ANNUAL REVIEWS

Annual Review of Statistics and Its Application

Nonparametric Spectral Analysis of Multivariate Time Series

Rainer von Sachs

Institute of Statistics, Biostatistics and Actuarial Sciences, Louvain Institute of Data Analysis and Modeling in Economics and Statistics, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium; email: rvs@uclouvain.be

Annu. Rev. Stat. Appl. 2020. 7:361–86

First published as a Review in Advance on November 18, 2019

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-031219-041138

Keywords

Fourier, wavelets, locally stationary time series, coherency, smoothing, positive-definiteness

Abstract

Spectral analysis of multivariate time series has been an active field of methodological and applied statistics for the past 50 years. Since the success of the fast Fourier transform algorithm, the analysis of serial auto- and cross-correlation in the frequency domain has helped us to understand the dynamics in many serially correlated data without necessarily needing to develop complex parametric models. In this work, we give a nonexhaustive review of the mostly recent nonparametric methods of spectral analysis of multivariate time series, with an emphasis on model-based approaches. We try to give insights into a variety of complimentary approaches for standard and less standard situations (such as nonstationary, replicated, or high-dimensional time series), discuss estimation aspects (such as smoothing over frequency), and include some examples stemming from life science applications (such as brain data).

1. INTRODUCTION

Spectral analysis of multivariate time series has been an active field of methodological and applied statistics for the past 50 years. Since the success of the fast Fourier transform (FFT) algorithm (Cooley & Tukey 1965), the analysis of serial auto- and cross-correlation in the frequency domain has helped us to understand the dynamics in many serially correlated data in a way that does not require the development of complex parametric models. On the one hand, studying the linear dependence structure of the components of a multivariate time series over time can reveal a first idea about aspects such as causality. Understanding the leads and lags between two components of the observed series is often of prime interest, for, e.g., economic indices. On the other hand, in particular for data from the life sciences (medical, geological, etc.), spectral analysis can serve as a preparatory tool for developing models such as those for the connectivity in the human brain measured via multi-channel electroencephalogram (EEG). Another application can be found for temperature (and even climate) observations showing pronounced temporal cycles of random nature.

Today, these second-order analyses are well developed and understood for stationary time series data, i.e., data with a linear dependence structure that does not change over time. However, they are not sufficient to realistically describe many real-life phenomena that show transients, regime changes over time, and so on. Hence, methodological research started, about 20 years ago, to generalize spectral analysis in a way that would still allow the use of the FFT (at least) locally over time. While these ideas had already been around in the engineering literature for quite a while, statisticians only later started to develop models that would accompany proposed modern methodologies (such as local FFT, spectrogram, Wigner–Ville, or wavelet transform), allowing the derivation of their theoretical properties.

In this overview, we concentrate on discussing a few of these recent, mostly model-based approaches in nonparametric spectral analysis of uni- and multivariate time series. We restrict ourselves to weakly dependent time series models (i.e., those with sufficiently fast decaying serial dependencies), but in order to go beyond stationarity, we also discuss time-varying spectral analysis in some detail. More specifically, we intend to address the following aspects of modern spectral analysis:

1. How do we smooth the (usually pretty irregularly appearing) raw Fourier-based estimators over frequency by methods that stem more generally from nonparametric curve estimation (see, e.g., Moulin 1994 on log-periodogram smoothing with wavelets; Neumann 1996 on classical wavelet smoothing; Rosen & Stoffer 2007 on spline smoothing; and Yuan et al. 2012, as well as Chau & von Sachs 2019, on preserving positive-definiteness of the resulting spectral estimators)?
2. What concepts do exist to render classical spectral analysis time-varying to adapt to the often inhomogeneous nature of the dynamics of time series (Dahlhaus 2012, Ombao et al. 2005, Neumann & von Sachs 1997, Nason et al. 2000)?
3. How can we perform spectral analysis in the presence of replicated time series in more complex situations of, e.g., supervised brain signal experiments (Chau & von Sachs 2016, Fiecas & Ombao 2016, Gorrostieta et al. 2019)?
4. What are the challenges of spectral analysis of high-dimensional time series (Böhm & von Sachs 2009, Fiecas & Ombao 2011, Fiecas & von Sachs 2014)?
5. What approaches exist that go beyond the classical Fourier-based spectral analyses (Davis et al. 2013, Dette et al. 2015), and how can we do spectral analysis for point processes (Roueff & von Sachs 2019)?

Our choice of topics to cover is certainly biased and cannot be exhaustive. We opt to discuss almost exclusively nonparametric approaches, which borrow strength from other fields of research (curve estimation, quantile regression, regularization methods, random effects modeling, etc.). As such, they allow for some theoretical development of the mostly asymptotic properties of the proposed estimation schemes. We give some critical views on the different methods, some of which have been developed from a more theoretical point of view of new model approaches, others of which are likely to be more suited to concrete applications.

With this choice, we apologize that we must omit many other interesting topics, such as semiparametric spectral analysis, goodness of fit of spectral distributions, testing, graphical time series models for partial correlation/coherence analysis, and many more (some additional references are provided in Section 7).

2. SOME BACKGROUND ON SPECTRAL ANALYSIS

There are many good books (Brillinger 1981, Brockwell & Davis 1991, Shumway & Stoffer 2006, and Koopmans 1974, to name but a few) that give a detailed introduction to spectral analysis of multivariate time series. In this section, we provide formal background and examples that lay the groundwork for the following review of modern nonparametric methods applicable to spectral analysis. We start by setting up notation for the treatment of stationary time series.

2.1. Stationary Time Series

Let $\mathbf{X}(t)$, $t = 1, \ldots, T$, be a discrete real-valued, zero-mean, weakly stationary time series of dimension P with an absolutely summable autocovariance function, elementwise for each entry of the following matrix (where $^\top$ denotes transposed vectors):

$$\mathbf{c}_X(h) = \mathbb{E}(\mathbf{X}(t)\mathbf{X}(t+h)^\top), \quad h = 0, \pm 1, \pm 2, \ldots.$$

Its elements $c_{ij}(h), i, j = 1, \ldots, P$, measure the linear dependency between components $X_i(t)$ and $X_j(t+h)$, which is invariant of time t, due to the stationarity of the underlying time series. As it is somewhat cumbersome to represent the whole (uni- or multivariate) autocovariance structure of the time series as an (infinite) sequence of time lag h, an alternative representation of this linear dependency is given in the frequency domain by the spectrum or spectral density of $\mathbf{X}(t)$.

The $P \times P$ spectral density matrix $\mathbf{f}(\omega)$ of $\mathbf{X}(t)$ is defined to be the Fourier transform (Fourier series) of $\mathbf{c}_X$, i.e.,

$$\mathbf{f}(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \mathbb{E}(\mathbf{X}(t)\mathbf{X}(t+h)^\top) \exp(-i\omega h), \qquad 1.$$

for (angular) frequencies $\omega \in [-\pi, \pi]$. The spectral density matrix is hence 2π-periodic and hermitian, i.e., complex conjugate symmetric about zero frequency. It is composed of the real-valued autospectra $f_{ii}(\omega), i = 1, \ldots, P$, on the diagonal and the (complex conjugate) cross-spectra on its off-diagonals $f_{ij}(\omega), i \neq j = 1, \ldots, P$.

The autospectrum $f_{ii}(\omega)$ represents the linear serial dependency in the frequency domain of component $X_i(t)$. It can be seen as arising in a decomposition of both the underlying process and its variance-covariance along sine and cosine waves with random amplitudes (and phase)

$$X_i(t) \approx \sum_k [U_{ik} \cos(\omega_k t) + V_{ik} \sin(\omega_k t)], \quad \omega_k = \frac{2\pi k}{T}, \qquad 2.$$

where U_{ik}, V_{ik} are independent and identically distributed (i.i.d.) $(0, \sigma_{ik}^2)$ such that

$$\mathrm{Var}\,(X_i(t)) \approx \sum_k \sigma_{ik}^2, \tag{3.}$$

and

$$c_{ii}(h) \approx \sum_k \sigma_{ik}^2 \cos(\omega_k h). \tag{4.}$$

Rigorously written, and generalized to all components of $\mathbf{X}(t)$, the last equation translates into the autocovariance matrix $\mathbf{c}_X(h)$ being the Fourier back transform of the spectral density,

$$\mathbf{c}_X(h) = \int_{-\pi}^{\pi} \mathbf{f}(\omega) \exp(i\omega h)\, \mathrm{d}\omega. \tag{5.}$$

Note, furthermore, that the cross-spectrum $f_{ij}(\omega)$ measures the linear dependency at all lags and leads between components $X_i(t)$ and $X_j(t)$. For example, if one component is an exact shift τ in time of the other, i.e., $X_i(t) = X_j(t + \tau)$, then the cross-correlation $\rho_{ij}(h)$ (i.e., the cross-covariance c_{ij}, normalized by the autocovariances c_{ii} and c_{jj}) has a pronounced peak (only) at $h = \tau$. Then the cross-spectrum is peaked at the frequency corresponding to this wave- or correlation length.

The following running example has been simulated using the R package `pdSpecEst` by Chau (2018). It represents a bivariate vector autoregressive-moving average (VARMA) process of order (2,2) based on Gaussian white noise. Its parameters have been chosen such that the first and second autospectra correspond to a low-frequency process such as MA(1) with a smooth peak at zero frequency and an AR(2) process with a sharper peak at frequency $\pi/2$ shared by the cross-spectra. Note that the plots in **Figure 1** only show the frequency range from $[0, \pi]$ as all spectra are symmetric about zero frequency. We also clearly observe the symmetry properties of real and imaginary parts of the cross-spectra f_{12} and f_{21}.

To estimate $\mathbf{f}(\omega)$ nonparametrically, one usually first converts the data $\mathbf{X}(t)$ from the time domain to the frequency domain using the discrete Fourier transform:

$$\mathbf{d}_X(\omega) = \sum_{t=1}^{T} \mathbf{X}(t) \exp(-i\omega t). \tag{6.}$$

In practice $\mathbf{d}_X(\omega)$ is fastly calculated [of order $T \log(T)$] via the famous FFT algorithm of Cooley & Tukey (1965) at the collection of grid or Fourier frequencies $\omega = \omega_\ell = \frac{2\pi\ell}{T}$, $\ell = -T/2, \ldots, T/2$.

