(\omega; \boldsymbol{\theta}_1).$$

We note that the hypotheses are nested, and that makes standard theory possible—however this is not a necessary requirement [43]. Assume that $\boldsymbol{\theta}_M$ is an l-vector and $\boldsymbol{\psi}$ is a p-vector.

Notation for Quadratic Approximations

Let us first define the Fisher information matrix

$$\mathcal{F}(\boldsymbol{\theta}_1) = \begin{bmatrix} \tilde{\mathcal{F}}(\boldsymbol{\theta}_1) & \mathcal{F}_\times(\boldsymbol{\theta}_1) \\ \mathcal{F}^T_\times(\boldsymbol{\theta}_1) & \mathcal{F}_\circ(\boldsymbol{\theta}_1) \end{bmatrix},$$

where we have decomposed $\mathcal{F}(\boldsymbol{\theta}_1)$ into 4 blocks to simplify calculations: $\tilde{\mathcal{F}}(\boldsymbol{\theta}_1)$ contains the negative of the expected mixed second derivatives of the log-likelihood function with respect to the parameters in $\boldsymbol{\theta}_M$, $\mathcal{F}_\times(\boldsymbol{\theta}_1)$ is the matrix with entries of the negative of the expectation of the second cross-derivatives between $\boldsymbol{\theta}_M$ and $\boldsymbol{\psi}$, and finally $\mathcal{F}_\circ(\boldsymbol{\theta}_1)$ contains the negative of the expectation of the mixed second derivatives of the log-likelihood function with respect to the parameters in $\boldsymbol{\psi}$. In parallel we define the observed Fisher information matrices as $\boldsymbol{F}$, which has been decomposed into $\tilde{\boldsymbol{F}}$, containing the negative of mixed second derivatives of the log-likelihood function with respect to the parameters in $\boldsymbol{\theta}_M$, $\boldsymbol{F}_\times(\boldsymbol{\theta}_1)$ is the matrix with entries of the negative of second cross-derivatives between $\boldsymbol{\theta}_M$ and $\boldsymbol{\psi}$, and finally $\boldsymbol{F}_\circ(\boldsymbol{\theta}_1)$ contains the negative of mixed second derivatives of the log-likelihood function with respect to the parameters in $\boldsymbol{\psi}$. Note that the observed Fisher information, and its expectation, as well as the corresponding submatrices are all a function of $\boldsymbol{\theta}_1 = [\boldsymbol{\theta}_M \ \boldsymbol{\psi}]^T$, where we recall also that $\boldsymbol{\theta}_0 = [\boldsymbol{\theta}_M \ \boldsymbol{\psi}_o]^T$.

We now also write

$$\hat{\boldsymbol{\theta}}_0 = \begin{bmatrix} \hat{\boldsymbol{\theta}}_{0M} & \boldsymbol{\psi}_o \end{bmatrix}^T, \quad \text{and} \quad \hat{\boldsymbol{\theta}}_1 = \begin{bmatrix} \hat{\boldsymbol{\theta}}_{1M} & \hat{\boldsymbol{\psi}} \end{bmatrix}^T,$$

for the parameter values that maximize (38), with the constraints of H_0, or without such constraints respectively.

Properties of the Whittle Estimators

We have already assumed that $\boldsymbol{F}(\boldsymbol{\theta}_0)$ is a continuous function of $\boldsymbol{\theta}_0$. We note that under the null $\hat{\boldsymbol{\theta}}_0 \xrightarrow{P} \boldsymbol{\theta}_0$, and $\hat{\boldsymbol{\theta}}_1 \xrightarrow{P} \boldsymbol{\theta}_0$, from Theorem 5.4 of [22]. Thus, with $N_{|\Omega|}$ denoting the cardinality of Ω, we rewrite the observed Fisher matrix evaluated at a point $\boldsymbol{\theta}'_0$ squeezed between $\hat{\boldsymbol{\theta}}_1$ and the true parameter value as

$$\boldsymbol{F}(\boldsymbol{\theta}'_0) = N_{|\Omega|}\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0) + N_{|\Omega|}\delta\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0),$$

for

$$\delta\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0) = N^{-1}_{|\Omega|}\boldsymbol{F}(\boldsymbol{\theta}'_0) - \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0).$$