Then, the periodogram matrix $\mathbf{I}_T(\omega)$ is

$$\mathbf{I}_T(\omega) = (2\pi)^{-1} T^{-1} \mathbf{d}_X(\omega) \mathbf{d}_X(\omega)^*, \tag{7.}$$

where $()^*$ denotes the complex conjugate transpose. It is well known that for weakly dependent processes [essentially, with elementwise absolutely summable $\mathbf{c}_X(h)$ over lag h] the periodogram matrix is an asymptotically unbiased but inconsistent estimator for $\mathbf{f}(\omega)$ (Brillinger 1981), i.e., as $T \to \infty$,

$$E(\mathbf{I}_T(\omega)) \to \mathbf{f}(\omega), \tag{8.}$$

$$\mathrm{Var}(\mathbf{I}_T(\omega)) \to \mathbf{f}^2(\omega), \quad \omega \neq 0, \pm\pi. \tag{9.}$$

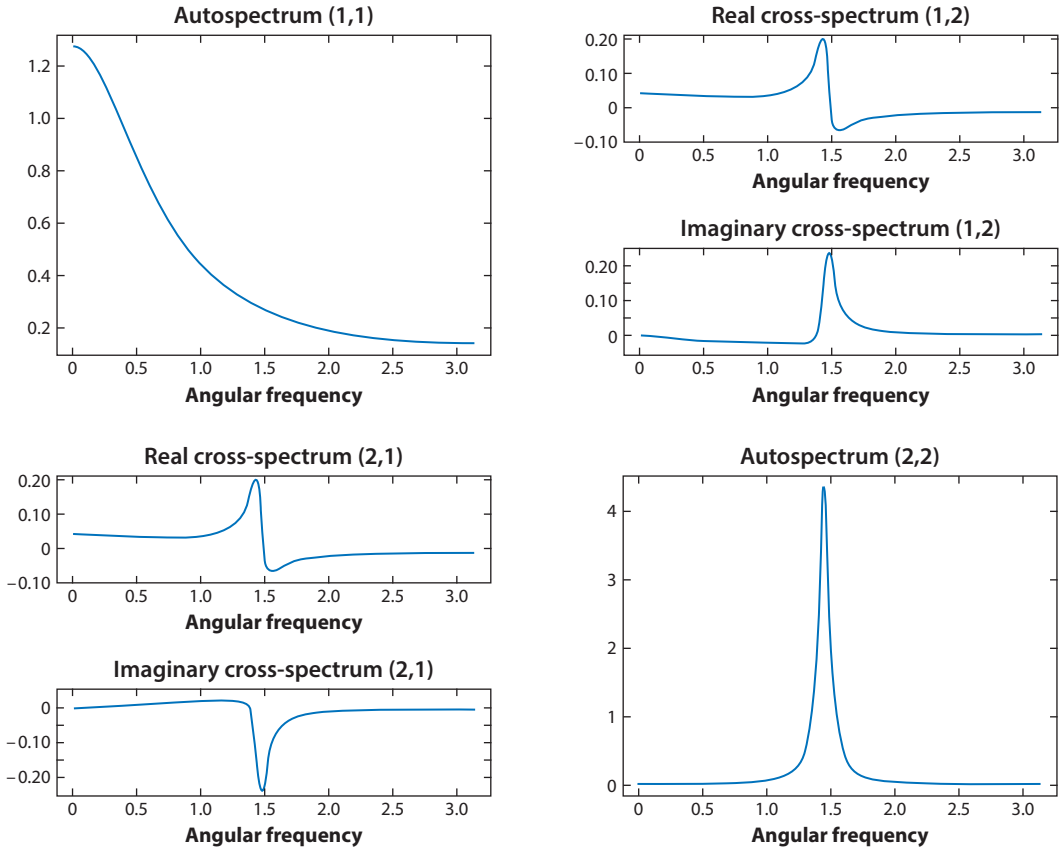

Figure 1

Population spectral matrix of a simulated VARMA(2,2) process. Abbreviation: VARMA, vector autoregressive-moving average.

In **Figure 2**, we show a periodogram for a simulated stretch of length $T = 4,096$ of the same VARMA process as before. As can be seen from the wildly erratic behavior of the periodogram, it is necessary to smooth each element (j,k) of $\mathbf{I}_T(\omega)$ over frequency. A popular nonparametric method is to use a smoothing kernel $K_T^{(jk)}(\cdot)$ whose smoothing span is $M_T^{(jk)}$:

$$\widetilde{f}_T^{(jk)}(\omega) = \int_{-\pi}^{\pi} K_T^{(jk)}(\omega - \theta)\, \mathbf{I}_T^{(jk)}(\theta)\, \mathrm{d}\theta, \qquad 10.$$

where the integration (expressing a convolution of the periodogram with the kernel weight that asymptotically concentrates on frequency ω) is replaced in practice by a summation over the grid of Fourier frequencies. The smoothing span $M_T^{(jk)}$ essentially represents the number of periodogram ordinates over which we smooth. As it tends asymptotically to infinity, however, by a slower rate than sample size T, under certain regularity conditions on the population spectrum $\mathbf{f}(\omega)$, the resulting smoothed periodogram matrix $\widetilde{\mathbf{f}}_T(\omega)$ becomes a consistent estimator of the latter (Brillinger 1981). The form of its (asymptotic) variance, as a consequence of Equation 9, presents another certain challenge,

$$\mathrm{Var}\left(\widetilde{f}_T^{(jk)}(\omega)\right) \sim \frac{1}{M_T^{(jk)}} \int_{-\pi}^{\pi} [K_T^{(jk)}(\theta)]^2 \mathrm{d}\theta\, [f^{(jk)}]^2(\omega). \qquad 11.$$

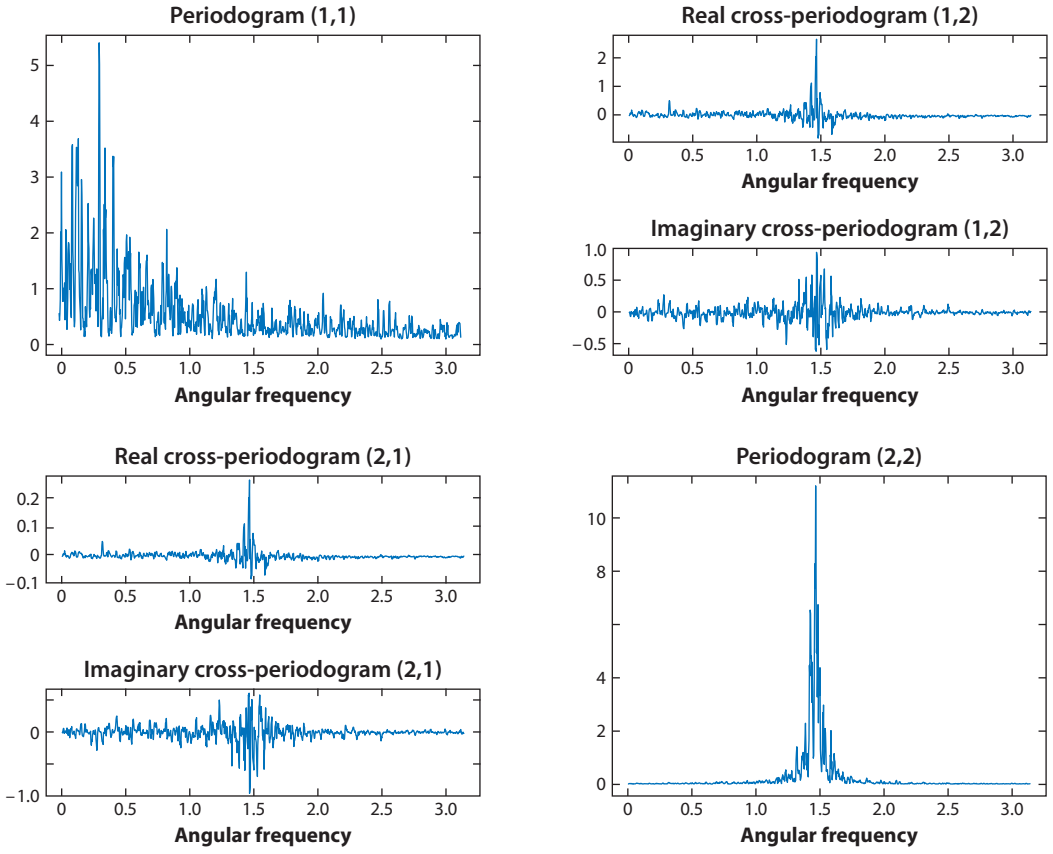

Figure 2

Periodogram matrix of the simulated VARMA(2,2) process with spectrum shown in **Figure 1**. Abbreviation: VARMA, vector autoregressive-moving average.

As we can observe, this variance, although asymptotically tending to zero, depends, again, on the (unknown) squared spectrum. This poses problem for questions of optimally choosing the span $M^{(jk)}$ [e.g., via mean-squared error (MSE) minimization, for which we refer to the discussion in Section 2.2], and also for subsequent inference (confidence intervals or tests).

Of course, many other smoothing techniques exist besides kernel smoothing of the periodogram matrix. Walden (2000) gives a nice overview on consistent multivariate spectral estimation, starting from the multitaper point of view. However, he adopts a unified view that includes many other techniques and that can also address the (non-)invertibility of the periodogram, which is a rank-1 matrix (as can be seen directly from Equation 7). Multitapers have been used, particularly by the engineering community, since the seminal work by Thomson (1982): This technique provides simultaneously for reducing spectral leakage and averaging over a series of uncorrelated periodogram estimates due to the construction of orthogonal taper windows, which are successively applied to the time series data before Fourier transforming.

Figure 3 shows an example for smoothing the periodogram matrix of **Figure 2** over frequency. Here, we used the more sophisticated wavelet method of Section 3.2. This method automatically adapts to the different degree of smoothness over frequency of the different components of the underlying population spectral matrix (displayed in **Figure 1**) despite using only one global