Using the continuous mapping (or Mann-Wald) theorem, it follows as $\hat{\boldsymbol{\theta}}_1 \xrightarrow{P} \boldsymbol{\theta}_0$,

$$\mathrm{E}\|\delta\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)\|^2_F = o(1),$$

where $o(1)$ is standard notation for a shrinking quantity. We can therefore deduce $\|\delta\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)\|^2_F$ is $o_P(1)$. This is the definition of $\delta\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0) = o_P(1)$, with the chosen matrix norm being the Frobenius norm. As $\delta\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)$ is of finite size, any sensibly chosen matrix norm, such as the trace norm will be equivalent in order to the Frobenius norm. We therefore may write

$$N^{-1}_{|\Omega|}\boldsymbol{F}(\boldsymbol{\theta}'_0) = \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)\{\mathbf{I}_{l+p} + o_P(1)\}.$$

The same argument holds for the reduced system excluding $\boldsymbol{\psi}$. Therefore for $\boldsymbol{\theta}^*_0$ between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_0$

$$N^{-1}_{|\Omega|}\tilde{\boldsymbol{F}}(\boldsymbol{\theta}^*_0) = \tilde{\boldsymbol{\mathcal{F}}}(\boldsymbol{\theta}_0)\{\mathbf{I}_l + o_P(1)\}.$$

We shall now link the parameter estimates to the score. Write $\nabla = (\partial/\partial\theta_{11} \ldots \partial/\partial\theta_{1l+p})^T$ for the gradient, and write $\nabla_M = (\partial/\partial\theta_{01} \ldots \partial/\partial\theta_{0l})^T$ for the gradient under the null model. We note that performing Taylor series of the score yields

$$\nabla_M \ell_W(\hat{\boldsymbol{\theta}}_0) = \nabla_M \ell_W(\boldsymbol{\theta}_0) - \tilde{\boldsymbol{F}}([\boldsymbol{\theta}''_{0M} \ \boldsymbol{\psi}_o]^T)\{\hat{\boldsymbol{\theta}}_{0M} - \boldsymbol{\theta}_{0M}\}$$

$$\nabla \ell_W(\hat{\boldsymbol{\theta}}_1) = \nabla \ell_W(\boldsymbol{\theta}_0) - \boldsymbol{F}(\boldsymbol{\theta}^\dagger_0)\{\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0\},$$

as we are using a Lagrange form of the Taylor series where the remainder $\boldsymbol{\theta}''_{0M}$ is squeezed between $\boldsymbol{\theta}_{0M}$ and $\hat{\boldsymbol{\theta}}_{0M}$, as $\hat{\boldsymbol{\theta}}_{0M} \xrightarrow{P} \boldsymbol{\theta}_{0M}$, and $\boldsymbol{\theta}^\dagger_0$ is squeezed between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_1$, as $\hat{\boldsymbol{\theta}}_1 \xrightarrow{P} \boldsymbol{\theta}_0$, under the null. This naturally produces a representation of

$$\hat{\boldsymbol{\theta}}_{0M} - \boldsymbol{\theta}_{0M} = [\tilde{\boldsymbol{\mathcal{F}}}(\boldsymbol{\theta}_0) + o_P(1)]^{-1} N^{-1}_{|\Omega|}\nabla_M \ell_W(\boldsymbol{\theta}_0), \quad (46)$$

as well as,

$$\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0 = [\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0) + o_P(1)]^{-1} N^{-1}_{|\Omega|}\nabla \ell_W(\boldsymbol{\theta}_0). \quad (47)$$

Calculating the difference between the two estimators for $\boldsymbol{\theta}_M$ from (46) and (47) we get that

$$
\begin{aligned}
\begin{pmatrix} \hat{\boldsymbol{\theta}}_{1M} - \hat{\boldsymbol{\theta}}_{0M} \\ \mathbf{0} \end{pmatrix} &= \begin{bmatrix} \mathbf{I}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} [\mathcal{F} + o_P(1)]^{-1} N_{|\Omega|}^{-1} \nabla \ell_W (\boldsymbol{\theta}_0) \\
&\quad - N_{|\Omega|}^{-1} \begin{bmatrix} \left\{ \tilde{\mathcal{F}} + o_P(1) \right\}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \nabla \ell_W (\boldsymbol{\theta}_0) \\
&= \left\{ -\begin{bmatrix} \tilde{\mathcal{F}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{I}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathcal{F}^{-1} \right\} \\
&\quad \times \mathcal{F} \left(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0 \right) \{ \mathbf{I}_{l+p} + o_P(1) \} \\
&= \left(\begin{bmatrix} \mathbf{I}_l & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{I}_l & \tilde{\mathcal{F}}^{-1} \mathcal{F}_\times \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right) \left\{ \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0 \right\} \\
&\quad \times \{ \mathbf{I}_{l+p} + o_P(1) \}.
\end{aligned}
$$

Note that all matrices and submatrices of the Fisher information above should be evaluated at the value $\boldsymbol{\theta}_0$. This has been omitted for brevity. This expression can then be simplified to

$$
\hat{\boldsymbol{\theta}}_{1M} = \hat{\boldsymbol{\theta}}_{0M} - \tilde{\mathcal{F}}^{-1} \mathcal{F}_\times \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right) \{ \mathbf{I}_p + o_P(1) \}. \tag{48}
$$

Distribution of the Quadratic Form

To determine the distribution of the likelihood ratio statistic we shall assume that 1) The process under observation is a Gaussian stationary process possessing a continuous spectrum, 2) $\boldsymbol{\theta}_0$ lies in an open ball in the parameter set $\boldsymbol{\Theta}$ and 3) the spectrum of the process satisfies certain regularity conditions which will be stated more carefully later in this section. It is not a necessary assumption that the process is Gaussian (see for example work by [50]), but it is sufficient and easy to state. We note that for the example in section VI the null value of the parameter is not in an open set but at the boundary, explaining why we use the parametric bootstrap.

We now return to the quantity W from (41), and implement an additional Taylor series for W, yet again with the Lagrange form of the remainder (assume $\boldsymbol{\theta}_0^*$ lies in a ball centered at $\hat{\boldsymbol{\theta}}_1$ less than $\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_1$ away from the center), an expansion that is possible because $\boldsymbol{\theta}_0$ lies in an open ball in the parameter set $\boldsymbol{\Theta}$. Note

$$
\begin{aligned}
\ell_W \left(\hat{\boldsymbol{\theta}}_0 \right) &= \ell_W \left(\hat{\boldsymbol{\theta}}_1 \right) + \left(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_1 \right)^T \nabla \ell_W \left(\hat{\boldsymbol{\theta}}_1 \right) \\
&\quad - \frac{1}{2} \cdot \left(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_1 \right)^T \boldsymbol{F} (\boldsymbol{\theta}_0^*) \left(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_1 \right). \tag{49}
\end{aligned}
$$

We can use that $\nabla \ell_W (\hat{\boldsymbol{\theta}}_1) = \mathbf{0}$ and then evaluate the variable W defined in (41)

$$
\begin{aligned}
W &= 2\ell_W \left(\hat{\boldsymbol{\theta}}_1 \right) - 2\ell_W \left(\hat{\boldsymbol{\theta}}_0 \right) \\
&= 2\ell_W \left(\hat{\boldsymbol{\theta}}_1 \right) - 2 \left[\ell_W \left(\hat{\boldsymbol{\theta}}_1 \right) \right. \\
&\quad \left. - \frac{1}{2} \cdot \left(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_1 \right)^T \boldsymbol{F} (\boldsymbol{\theta}_0^*) \left(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_1 \right) \right] \\
&= N_{|\Omega|} \left(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_1 \right)^T \mathcal{F} (\boldsymbol{\theta}_0) \left(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_1 \right) (\mathbf{I}_{l+p} + o_P(1)),
\end{aligned}
$$

as $\boldsymbol{\theta}_0^*$ is squeezed between $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_1$, again using the continuous mapping theorem. Substituting in (48) then yields