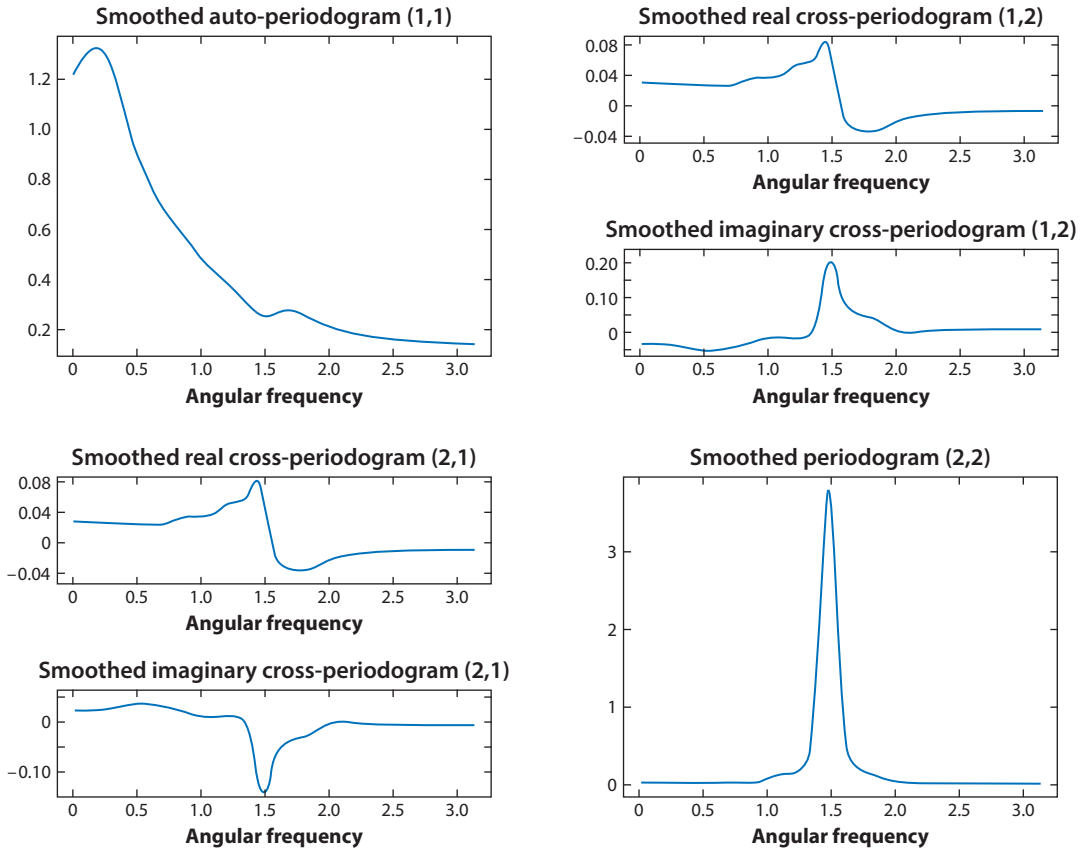

Figure 3

Smoothed periodogram matrix of the simulated VARMA(2,2) process with spectrum shown in **Figure 1**. Abbreviation: VARMA, vector autoregressive-moving average.

smoothing parameter. For information on its choice, as well as details on the actual wavelet smoother, we refer the reader to Sections 3.1 and 3.2. Furthermore, in order to do inference, asymptotic normality of the smoothed periodogram can be used (more information on this is provided in Section 2.3.2).

For the analysis of the cross-spectral information, rather than looking at the real and imaginary part of the off-diagonal elements of the estimated spectral density matrix, it is more useful to look at the amplitude and phase of these elements (see, e.g., Brockwell & Davis 1991, section 11.6). A more direct measure of the amount of cross-correlation of two components of $\mathbf{X}(t)$ in the frequency domain is the spectral coherence at a given frequency ω:

$$\kappa_{ij}(\omega) = f_{ij}(\omega)/[f_{ii}(\omega)f_{ij}(\omega)]^{1/2}.$$

Estimators of this quantity can be obtained by replacing the population spectra by frequency smoothed cross- and auto-periodograms, the respective elements of Equation 7. It is more common to look at the squared coherency function $|\kappa_{ij}(\omega)|^2$, the analog of the squared correlation in the frequency domain. Here, a value near one indicates a strong linear relationship at frequency ω between $dZ_i(\omega)$ and $dZ_j(\omega)$, the increment processes of X_{ti} and X_{tj} in the multivariate analog of

the spectral decomposition (Equation 13) given below. The reader is referred to Section 2.3.2 for an example of time-varying coherence analysis of a multivariate EEG (see **Figure 5**, below). As a first illustration, we refer to a simple bivariate example taken from Brockwell & Davis (1991, their example 11.6.1), where $X_{t1} = Z_t$, $X_{t2} = Z_t + 0.75Z_{t-10}$, for a given white noise process $\{Z_t\} \sim \mathrm{WN}(0, 1)$. In this case, $|\kappa_{ij}(\omega)|^2 = 1$ for all frequencies ω, which is actually more generally the case if $\{X_{t1}\}$ and $\{X_{t2}\}$ are related by any time-invariant linear filter.

For more details on spectral coherence analysis and another (econometrics) example about a price time series led by a market indicator series by several units in time, we refer again to the book by Brockwell & Davis (1991) and example 11.6.3 therein. A more comprehensive analysis of a multivariate monthly mean temperature time series of dimension 14 can be found in Brillinger (1981, e.g., figure 7.8.9). The general interest is to identify the frequencies (or frequency bands) that drive the linear relationship between two components of the multivariate time series. For this, proper inference on the squared coherency has to be performed. One possibility is based on, again, asymptotic normality (with an appropriate bias correction; see Koopmans 1974) for tests or confidence intervals, frequency by frequency (see, e.g., Brockwell & Davis 1991, figure 11.7, and figure 11.8 for the estimated phase).

Finally, it is also common to look at partial coherence (as in Park et al. 2014) as the frequency domain analog of partial correlation. For this concept, widely used for studying brain connectivity, for example, we refer to Brillinger (1981, sections 8.3 and 8.4).

2.2. The Challenges in Smoothing the Periodogram Matrix

Classical textbooks on modern spectral analysis suggest using one of the standard nonparametric curve smoothing techniques, such as the kernel, nearest neighbor, spline, or more recently, wavelet techniques. It is not the goal of this review to furnish any details on these well-documented methods. However, we provide some preliminary comments in order to lay the groundwork for treating the nonstandard but often very realistic situation of, possibly also time-varying, spectra showing spatially localized features (over frequency and/or over time). This is as opposed to handling the much easier situation of smoothing the log-periodogram of a univariate stationary time series based on the well-known Wahba approximation (Wahba 1980):

$$\log(I_T(\omega)) \approx \log(f(\omega)) - \gamma + e(\omega), \qquad 12.$$

with γ denoting the Euler-Mascheroni constant ($= 0.57722...$) and with mean centered, constant variance i.i.d. errors $e(\omega)$. Obviously this classical relationship of noisy observations modeled as a population target plus homoscedastic additive noise fails to hold more generally. This makes it considerably more difficult to smooth multivariate periodograms over frequency (see the discussion around Equation 11):

1. The first reason is that, as is the univariate periodogram, the elements of the $\mathbf{I}_T(\omega)$ are known to be somewhat nonnormally distributed (i.e., Wishart in the multivariate case as a generalization of χ^2 in the univariate case), and this only approximately so, for sufficiently large T. Hence, any (linear) smoothing technique that essentially averages periodograms either over neighboring Fourier frequencies or, in the stationary case, over neighboring blocks over time (Welch 1967) works under the paradigm of an asymptotically, not too quickly growing window. Essentially, consistency can be achieved if the smoothing span $M_T \to \infty$ but $M_T/T \to 0$ as $T \to \infty$. Usually asymptotic normality of the spectral estimators also follows under these conditions. In theory, these results are completely analogous to classical curve estimation under an additive signal-plus-noise model (such as the one in Equation 12) due to the (asymptotic) uncorrelatedness of the $\mathbf{I}_T(\omega_\ell)$ over the grid

of Fourier frequencies. Nevertheless, in practice it might be interesting to work with an intermediate approximation based on weighted sums of χ^2 ordinates instead of basing inference on asymptotic normality of the smoothed periodogram. This is referred to as the Satterthwaite approximation (see, e.g., Brockwell & Davis 1991, chapter 10.5, along with figure 10.9 therein).

2. A second, and more severe, complication for periodogram smoothing comes from the heteroscedasticity of the (smoothed) periodogram, as its asymptotic variance depends on the squared value of the spectral density at frequency ω (see Equations 9 and 11). This poses problems for choosing the smoothing parameter M_T (or the bandwidth) via MSE minimization, whether in the case of classical kernel smoothing or even for assessing the variance of the empirical wavelet coefficients in case of a wavelet smoother (Neumann 1996), which we revisit in Section 3.1.
3. A final, and most challenging, obstacle comes from the nature of a spectral density matrix to be positive definite (PD), a property that is widely used in time series analysis for exploiting the link between autocovariance and its Fourier transform (see Equation 5). Hence, one wishes the estimated spectrum to remain PD—which is challenging due to the fact the periodogram is a rank-1 matrix (as can be seen directly from Equation 7). We come back to this challenge in Section 3.2 on positive-definiteness-preserving wavelet estimation.

In the next subsection, these aspects become even more challenging when, e.g., trying to smooth a time-varying periodogram, such as in Equation 14, not only over frequency ω but also over time. This can be done, e.g., by a two-fold kernel smoother, as sketched out below in Equation 15.

2.3. Time-Varying Spectral Analysis

The spectral analysis discussed above is shift invariant with respect to time; however, many time series data sets (speed and sound, geophysical or climate data, EEG or electrocardiogram data, etc.) show a time-varying second-order behavior, possibly even with abrupt transients. Many approaches of generalizing to time-dependent spectral analysis can be found in the literature; one of the first in statistical time series analysis was also the most prominent (Priestley 1965). For this review, we focus on two more recent and complementary representatives, which both are statistical and somewhat model-based approaches built on the idea of a short-time Fourier transform (using a windowed or segmented periodogram). Other important and statistically interesting approaches do exist, such as those based on the Wigner–Ville spectrum, a very localized time-frequency spectrum that does not require choosing an a priori segmentation. For details on this approach, which requires more sophisticated smoothing operations over time and frequency, we refer the reader to Martin & Flandrin (1995) and Neumann & von Sachs (1997); Dahlhaus (2000a) provides a parametric approach that is based on the Whittle likelihood.

2.3.1. Locally stationary time series. We start from the rigorous analog of the motivating Equation 2. The so-called Cramér spectral representation of a stationary time series $X(t)$ with spectral density $f(\omega)$ reads in its univariate version as

$$X(t) = \int_{-\pi}^{\pi} A(\omega) \exp(i\omega t)\, \mathrm{d}Z(\omega). \qquad 13.$$

In this equation, the transfer function A is related to the spectrum $f(\omega)$ by the relation $f(\omega) = |A(\omega)|^2$, whereas $\mathrm{d}Z(\omega)$ denotes an orthonormal increment process, i.e., a process with variance one and uncorrelated increments: $\mathrm{Cov}(\mathrm{d}Z(\omega), \mathrm{d}Z^*(\omega')) = \delta(\omega, \omega')$. [Note that often in

the literature we find this representation given in a form based on orthogonal increments, i.e., normalized such that $\mathrm{Var}(dZ(\omega)) = f(\omega)$.] The multivariate spectral representation of $\mathbf{X}(t)$ is analogous (see Brockwell & Davis 1991, chapter 11.8) and is not shown here to simplify the exposition.