$$
\begin{aligned}
W &= N_{|\Omega|} \begin{bmatrix} -\tilde{\mathcal{F}}^{-1} \mathcal{F}_\times \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right) \\ \hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \end{bmatrix}^T \mathcal{F} (\boldsymbol{\theta}_0) \\
&\quad \times \begin{bmatrix} -\tilde{\mathcal{F}}^{-1} \mathcal{F}_\times \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right) \\ \hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \end{bmatrix} \{ 1 + o_P(1) \} \\
&= N_{|\Omega|} \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right)^T \begin{bmatrix} -\mathcal{F}_\times^T \tilde{\mathcal{F}}^{-1} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \tilde{\mathcal{F}} & \mathcal{F}_\times \\ \mathcal{F}_\times^T & \mathcal{F}_\circ \end{bmatrix} \\
&\quad \times \begin{bmatrix} -\tilde{\mathcal{F}}^{-1} \mathcal{F}_\times \\ \mathbf{I}_p \end{bmatrix} \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right) \{ 1 + o_P(1) \} \\
&= N_{|\Omega|} \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right)^T \\
&\quad \times \begin{bmatrix} -\mathcal{F}_\times^T \tilde{\mathcal{F}}^{-1} \tilde{\mathcal{F}} + \mathcal{F}_\times^T & -\mathcal{F}_\times^T \tilde{\mathcal{F}}^{-1} \mathcal{F}_\times + \mathcal{F}_\circ \end{bmatrix} \\
&\quad \times \begin{bmatrix} -\tilde{\mathcal{F}}^{-1} \mathcal{F}_\times \\ \mathbf{I}_p \end{bmatrix} \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right) \{ 1 + o_P(1) \} \\
&= N_{|\Omega|} \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right)^T \begin{bmatrix} \mathcal{F}_\circ - \mathcal{F}_\times^T \tilde{\mathcal{F}}^{-1} \mathcal{F}_\times \end{bmatrix} \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right) \\
&\quad \times \{ 1 + o_P(1) \}.
\end{aligned}
$$

The final step comes from observing that from [22, Theorem 5.5], writing $\mathcal{G} = \mathcal{F}^{-1}$, and partitioning the matrix into

$$
\mathcal{G} = \begin{bmatrix} \mathcal{G}_1 & \mathcal{G}_2 \\ \mathcal{G}_2^T & \mathcal{G}_3 \end{bmatrix},
$$

we arrive at

$$
N_{|\Omega|}^{1/2} \left\{ \hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right\} \sim N (\mathbf{0}, \mathcal{G}_3). \tag{50}
$$

For [22, Theorem 5.5] to hold, we need to assume that the observed process is Gaussian, and has a spectrum $S(\omega)$, which for two distinct parameter values are not equal for almost all frequencies, and the inverse spectrum as well as its derivatives with respect to all parameter components are continuous both in frequency and parameter components. Normality can be derived under other assumptions, see e.g. [50], but this simplifies the statement of the result.

Note that when giving the normality result from [22, Theorem 5.5] we would use $\boldsymbol{\psi}$ rather than $\boldsymbol{\psi}_o$, but as the distribution is determined under the assumption that the null holds, we can replace $\boldsymbol{\psi}$ by $\boldsymbol{\psi}_o$. Noting that

$$
\mathcal{F}_\circ - \mathcal{F}_\times^T \tilde{\mathcal{F}}^{-1} \mathcal{F}_\times = \mathcal{G}_3^{-1}, \tag{51}
$$

we therefore directly arrive at the result by defining the new random vector

$$
\boldsymbol{Z}_1 = N_{|\Omega|}^{1/2} \left(\mathcal{F}_\circ - \mathcal{F}_\times^T \tilde{\mathcal{F}}^{-1} \mathcal{F}_\times \right)^{1/2} \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right), \tag{52}
$$

where we may determine that asymptotically

$$
\boldsymbol{Z}_1 = \mathcal{G}_3^{-1/2} N_{|\Omega|}^{1/2} \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_o \right) \sim N (\mathbf{0}, \mathbf{I}_p),
$$

from which the asymptotic result $\boldsymbol{Z}_1^T \boldsymbol{Z}_1 \sim \chi_p^2$ follows.

REFERENCES

[1] P. J. Brands, A. P. Hoeks, L. A. F. Ledoux, and R. S. Reneman, "A radio frequency domain complex cross-correlation model to estimate blood flow velocity and tissue motion by means of ultrasound," *Ultrasound Med. Biol.*, vol. 23, no. 6, pp. 911–920, 1997.

[2] J. Gonella, "A rotary-component method for analysing meteorological and oceanographic vector time series," *Deep-Sea Res.*, vol. 19, no. 12, pp. 833–846, 1972.

[3] R. Lumpkin and M. Pazos, "Measuring surface currents with surface velocity program drifters: The instrument, its data, and some results," in *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*. Cambridge, U.K.: Cambridge Univ. Press, 2007, ch. 2, pp. 39–67.

[4] A. M. Walker, "Periodogram analysis for complex-valued time series," in *Developments in Time Series Analysis*, T. Subba Rao, Ed. London, U.K.: Chapman & Hall, 1993, pp. 149–163.

[5] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. Hoboken, NJ, USA: Wiley, 2009.

[6] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

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

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

[9] A. G. Davenport, "The spectrum of horizontal gustiness near the ground in high winds," *Quart. J. Roy. Meteorol. Soc.*, vol. 87, no. 372, pp. 194–211, 1961.

[10] J. Calman, "On the interpretation of ocean current spectra. part II: Testing dynamical hypotheses," *J. Phys. Oceanogr.*, vol. 8, no. 4, pp. 644–652, 1978.

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

[12] P. Rubin-Delanchy and A. T. Walden, "Kinematics of complex-valued time series," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4189–4198, Sep. 2008.

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

[14] A. M. Sykulski, S. C. Olhede, and J. M. Lilly, "A widely linear complex autoregressive process of order one," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6200–6210, Dec. 2016.

[15] B. V. Hamon and E. J. Hannan, "Spectral estimation of time delay for dispersive and non-dispersive systems," *J. Roy. Statist. Soc. C*, vol. 23, no. 2, pp. 134–142, 1974.

[16] P. M. Robinson, "Gaussian semiparametric estimation of long range dependence," *Ann. Statist.*, vol. 23, no. 5, pp. 1630–1661, 1995.

[17] D. Gabor, "Theory of communication. Part 1: The analysis of information," *J. Inst. Elect. Eng., Part III, Radio Commun. Eng.*, vol. 93, no. 26, pp. 429–457, Nov. 1946.

[18] L. Cohen, *Time-Frequency Analysis*, vol. 778, Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.

[19] T. Gneiting, W. Kleiber, and M. Schlather, "Matérn cross-covariance functions for multivariate random fields," *J. Amer. Statist. Soc.*, vol. 105, no. 491, pp. 1167–1177, 2010.

[20] P. J. Schreier, "Polarization ellipse analysis of nonstationary random signals," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4330–4339, Sep. 2008.

[21] P. Whittle, "The analysis of multiple stationary time series," *J. Roy. Statist. Soc. B*, vol. 15, no. 1, pp. 125–139, 1953.

[22] K. O. Dzhaparidze and A. M. Yaglom, "Spectrum parameter estimation in time series analysis," in *Developments in Statistics*, P. R. Krishnaiah, Ed., New York, NY, USA: Academic, 1983, pp. 1–96.

[23] P. Whittle, "Estimation and information in stationary time series," *Arkiv för Matematik*, vol. 2, no. 5, pp. 423–434, 1953.

[24] R. Dahlhaus, "Small sample effects in time series analysis: A new asymptotic theory and a new estimate," *Ann. Statist.*, vol. 16, no. 2, pp. 808–841, 1988.

[25] S. L. Marple Jr, "Computing the discrete-time analytic signal via FFT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2600–2603, Sep. 1999.

[26] C. Velasco and P. M. Robinson, "Whittle pseudo-maximum likelihood estimation for nonstationary time series," *J. Amer. Statist. Soc.*, vol. 95, no. 452, pp. 1229–1243, 2000.

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

[28] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge, U.K.: Cambridge Univ. Press, 1993.

[29] J. R. Glover, "Adaptive noise canceling applied to sinusoidal interferences," *IEEE Trans. Acoust. Speech*, vol. AS-25, no. 6, pp. 484–491, Dec. 1977.

[30] B. Hamon and E. J. Hannan, "Estimating relations between time series," *J. Geophys. Res.*, vol. 68, no. 21, pp. 6033–6041, 1963.

[31] G. Wahba, "Automatic smoothing of the log periodogram," *J. Amer. Statist. Assoc.*, vol. 75, no. 369, pp. 122–132, 1980.

[32] P. M. Robinson, "Semiparametric analysis of long-memory time series," *Ann. Statist.*, vol. 22, no. 1, pp. 515–539, 1994.

[33] R. Lumpkin and S. Elipot, "Surface drifter pair spreading in the North Atlantic," *J. Geophys. Res.-Oceans*, vol. 115, no. C12, 2010, Art. no. C12017.

[34] 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.

[35] C. M. Hurvich, R. Deo, and J. Brodsky, "The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series," *J. Time Ser. Anal.*, vol. 19, no. 1, pp. 19–46, 1998.