Dahlhaus wrote a series of seminal papers on locally stationary time series, of which we cite only the first two related to the frequency domain: Dahlhaus (1997) for the univariate approach and Dahlhaus (2000a) for the case of multivariate processes. An excellent overview of his and all related historical work in the literature is given by Dahlhaus (2012). Motivated by Priestley (1965), Dahlhaus came up with the class of locally stationary processes by imposing a smooth variation in time t on the now time-dependent transfer function $A_t(\omega)$, resulting in a model quantity $A(u, \omega)$ [with $A(t/T, \omega) \approx A_t(\omega)$ for large T], which is defined to live on rescaled time $u = t/T \in [0, 1]$. With this trick, the underlying observed processes is embedded into a doubly indexed sequence of processes $\{X_T(t)\}$. Its time-changing spectral content is now characterized by the sequence $A_{t,T}(\omega)$. This allows for the property that the limiting spectrum $f(u, \omega) := |A(u, \omega)|^2$, for each fixed u, represents the spectral density of a truly stationary process that approximates in a controlled way the spectral density $f_{t,T}(\omega) = |A_{t,T}(\omega)|^2$ of $\{X_T(t)\}$. The advantage of this model-based approach is that the uniquely defined evolutionary spectrum $f(u, \omega)$ can be estimated, as a function defined on the (rescaled) time-frequency plane $[0, 1] \times [-\pi, \pi]$, from sequences of time-localized periodograms along segments of length N (smaller than T):

$$I_{N,T}(u, \omega) = (2\pi)^{-1} N^{-1} \left| \sum_{s=-N/2+1}^{N/2} X_T([uT] + s) \exp(-i\omega s) \right|^2, \; t = [uT] \in [N/2, T - N/2]. \quad 14.$$

The direct generalization of smoothing periodograms of stationary time series given by Equation 10 amounts to simply applying ideas of (kernel-based) smoothing of two-dimensional nonparametric regression curves. In the case of a univariate time series this results in the estimator

$$\widehat{f}(u, \omega) = \sum_i K_b^t(u - u_i) \int_{-\pi}^{\pi} K_h^f(\omega - \theta)\, I_{N,T}(u_i, \theta)\, d\theta, \quad 15.$$

where K^t and K^f denote some classical kernel weights in time and frequency, with bandwidths b and h respectively (e.g., $h = h_T = M_T^{-1}$, the reciprocal of the aforementioned smoothing span M_T) and using the common notation $K_b(\cdot) := b^{-1} K(\cdot/b)$. Subsequent theories of consistency, for bandwidths b and h tending sufficiently fast to zero with sample size T (with rates of convergence given by, e.g., Dahlhaus 2012, theorem 4.7), and also inference, have thus been made possible.

Concerning other concepts of local stationarity, we mention a series of papers by Wu and coauthors. Wu & Zhou (2011) is the first paper to apply a powerful new concept based on Bernoulli shift processes, in combination with the functional dependence measure, to describe an alternative dependence structure for local stationarity. Furthermore, Wu & Zaffaroni (2018) establish the asymptotic theory for spectral density estimation of general stationary multivariate time series. Zhang & Wu (D. Zhang and W.B. Wu, manuscript in preparation) treat high-dimensional locally stationary processes. A recent work by Dahlhaus et al. (2019) gives a generalization toward treating nonlinear processes by combining the two aforementioned concepts of local stationarity with the use of stationary approximations and derivative processes (as first introduced in the context of time-varying autoregressive conditional heteroscedasticity processes by Dahlhaus & Subba Rao 2006).

2.3.2. Smoothed local complex exponentials analysis—a particular localized periodogram approach. A different approach, which is less motivated from nonparametric estimation theory, is directly built on a computationally efficient modification of the time-localized periodogram

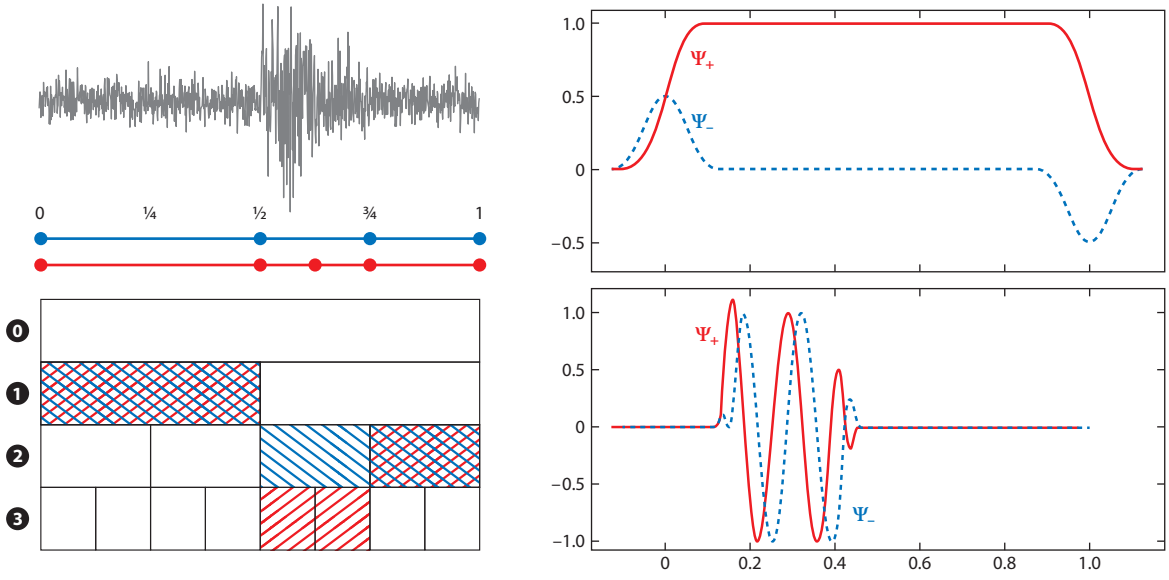

Figure 4

(*Left*) Illustration of adaptively segmenting a nonstationary (electroencephalogram) time series signal. (*Right*) Smoothed local complex exponentials (SLEX) window function (*top*) and a time-localized waveform of the SLEX basis (*bottom*).

(Equation 14). The idea of a series of papers by Ombao and coworkers (Ombao et al. 2001, 2002, 2005, as well as Huang et al. 2004 and Böhm et al. 2010) is to split the observation interval $[0, T]$ in the time domain into a hierarchy of dyadic subintervals and compute particularly windowed periodograms on those, in a fast structured way. Related to this approach are similar ideas developed by, among others, Adak (1998) and Davis et al. (2006), and in a Bayesian framework, based on reversible jump Markov chain Monte Carlo methods, by Rosen et al. (2012), thus circumventing the dyadic restriction, though without treating cross-spectral analysis.

In a fashion similar to the construction of ordinary periodograms, Ombao et al. project the data on a particular local Fourier basis called SLEX (smoothed local complex exponentials). In order to get nonredundant information in this collection of SLEX periodograms, and to allow fast computation along the dyadic blocks of the resulting SLEX library, a specific window function (the SLEX window) is used, which acts like a taper. This smooth window not only reduces the spectral bias that would occur with rectangular windows but also preserves orthogonality (uncorrelatedness) between the segmented periodograms at each Fourier frequency (see **Figure 4**).

The key ingredient of this method is the use of an entropy-related cost function in the time-frequency domain. Its optimum delivers the best-adapted segmentation of the underlying time series, taking both (estimated and frequency-smoothed) auto- and cross-spectral content into account. The method, originally developed for bivariate processes (Ombao et al. 2001), allows a truly multivariate extension based on principal component analysis (PCA) in the SLEX frequency domain (Ombao et al. 2005). In **Figure 5**, we show the idea behind this method in a simplified scenario (for exposure) to only a bivariate set of transient EEG signals of length $T = 2^{15}$ recorded with sampling rate 100 Hz at two different positions on the scalp (the channels corresponding to left parietal lobe P3 and left temporal lobe T3) during an epileptic seizure.

Note how both the spectral structure and the coherency change over time according to the region of onset of the seizure. Moreover, starting from this moment, the two channels T3 and P3,

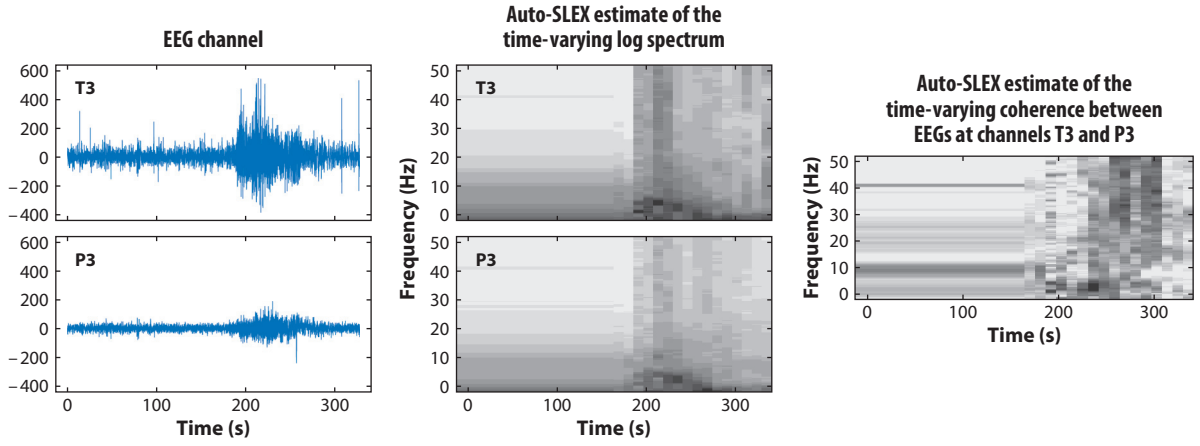

Figure 5

Bivariate SLEX analysis of EEG T3 and P3 channels during an epileptic seizure: autospectra and coherency estimators. Grayscale amplitudes vary from white (low) to black (high) values of the time-varying estimators. Abbreviations: EEG, electroencephalogram; SLEX, smoothed local complex exponentials.

which had previously been in coherent states along a low-frequency band (~10 Hz), become less uniformly coherent, with spectral power shifting to higher frequency. We refer the reader again to Ombao et al. (2005) for a more comprehensive analysis of the full 18-dimensional EEG signal, using a SLEX analysis on the first few principal components of this multivariate data set in the frequency domain.

The SLEX method obviously suffers from the restriction to only provide dyadic partitionings. However, its direct analog to stationary spectral analysis (on each of the chosen dyadic blocks in time) makes it appealing and intuitive for practitioners such as medical doctors in brain data analysis.

3. MODERN NONPARAMETRIC METHODS FOR SMOOTHING THE PERIODOGRAM

Having introduced the challenges in smoothing the periodogram matrix in Section 2.2, we now discuss some solutions based on modern nonparametric smoothing methods (kernel, spline, wavelet, and so on) adapted to the situation at hand. Section 3.1 aims to provide the reader with a nonexhaustive overview of generic smoothers over frequency (not specific to the multivariate nature of the periodogram matrix), with a certain emphasis on the author's own favorite (wavelet) smoothers. In Section 3.2, we illustrate a recent potential solution to the most challenging problem in periodogram smoothing, the problem of preserving positive-definiteness of the resulting spectral estimator.

3.1. Using Wavelet Thresholding Compared with Linear Smoothers (Kernel, Spline, and Others)

Classical nonparametric curve smoothers have been successfully used to address the problem of smoothing the periodogram matrix. Numerous methods exist, such as kernels with a global or local bandwidth, nearest-neighbor methods, smoothing splines, or local polynomials (see Fan & Gijbels 1992 for an excellent overview), in both frequentist and Bayesian versions. It is not the goal of this review article to compare them to each other or discuss how to optimally choose their smoothing

parameter; we simply note that one can directly borrow strength from what has been developed in the classical additive signal-plus-noise model $Y_t = m(x_t) + \varepsilon_t$ (compare also Equation 12) if one takes into account the additional challenges discussed in Section 2.2. We recall, however, that the heteroscedasticity of the periodogram makes it potentially rather difficult to produce visually appealing plots of smooth spectral estimates without oversmoothing local structure (e.g., in the peaks or troughs, as in the spectra shown in **Figure 1**).

A solution to this problem would be to work with a local smoothing parameter (Bühlmann 1996). However, this involves a large amount of computational effort in the case of a multivariate time series with high dimensionality P, and it also makes the data-driven smoothing method highly nonlinear in the data.

For this reason, working with a different, nonlinear smoother based on wavelet technology for univariate periodograms began to attract interest about 20 years ago (see, e.g., Neumann 1996, Gao 1997, and Moulin 1994 for log-periodograms under the model in Equation 12). For details on wavelet smoothing, readers may refer to the cited literature; here, we summarize what is necessary for grasping the ideas relevant to this review. Whereas linear wavelet smoothing is essentially equivalent to kernel smoothing with a fixed bandwidth, the full potential of wavelets arises only if combined with thresholding the wavelet coefficients. These coefficients are, in general, obtained by projecting the observed data, here the periodogram, onto a given wavelet basis, e.g., orthogonal Daubechies wavelets with compact support (Nason 2008) and then suppressing those that are too small in amplitude. Setting these to zero should get rid of the noise in the data because, under the model assumption of a sparse signal (the true underlying curve, here the spectrum, has few nonzero coefficients), this signal is compressed into the few large coefficients that survive thresholding. Of course, the choice of an optimal threshold is crucial and has been object of quite a bit of research. For periodogram smoothing, it represents an additional challenge due to its (finite sample) nonnormality and also its heteroscedasticity (as discussed in Section 2.2). This essentially calls for level- and location-dependent thresholds: Each wavelet coefficient will be thresholded adaptively by its own rule. Unfortunately, and similarly to Equation 11, a plug-in estimator of the square of the unknown spectrum would be necessary, as in Neumann (1996); a recent alternative could be the method proposed by Subba Rao (2018).

Various refined wavelet threshold schemes, e.g., based on nondecimated wavelets (Nason & Silverman 1995), tree thresholding (Freyermuth et al. 2010), or the wavelet-Fisz approach (Fryzlewicz et al. 2008), have been proposed to better cope with this aforementioned problem. From **Figure 3**, one can get an impression of the potential of this methodology to produce smoothed versions of potentially very irregular curves by choosing, essentially, only one smoothing parameter (the threshold).

3.2. Spectral Estimators Preserving Positive-Definiteness

For smoothing the multivariate periodogram, the question of how to optimally choose the smoothing parameter gives us the additional challenge of balancing between full flexibility (each element of the periodogram matrix, potentially at each frequency, gets its own optimal smoothing parameter to choose) and preserving positive-definiteness. Remember that positive-definiteness, similarly to positivity of the spectral density for univariate time series, is important for both computation and interpretation (after application of a PCA, for example, the spectrum of each component should remain positive). As we have shown that wavelet smoothers can be spatially adaptive to local structure without requiring a local smoothing parameter, we now investigate their suitability for preserving positive-definiteness. In a series of papers, Chau & von Sachs (2018, 2019) and Chau et al. (2019) suggested embedding the problem more generally into curve estimation on

Riemannian manifolds—an elegant approach that intrinsically preserves, among other interesting properties, positive-definiteness of the resulting estimators. Partly inspired by the related work of Rahman et al. (2005), Chau and von Sachs successfully implemented their idea by constructing a wavelet-based algorithm that operates directly on matrix-valued curve denoising (or curve-valued matrix estimation). In other words, their constructed wavelet transform is no longer scalar (as for univariate curve estimation problems) but matrix-valued. It nicely uses the structure of the space of symmetric (or Hermitian) PD matrices, which is in fact a Riemannian manifold (this property is the key to success, as discussed below). Distances in this space are no longer Euclidean, as they are in classical wavelet-based algorithms, and hence second-generation-type wavelet algorithms (Jansen & Oonincx 2005) need to be used: Chau and von Sachs found out that, contrary to the approach of Rahman et al. (2005) built on midpoint interpolation, using average interpolation algorithms mimics and transfers many of the advantageous properties of scalar to matrix-valued wavelet thresholding. Their generic approach, which is not tailored either to the spectral estimation problem or to using wavelets, allows for the use of other distances such as the log-Euclidean or the Cholesky (see also Yuan et al. 2012). However, only the Riemannian distance shares all desirable properties to provide, beyond positive-definiteness, for equivariant estimators with respect to affine transformations of the original time series data: permutation of the components of the multivariate series, or more general transformations of the coordinate system. This property, important in practice, has not been shared by alternative, non-wavelet-based approaches such as those of Krafty & Collinge (2013), Rosen & Stoffer (2007), or Dai & Guo (2004) and Guo & Dai (2006) (the latter two are based on smoothing the Cholesky square root of the periodogram matrix).

To illustrate the properties of the aforementioned method (Chau & von Sachs 2019), we revisit the bivariate brain data set (local field potentials), which is also used in Section 4 in the context of replicated time series analysis where we provide a detailed description of these data. **Figure 6** shows both the auto- and cross-spectral elements of the noisy periodograms and of the denoising wavelet threshold estimator, averaged over each of 3 different subsets (start, middle, and end trials) composed of 10 trials each (as a subset of the original much longer set of trial replicates). Note again that both periodograms and wavelet estimators are constructed on the Riemannian manifold of Hermitian PD matrices.

A more comprehensive study developed by Chau & von Sachs (2019) shows, in addition, how building on the Riemann distance allows the construction of estimators that are stable (i.e., invariant) to permutations and change of coordinate system of the underlying time series.

An additional advantage of having implemented the general ideas of Chau and von Sachs on positive-definiteness-preserving estimation via nonlinear wavelet methodology is the proposal of a global smoothing parameter—a kind of universal threshold based on the information in the trace of the spectral (and periodogram) matrix. To achieve this, the authors developed an additive signal-plus-noise model akin to traditional scalar curve estimation (and Equation 12) that respects the theoretical justification for successful denoising. However, more research on this recent approach is necessary to allow for a more flexible threshold choice, in the direction of potentially smoothing each element of the periodogram matrix by its own appropriate threshold. Interestingly, the off-diagonal cross-spectral structure often does not show less smoothness over frequency than the autospectra (see **Figure 1**), which saves us from needing to be too ambitious here. Finally, Chau & von Sachs (2018) provide a PD time-dependent spectral analysis built on this methodology.

4. REPLICATED TIME SERIES

Spectral analysis in the presence of time series replicates has recently become an important topic with applications in medicine, neuroscience, and beyond. Consider medical studies where subjects

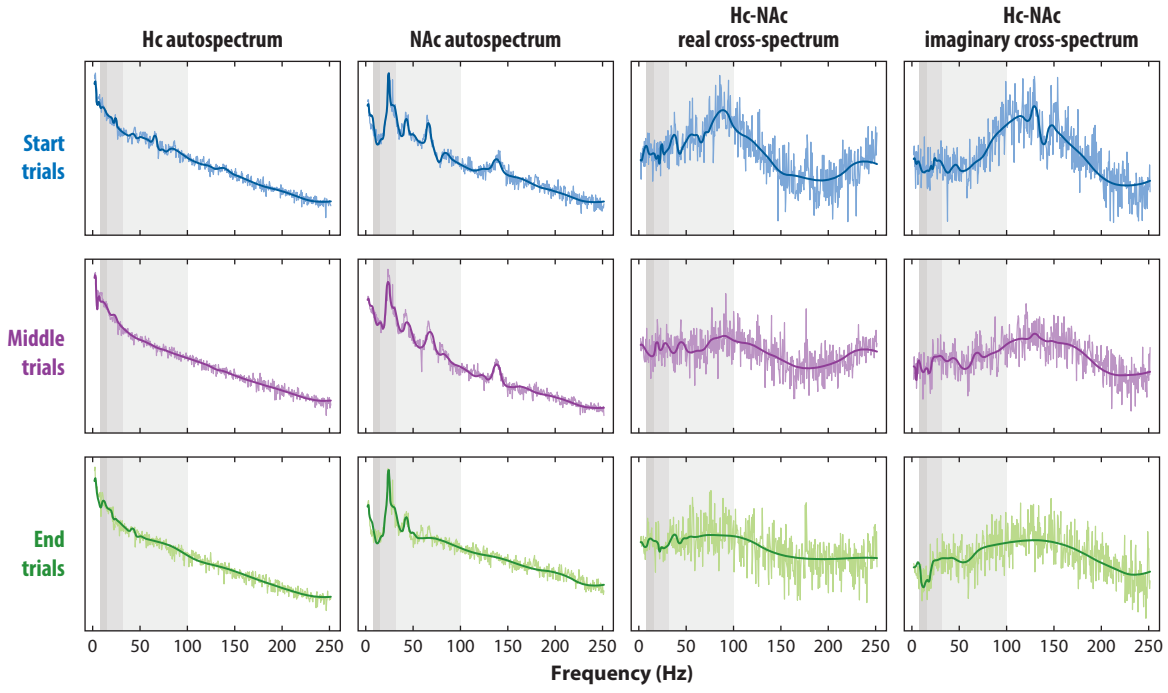

Figure 6

Auto- and cross-spectral periodograms and positive definite threshold wavelet estimators of three different trials of the bivariate Hc/NAc LFP series shown in **Figure 7**. Abbreviations: Hc, hippocampus; LFP, local field potential; NAc, nucleus accumbens.

(animals or human patients) are exposed to supervised experiments or trials, the outcome of which is monitored as a function of the conditions. Due to the availability of more than one time series string, in principal, it becomes possible to estimate the characteristic quantities of more complicated time series models—something virtually impossible if one deals with only one time series set. However, assuming independent replicates is not always very realistic in practice. Hence, Chau & von Sachs (2016) tried, in a functional mixed-effects model in the frequency domain, to take possible correlations between the replicates into account. In this section, we discuss a few approaches to replicated time series analysis in the frequency domain that reflect the recent evolution of these kinds of models.