[36] S. C. Olhede, A. M. Sykulski, and G. A. Pavliotis, "Frequency domain estimation of integrated volatility for Itô processes in the presence of market-microstructure noise," *Multiscale Model. Simul.*, vol. 8, no. 2, pp. 393–427, 2009.

[37] N. Cressie, *Statistics for Spatial Data*. Hoboken, NJ, USA: Wiley, 1993.

[38] E. Masry, "Alias-free sampling: An alternative conceptualization and its applications," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 317–324, May 1978.

[39] E. Ollila and V. Koivunen, "Generalized complex elliptical distributions," in *Proc. Sensor Array Multichannel Signal Process. Workshop*, 2004, pp. 460–464.

[40] A. T. Walden and P. Rubin-Delanchy, "On testing for impropriety of complex-valued Gaussian vectors," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 825–834, Mar. 2009.

[41] S. Chandna and A. T. Walden, "A frequency domain test for propriety of complex-valued vector time series," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1425–1436, Mar. 2017.

[42] J. K. Tugnait and S. A. Bhaskar, "Testing for impropriety of multivariate complex random processes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2016, pp. 4264–4268.

[43] Q. H. Vuong, "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica*, vol. 57, no. 2, pp. 307–333, 1989.

[44] C. M. Hurvich and C. L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, no. 2, pp. 297–307, 1989.

[45] G. K. Vallis, *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[46] J. M. Lilly, A. M. Sykulski, J. J. Early, and S. C. Olhede, "Fractional Brownian motion, the Matérn process, and stochastic modeling of turbulent dispersion," *Nonlin. Processes Geophys. Discuss.*, doi: 10.5194/npg-2017-15, in review, 2017.

[47] J. H. LaCasce, "Statistics from Lagrangian observations," *Prog. Oceanogr.*, vol. 77, no. 1, pp. 1–29, 2008.

[48] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, Sep. 1982.

[49] P. B. Rhines, "Waves and turbulence on a beta-plane," *J. Fluid. Mech.*, vol. 69, no. 3, pp. 417–443, 1975.

[50] D. R. Brillinger, *Time Series: Data Analysis and Theory*. Philadelphia, PA, USA: SIAM, 2001.

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

From 2007 to 2008, he was a Corporate Analyst in Lazard, London, U.K., specializing in corporate mergers and acquisitions. From 2011 to 2014, after receiving the Ph.D. degree, he was a Postdoctoral Research Associate in the Department of Statistical Science, University College London. Since 2014, he has been a Research Fellow in a joint position between University College London, London, and NorthWest Research Associates, Redmond, WA, USA. He is currently funded by a Marie Curie Research Fellowship from the European Commission. In April 2017, he will be starting a new position as a Lecturer in Data Science and Statistics at Lancaster University, Lancashire, U.K. 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.

He is a member of the European Geophysical Union and a Fellow of the Royal Statistical Society.

Jonathan M. Lilly (M'05–SM'12) was born in Lansing, MI, USA, in 1972. He received the B.S. degree in geology and geophysics from Yale University, New Haven, CT, USA, in 1994, and the M.S. and Ph.D. degrees in physical oceanography from the University of Washington (UW), Seattle, WA, USA, in 1997 and 2002, respectively.

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

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

Sofia Charlotta 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 a Senior Lecturer, from 2006 to 2007, in the Mathematics Department, Imperial College London. In 2007, she joined the Department of Statistical Science, University College London, London, U.K., where she is currently a Professor of statistics, an Honorary Professor of computer science, and the Director of Research in the Department of Statistical Science. Her research interests include the analysis of network data, nonstationary time series, inhomogeneous random fields, and multiscale methods.

She is a member of the Programme Committee of the International Centre for Mathematical Sciences. She 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.

Jeffrey J. Early was born in Redmond, WA, USA, in 1980 and raised in Juneau, AK, USA. He received the B.S. degree in mathematics and physics from the University of Puget Sound, Tacoma, WA, USA, in 2002, the M.S. degree in physics from University of Oregon, Eugene, OR, USA, in 2004, and the Ph.D. degree in oceanography from Oregon State University, Corvallis, OR, USA, in 2009.

In 2011, he joined NorthWest Research Associates in Redmond, WA, USA, as a Postdoctoral Researcher, where he has been a Research Scientist since 2012. He also runs a small software company, Early Innovations, that sells photo annotation and geotagging software. His research interests include the dynamics of oceanic eddies and internal waves, numerical modeling methods, and symmetry solutions to partial differential equations.