The classical instance of replicated (multivariate) time series analysis starts from the idea that one disposes of (approximately) independent copies of the same underlying data-generating process such that one can construct ensemble averages to better estimate the spectral structure via (smoothed) periodograms and associated coherence and phase estimators. This allows, in particular, the application of less stringent smoothing over frequency (e.g., just enough to make the periodogram matrix full rank) and hence a gain in frequency resolution, and also enables the treatment of data with a time-varying spectral content without needing to smooth much over time. In statistical terms, this situation would typically be described by a fixed effects model. However, assuming variability within the subject population (still independent) by adding random effects is more realistic and will lead to a mixed effects model. This is widely used in the time domain, but there have been far fewer attempts to do this directly in the frequency domain via a functional mixed effects model (Krafty et al. 2011): Typically, for any of the S replicated time series, the sth subject-specific log-periodogram $Y^s(\omega) := \log I_T(\omega)$, at a given frequency ω, is modeled to be a

superposition of a population log-spectrum $h(\omega)$ (the mean or fixed effect), plus a subject-specific random effect $U^s(\omega)$ at this frequency, plus a nonsystematic (i.i.d.) error $e^s(\omega)$:

$$Y^s(\omega) \;=\; h(\omega) + U^s(\omega) + e^s(\omega), \quad s = 1, \ldots, S.$$

Here, the errors $e^s(\omega)$ are independent between different replicates s and independent of the functional random effects $U^s(\omega)$. For each $s = 1, \ldots, S$ they follow the same approximative $\log(\chi^2_2/2)$ distribution as the error in the Wahba approximation of Equation 12.

In the context of wavelet smoothing of the observed (log-)periodograms, Freyermuth et al. (2010) and Chau & von Sachs (2016), allowing for a correlation structure within the subject population $\{U^s(\omega)\}$, made some progress on transferring the advantages of nonlinear wavelet thresholding (Section 3.1) to this more refined setup. They developed some proper theory of including the mixed-effects situation; however, rendering this promising approach fully multivariate is still an open problem, as one would need to give up working with the particularly attractive log-periodogram.

Some recent, more practical work on replicated time series analysis in the frequency domain can be found, for example, in Gorrostieta et al. (2012) and Fiecas & Ombao (2016). As an alternative to treating time-varying spectral analysis as exposed in Section 2.3, in Fiecas & Ombao (2016), the replicates are provided by cutting up the timeline of an associative learning experiment of an animal in a brain study into a series of (time-ordered) trials. Then a model of these trial-replicated spectral analyses is developed. The authors can clearly trace the evolution of the animal acquiring learned patterns over the time-ordered trials, which shows up in the evolution of estimated spectral structure (changing linear dependencies) of the local field potential data of the animal's brain. (A different, though related, analysis is provided by the estimators in **Figure 6**.) To give an illustration of the nature of these data, in **Figure 7** we show, on a bivariate subset, three different trials of 2 s duration each, recorded in the hippocampus and the nucleus accumbens regions of the animal's brain.

For the same experiment, using a more classical approach of a sequence of piecewise stationary segments of these same brain data, Gorrostieta et al. (2012) previously had studied a more complex frequency model, within a harmonizable process: It turned out that standard coherence does not adequately model many biological signals with complex dependence structures—such as cross-oscillatory interactions between a low-frequency component in one signal and a

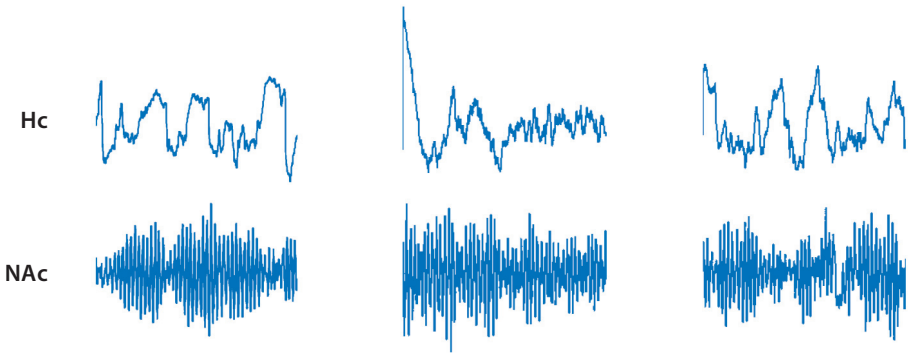

Figure 7

A subset of three different trials of a recorded LFP bivariate brain data set. Abbreviations: Hc, hippocampus; LFP, local field potential; NAc, nucleus accumbens.

high-frequency component in another—which is possible using harmonizable processes (revisited in Section 6.1).

5. HIGH-DIMENSIONAL TIME SERIES

As soon as the dimensionality P gets into the order of the sample size T or even larger, the well-known problem arises that the spectral density matrix, and in particular its estimators, become nonregular (and hence noninvertible). To face this problem, different regularization approaches have been proposed, mostly in the context of the related problem of regularization of high-dimensional covariance matrices. Whereas one popular method is based on banding, tapering, or thresholding (discussed, among many other methods, in the excellent overview by Pourahmadi 2011), we focus on the complementary approach based on shrinkage. This is supposed to work without any assumption of sparsity—which is typically rarely fulfilled for spectral density matrices, as opposed to covariance matrices, for which the first type of methods seemingly work well. Shrinkage-based methods, which essentially amount to adding a ridge (e.g., a multiple of the identity matrix) to the covariance matrix, were first developed in the seminal work of Ledoit & Wolf (2004). In this review, we concentrate on briefly discussing the linear shrinkage approach, which, being particularly simple, has entered into spectral analysis in a series of papers. Its nonlinear version, developed in Ledoit & Wolf (2012), is an interesting but rather impractical generalization related to random matrix theory (a practical improvement was recently proposed, however, in Ledoit & Wolf 2017). This research has triggered a variety of related alternative regularization approaches, but discussing them would be outside the scope of this review.

Linear shrinkage of covariance or spectral density matrices amounts to finding the (asymptotic) MSE optimal linear combination between the given empirical covariance or periodogram matrix and a prespecified target that regularizes the former. The solution enjoys good statistical and numerical properties as it considerably reduces the spread of the eigenvalues. It thus improves the condition number (the ratio of the largest to the smallest eigenvalue), in particular lifting the smallest eigenvalues away from zero. Böhm & von Sachs (2009) implemented this approach by taking on the original idea of adding a multiple of the identity matrix to the (smoothed) periodogram. This strategy is advised in absence of any a priori information. In structural shrinkage, as in Böhm & von Sachs (2008), the point of view is slightly changed by assuming that the spectral structure of the data is induced by underlying factors. However, in contrast to actual factor modeling suffering from the need to choose the number of factors, in this model-free approach, the final estimator is the asymptotically MSE-optimal linear combination of the smoothed periodogram and the parametric estimator based on an underfitting (and hence deliberately misspecified) factor model. Finally, Böhm et al. (2010) combined the aforementioned SLEX estimation method for fitting piecewise-stationary models (as discussed in Section 2.3.2) with a shrinkage estimator for the spectral density matrix used in the SLEX algorithm.

Further applications of spectral shrinkage arise, for example, in several papers by Fiecas and colleagues. On the one hand, Fiecas & von Sachs (2014) developed a time-frequency-toggle bootstrap method based on Kirch & Politis (2011) to choose the optimal shrinkage parameter in the mixture of a kernel-smoothed high-dimensional periodogram with any suitable regularization target. Following the idea of Böhm & von Sachs (2008), on the other hand, Fiecas & Ombao (2011) developed a generalized shrinkage estimator for the analysis of functional connectivity of brain signals. Here, the proposed spectral estimator is a weighted average of a parametric estimator and a nonparametric estimator to balance good frequency resolution (of the former) and robustness against model misspecification (of the latter). To complete the instances of shrinkage in high-dimensional

time series, in Fiecas et al. (2017), a linearly shrunken covariance matrix estimator is used within a multivariate hidden Markov chain for regime switching in a simple multivariate volatility model.

6. BEYOND CLASSICAL FOURIER-BASED SPECTRAL ANALYSIS

In this section, we briefly discuss a nonexhaustive series of approaches to spectral analysis that are not based on classical Fourier analysis.

6.1. Dual-Frequency Spectral Analysis for Harmonizable Processes

Motivated by a refined analysis of EEG data recorded in a motor intention experiment, Gorrostieta et al. (2012)—and Gorrostieta et al. (2019), for the more realistic case of time-varying spectral structure—allow for modeling cross-oscillatory interactions between, e.g., a low-frequency component in one signal and a high-frequency component in another. This notion of cross-dependence between regions (single frequencies or bands) in the frequency domain has been made possible by replacing the classical spectral representation (Equation 13) of stationary processes by the one of harmonizable processes. For these, the increment processes are not orthogonal over frequency; rather, they fulfill

$$\mathrm{Cov}(\mathrm{d}Z(\omega_1), \mathrm{d}Z(\omega_2)) = f(\omega_1, \omega_2)\, \mathrm{d}\omega_1 \mathrm{d}\omega_2, \qquad 16.$$

where $f(\omega_1, \omega_2)$ is the so-called Loève spectrum (Lii & Rosenblatt 2002). In Gorrostieta et al. (2019), this dual-frequency spectrum is rendered time varying, and an evolutionary dual-frequency coherence, as well as time-localized estimators based on dual-frequency local periodograms of replicated time series courses, is developed. In the applications, the proposed method uncovers new and interesting cross-oscillatory interactions that have been overlooked by the standard approaches. We also refer readers to the discussion of the aspect of replicated time series analysis in Section 4.

6.2. Wavelet Spectra

It seems to be an obvious idea to try to replace Fourier analysis of time series by something that is similar in spirit but more localized in time. Having learned about the success story of wavelets in statistics (primarily for curve estimation; see also Section 3.1), several researchers at the end of the past century explored their use as an alternative to a (local) Fourier analysis. Nason et al. (2000) came up with a first rigorous model of a stochastic representation of a time series as an analog to the well-known Cramér spectral representation (Equation 13). In their approach, the Fourier basis is replaced by a (particular) wavelet basis $\{\psi_{jk}(t)\}$, localized at (dyadic) locations $k = 0, \ldots, 2^j - 1$, on successively finer scales 2^{-j}, $j \geq 0$,

$$X(t) = \sum_j \sum_k w_{jk}\, \psi_{jk}(t)\, \xi_{jk}. \qquad 17.$$

Here, the amplitudes w_{jk} take the role of the localized amplitude functions $A_t(\omega)$ discussed in Section 2.3.1, whereas the sequence ξ_{jk} (uncorrelated over j and k) can be seen as a discrete version of the increment process $\mathrm{d}Z(\omega)$ of a time-dependent Cramér spectral representation. In this sense, a scale j replaces (the reciprocal of a discretized) frequency ω, whereas local time t can be seen to correspond to a localization k on scale j in the typical dyadic scale-location plane of wavelets. The interpretation of this model is, roughly speaking, that the amplitude w_{jk} is large if at time

$t = k$ there is high correlation of X_k with $X_{k-\tau}$ or $X_{k+\tau}$ for some τ that matches the wavelength of $\psi_{jk}(t)$, which itself is proportional to 2^{-j}.

To be able to properly include stationary models into their representation (Equation 17) Nason and von Sachs had to use, as building blocks, nondecimated wavelets (Nason & Silverman 1995). These are variants of classical orthogonal wavelets (such as the Daubechies family of compact support) made shift-invariant (shift invariance of the Fourier basis being the key to representing stationarity). The authors found a way to control for the resulting redundancy in this new spectral representation with respect to overlapping and hence no longer orthogonal basis functions. They introduced what they called the evolutionary wavelet spectrum, in the class of locally stationary wavelet processes. Basically, this wavelet spectrum $S_j(u)$ at local or rescaled time $u \in [0, 1]$ (see Section 2.3.1) amounts to w_{jk}^2 at localization $u = k/T$. However, similarly to Dahlhaus's theory of evolutionary spectra (Dahlhaus 2012), they work with some doubly indexed array of processes arising from Equation 17 to allow for uniquely defined limiting spectra, accompanied by a theory of their consistent estimation. The latter is based on the wavelet periodogram I_{jk}, basically the squared wavelet coefficients at scale j and location k of the time series $X(t)$. As in Fourier spectral analysis, one needs to work with a smoothed version of the (wavelet) periodogram, where smoothing is over neighboring time locations (after having applied a bias correction to reduce the redundancy over wavelet scales). The evolutionary wavelet spectrum $S_j(u)$, quite analogously to the Fourier spectrum, also arises in a spectral representation of a (local) autocovariance function of $X(t)$ with respect to what are called autocorrelation wavelets, $\Psi_j(\tau) := \sum_k \psi_{jk}(0)\psi_{jk}(\tau)$.

This seminal research by Nason and von Sachs, though somewhat theoretical in flavor, triggered a long series of subsequent work on both modeling and practical application of the idea of a wavelet spectral representation that could be estimated from the time series data. For this review, we content ourselves to cite the first instances of the bi- and multivariate generalizations of Equation 17 delivered by Sanderson et al. (2010) and Park et al. (2014): Both approaches attempt to build up a cross-spectral analysis, including what could be called wavelet coherence. Although this concept is interesting in itself, it remains questionable whether the benefit of working with a truly local (compared with the nonlocal Fourier) basis is not somewhat overruled by the fact that phase interpretation is much more difficult in the wavelet domain (in comparison with the elegant property of Fourier functions, with linear phase, to give direct information about lead and lag of two components of a multivariate time series).

An interesting field of application for this kind of (multivariate) wavelet spectral analysis, however, consists in (multiple) change-point detection, in particular for high-dimensional time series. In Cho & Fryzlewicz (2015), and in Barigozzi et al. (2018) in the context of factor analysis, a well-established method for detection of breaks in the mean structure of data, the cumulative sum (CUSUM) method, is successfully applied, via combination with (sparsified) binary segmentation, to the series of wavelet periodograms and cross-periodograms, in order to detect changes in the second-order structure of a possibly high-dimensional serially correlated multivariate series.

6.3. Spectral Analysis for Locally Stationary Hawkes Processes

Locally stationary (multivariate) Hawkes processes are a class of time-inhomogeneous (self-exciting) point processes, spectral analysis for which has been developed in Roueff et al. (2016) and Roueff & von Sachs (2019). Their first work addresses the generalization of stationary Hawkes processes in order to allow for a time-evolving second-order analysis. We recall that stationary linear Hawkes processes with fertility function p can be described, on the one hand, by

their conditional intensity function $\lambda(t)$, which is given by

$$\lambda(t) = \lambda_c + \int_{-\infty}^{t^-} p(t-s)\, N(\mathrm{d}s) \;=\; \lambda_c + \sum_{t_i<t} p(t-t_i),$$

where the integral of p with respect to the counting process N is to be understood as a sum of Dirac masses at (random) points $\{t_i\}_i$. On the other hand, linear self-exciting processes can also be viewed as clusters of point processes, which are created by the dynamics of the interplay between immigrant arrivals and their offspring production. This self-exciting mechanism is used for modeling phenomena in, e.g., seismological, genome, and brain data analysis but also for high-frequency financial data. In order to develop spectral analysis for the dynamics of Hawkes processes, we also briefly recall the analog of Equation 13, i.e., the spectral decomposition of the variance of a point process N, allowing us to subsequently define its Bartlett spectrum $f(\omega)$ via

$$\mathrm{Var}(N(g)) \;=\; \int \left|\widehat{g}(\omega)\right|^2 f(\omega)\,\mathrm{d}\omega,$$

where $\hat{g}$ denotes the Fourier transform of a test function g (e.g., a smooth indicator or window function over a given time interval), i.e., $\widehat{g}(\omega) = \int g(t)\exp(-it\omega)\mathrm{d}t$. For stationary Hawkes processes with immigrant intensity λ_c and fertility function p, the Bartlett spectral density is then given by

$$f(\omega) = \frac{\lambda_c}{2\pi(1-\int p)}\,\left|1-\widehat{p}(\omega)\right|^{-2}.$$

Motivated by the concept of locally stationary autoregressive processes as developed by Dahlhaus (2012), inherently different techniques are applied in the work of Roueff and von Sachs to describe the time-varying dynamics of self-exciting point processes. For this, the time-dependent Bartlett spectrum $f(t,\omega)$ is modeled via now time-dependent intensity $\lambda_c(t)$ and time-varying fertility $p(t;t-s)$. In particular, a stationary approximation of the Laplace functional of a locally stationary Hawkes process is derived. With this, it has been possible to rigorously define a local mean density function and a local Bartlett spectrum. A rather complete theory and some working algorithms for estimating these objects (including rates of convergence) were delivered in the second work of Roueff & von Sachs (2019), which also presented an application to order book tick-by-tick data modeled previously by homogeneous point processes.

6.4. Quantile Spectral Analysis

Obviously, classical spectral analysis, based on the Fourier analysis of the auto- and cross-covariance function, only takes the structure of linear dependencies in a given time series into account. A variety of approaches attempt to remedy this drawback, to allow for more general rather than just second-order stationary processes, and to model and estimate dependencies in the tails of the distribution of given time series: These so-called extremes are important for highly non-Gaussian time series as they arise in climatology, and also more generally in spatial-temporal data. Quantile spectral analysis was first described in Parzen (1985). However, only in a more recent series of articles, starting from Dette et al. (2015), have Kley, Volgushev, Birr, and coworkers developed rigorous modern statistical concepts in this field. They embed them into locally stationary process theory (Birr et al. 2017) and cover cross-spectral analysis and copula spectral densities as well. In a nutshell, this type of spectral analysis is based on Fourier transforms of the following two types of generalized covariance. We give them only for the case of autocovariance

of a univariate strictly stationary process $\{X_t\}_t$, but obvious modifications for the multivariate case, and also time-varying versions of it, do exist:

$$\gamma_k(x_1, x_2) := \mathrm{Cov}(I\{X_t \le x_1\}, I\{X_{t-k} \le x_2\}), \quad x_i \in \mathbf{R},$$

called the Laplace covariance-kernel, and the copula covariance-kernel,

$$\gamma_k^U(\tau_1, \tau_2) := \mathrm{Cov}(I\{U_t \le \tau_1\}, I\{U_{t-k} \le \tau_2\}), \quad \tau_i \in (0, 1),$$

where $U_t := F(X_t)$, with F being the marginal distribution of $\{X_t\}_t$, and where $I\{B\}$ stands for the indicator function of the set B.

A more specific frequency domain approach for heavy-tailed time series has been developed in parallel by Davis, Mikosch, and various coauthors. They invented the concept of extremograms (of uni- and multivariate time series); we refer readers to the overview in Davis et al. (2013) of recent measures of serial extremal dependence in a strictly stationary time series as well as their estimation. For completeness, here we recall the definition of the extremogram in the univariate case, essentially akin an (asymptotic) correlogram for extreme events; its generalization to the cross-extremogram can be found in the cited literature. In fact, the extremogram is defined as a limiting sequence given by

$$\gamma_{AB}(h) = \lim_{T\to\infty} T\ \mathrm{Cov}(I\{a_T^{-1}X_0 \in A\}, I\{a_T^{-1}X_h \in B\}), \quad h \ge 0.$$

Here, (a_T) is a suitably chosen normalization sequence and A, B are two fixed sets bounded away from zero. The events $\{X_0 \in a_T A\}$ and $\{X_h \in a_T B\}$ are considered as extreme ones, and $\gamma_{AB}(h)$ measures the influence of the time zero extremal event $\{X_0 \in a_T A\}$ on the extremal event $\{X_h \in a_T B\}$, h lags apart. The choice of (a_T) depends on the situation at hand, e.g., via $T\ \mathrm{P}(|X| > a_T) \sim 1$, which leads to $\gamma_{AB}(h) = \lim_{T\to\infty} T\ \mathrm{P}(X_0 \in a_T A, X_h \in a_T B)$. Motivating examples of extremograms are the limiting conditional probabilities $\lim_{T\to\infty} T\ \mathrm{P}(X_h \in a_T B | X_0 \in a_T A)$. With creative choices of A and B, one can investigate interesting sources of extremal dependence that may arise not only in the upper and lower tails but also in other extreme regions of the sample space.

The practical assessment of this limiting approach is less direct, and for all of the cited methods, a practitioner might run into the problem of not having at his or her disposal sufficiently many time series observations while trying to estimate the dependence structure in the very high (or very low) quantiles of the time series distribution.

7. REMAINING ASPECTS AND CONCLUSION

With this nonexhaustive review on (multivariate) spectral analysis, we have tried to alert the reader to the existing challenges in this area and how to (partially) address them. We have focused on the author's experience with various approaches to smoothing the (uni- or multivariate) periodogram statistics, over frequency, ideally preserving the important property of a PD spectral estimator. On a second note, we introduced the reader to some aspects of time-varying spectral estimation, insisting on the limitations of classical spectral analysis under the restrictive assumption of stationary time observations. Finally, we mentioned a few approaches that go beyond Fourier analysis of classical time series (auto- and cross-)covariances, fields that certainly need more development to enter into the world of practitioners. Before concluding, we briefly mention a number of important issues not addressed here, for which we apologize: parametric spectral analysis based on

VARMA modeling (still very much used in fields such as econometrics) and semiparametric approaches based on, e.g., the Whittle likelihood [see, again, Dahlhaus (2012), and also Dahlhaus & Neumann (2001) on fitting semiparametric models to nonstationary time series and Dahlhaus & Polonik (2006) on nonparametric quasi maximum likelihood estimation for locally stationary time series]. Another developing field that has become important for high-dimensional data modeling is that of graphical models for time series (Dahlhaus 2000b, Eichler 2012). We also have not discussed the very active field of functional time series, for which most of the ideas of spectral analysis can be transferred to analyzing curve-valued time series; however, the necessary developments for this are highly nontrivial (Panaretos & Tavakoli 2013, Hörmann et al. 2015, van Delft & Eichler 2018). We have also omitted aspects of outlier-robust spectral analysis (von Sachs 1994) and what to do with missing values—both of which are extremely important for practical time series analysts. Another important application of spectral analysis is discrimination of time series in the frequency domain (Kakizawa et al. 1998, Shumway 2003, Huang et al. 2004, Sakiyama & Taniguchi 2004).

And finally, again from a more econometric point of view, we mention the use of spectral estimation for dynamic factor analysis, as in the series of papers by Forni et al. (starting with Forni et al. 2000), essentially developed for dimension reduction. In this approach, a time series panel is explained by a superposition of a few (uncorrelated) driving factors (with potentially time-varying loadings) and an idiosyncratic structure—possibly as general as a locally stationary process (Eichler et al. 2011). Bayesian spectral analysis (see Rosen et al. 2009, and Rosen et al. 2012, which also provides both Matlab and R code) is an important field that we have not discussed explicitly, as is spectral analysis of qualitative or categorical time series (Krafty et al. 2012, Stoffer 2015).

Clearly, this review emphasizes existing model approaches rather than statistical estimation (apart from the specific aspect of periodogram smoothing) and inference. Of course, the latter is equally important, and a more comprehensive overview about different inference methods would be a rewarding project for a different occasion. This should include a variety of recent mathematical-statistical work on testing problems related to spectral analysis (see, e.g., Dette & Paparoditis 2009). Finally, researchers who are experts in resampling methods such as the bootstrap have contributed important methodology to spectral estimation (see Franke & Härdle 1992, Jentsch & Kreiss 2010, Kirch & Politis 2011, and the excellent overview in Kreiss & Paparoditis 2020). In this field, due to the challenging variance-covariance structure of spectral estimators (compare Equation 11 on the variance of the kernel-smoothed periodogram), resampling methods turn out to be a very promising alternative to, e.g., suboptimal plug-in methods.

Concluding this certainly imperfect review, let us ask the question, "Where do we go from here?" With this rather impressive series of methodological contributions to the field of spectral analysis (and related questions), it remains very important that the proposed methodologies are converted into practical, working algorithms, with precise guidelines for how and in which situation they can safely be applied. In this review, we tried to give some examples of common but nonstandard situations that present challenges that require practical solutions. We discussed examples including replicated and nonstationary multivariate time series data and/or high-dimensional serially correlated data, possibly with outliers and/or missing values. There is still a need to better understand the developed algorithms and their properties (such as preserving positive-definiteness, or their regularization capacity for high-dimensional data), and for increased interaction with practitioners. Whereas time series analysis in general found its way some time ago into the world of application, spectral analysis still tends to appear somewhat too complicated to be more widely used. A prominent exception is, among others, brain data modeling and analysis in neuroscience. Here, however, a new challenge arose when researchers combined more classical spectral analysis of EEG data, known to give high frequency resolution, with fMRI (functional magnetic resonance

imaging) data, which are complementary as they enjoy better spatial-temporal resolution (Wang et al. 2017). It remains a challenge to further explore this combination from the statistical side.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The author thanks J. Chau and M. Fiecas for help preparing this review. More generally, he expresses his gratitude to his colleagues—and in particular his numerous collaborators—in this field for having made his scientific life so enjoyable over the past couple of decades (full of mutual scientific visits and occasions to discuss on behalf of stimulating meetings). At the same time, he apologizes for any missing reference he should have cherished.

LITERATURE CITED

Adak S. 1998. Time-dependent spectral analysis of non-stationary time series. *J. Am. Stat. Assoc.* 93:1488–501

Barigozzi M, Cho H, Fryzlewicz P. 2018. Simultaneous multiple change-point and factor analysis for high-dimensional time series. *J. Econometr.* 206:187–225

Birr S, Volgushev S, Hallin M, Kley T, Dette H. 2017. Quantile spectral analysis for locally stationary time series. *J. R. Stat. Soc. B* 79:1619–43

Böhm H, Ombao H, von Sachs R, Sanes J. 2010. Classification of multivariate non-stationary signals: the SLEX-shrinkage approach. *J. Stat. Plan. Inference* 140:3754–63

Böhm H, von Sachs R. 2008. Structural shrinkage of nonparametric spectral estimators for multivariate time series. *Electron. J. Stat.* 2:696–721

Böhm H, von Sachs R. 2009. Shrinkage estimation in the frequency domain of multivariate time series. *J. Multivar. Anal.* 100:913–35

Brillinger D. 1981. *Time Series: Data Analysis and Theory*. San Francisco: Holden-Day

Brockwell P, Davis R. 1991. *Time Series: Theory and Methods*. New York: Springer

Bühlmann P. 1996. Locally adaptive lag-window spectral estimation. *J. Time Ser. Anal.* 17:247–70

Chau J. 2018. `pdSpecEst`: Positive-definite wavelet-based multivariate spectral analysis. *R package*, version 1.2.3. **https://CRAN.R-project.org/package=pdSpecEst**

Chau J, Ombao H, von Sachs R. 2019. Intrinsic data depth for Hermitian positive definite matrices. *J. Comput. Graph. Stat.* 28:427–39

Chau J, von Sachs R. 2016. Functional mixed effects wavelet estimation for spectra of replicated time series. *Electron. J. Stat.* 10:2461–510

Chau J, von Sachs R. 2019. Intrinsic wavelet regression for curves of Hermitian positive definite matrices. *J. Am. Stat. Assoc.* **https://doi.org/10.1080/01621459.2019.1700129**

Chau J, von Sachs R. 2018. Intrinsic wavelet regression for surfaces of Hermitian positive definite matrices. arXiv:1808.08764 [stat.ME]

Cho H, Fryzlewicz P. 2015. Multiple change-point detection for high-dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. B* 77:475–507

Cooley J, Tukey J. 1965. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* 90:297–301

Dahlhaus R. 1997. Fitting time series models to nonstationary processes. *Ann. Stat.* 25:1–37

Dahlhaus R. 2000a. A likelihood approximation for locally stationary processes. *Ann. Stat.* 28:1762–94

Dahlhaus R. 2000b. Graphical interaction models for multivariate time series. *Metrika* 51:157–72

Dahlhaus R. 2012. Locally stationary processes. In *Time Series Analysis: Methods and Applications*, ed. T Subba Rao, S Subba Rao, CR Rao, pp. 351–413. Amsterdam: Elsevier

Dahlhaus R, Neumann M. 2001. Locally adaptive fitting of semiparametric models to nonstationary time series. *Stoch. Process. Appl.* 91:277–308

Dahlhaus R, Polonik W. 2006. Nonparametric quasi maximum likelihood estimation for Gaussian locally stationary processes. *Ann. Stat.* 34:2790–824

Dahlhaus R, Richter S, Wu WB. 2019. Towards a general theory for nonlinear locally stationary processes. *Bernoulli* 25:1013–44

Dahlhaus R, Subba Rao S. 2006. Statistical inference for time-varying ARCH processes. *Ann. Stat.* 34:1075–114

Dai M, Guo W. 2004. Multivariate spectral analysis using Cholesky decomposition. *Biometrika* 91:629–43

Davis R, Lee TCM, Rodriguez-Yam G. 2006. Structural break estimation for non-stationary time series models. *J. Am. Stat. Assoc.* 101:223–39

Davis R, Mikosch T, Zhao Y. 2013. Measures of serial extremal dependence and their estimation. *Stoch. Process. Appl.* 123:2575–602

Dette H, Hallin M, Kley T, Volgushev S. 2015. Of copulas, quantiles, ranks and spectra: An L_1−approach to spectral analysis. *Bernoulli* 21:781–831

Dette H, Paparoditis E. 2009. Bootstrapping frequency domain tests in multivariate time series with an application to comparing spectral densities. *J. R. Stat. Soc. B* 71:831–57

Eichler M. 2012. Graphical modelling of multivariate time series. *Probab. Theory Relat. Fields* 153:233–68

Eichler M, Motta G, von Sachs R. 2011. Fitting dynamic factor models to non-stationary time series. *J. Econom.* 163:51–70

Fan J, Gijbels I. 1992. *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall

Fiecas M, Franke J, von Sachs R, Tadjuidje J. 2017. Shrinkage estimation for multivariate hidden Markov mixture models. *J. Am. Stat. Assoc.* 112:424–35

Fiecas M, Ombao H. 2011. The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *Ann. Appl. Stat.* 5:1102–25

Fiecas M, Ombao H. 2016. Modeling the evolution of dynamic brain processes during an associative learning experiment. *J. Am. Stat. Assoc.* 111:1440–53

Fiecas M, von Sachs R. 2014. Data-driven shrinkage of the spectral density matrix of a high-dimensional time series. *Electron. J. Stat.* 8:2975–3003

Forni M, Hallin M, Lippi M, Reichlin L. 2000. Fitting dynamic factor models to non-stationary time series. *Rev. Econ. Stat.* 82:540–54

Franke J, Härdle W. 1992. On bootstrapping kernel spectral estimates. *Ann. Stat.* 20:121–45

Freyermuth JM, Ombao H, von Sachs R. 2010. Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series. *J. Am. Stat. Assoc.* 105:634–46

Fryzlewicz P, Nason G, von Sachs R. 2008. A wavelet-Fisz approach to spectrum estimation. *J. Time Ser. Anal.* 29:868–80

Gao H. 1997. Choice of thresholds for wavelet shrinkage estimate of the spectrum. *J. Time Ser. Anal.* 18:231–51

Gorrostieta C, Ombao H, Prado R, Patel S, Eskandar E. 2012. Exploring dependence between brain signals in a monkey during learning. *J. Time Ser. Anal.* 33:771–78

Gorrostieta C, Ombao H, von Sachs R. 2019. Time-dependent dual-frequency coherence in multivariate nonstationary time series. *J. Time Ser. Anal.* 40:3–22

Guo W, Dai M. 2006. Multivariate time-dependent spectral analysis using Cholesky decomposition. *Stat. Sin.* 16:825–45

Hörmann S, Kidzin'ski L, Hallin M. 2015. Dynamic functional principal components. *J. R. Stat. Soc. B* 77:319–48

Huang HY, Ombao H, Stoffer D. 2004. Discrimination and classification of nonstationary time series using the SLEX model. *J. Am. Stat. Assoc.* 99:763–74

Jansen M, Oonincx P. 2005. *Second Generation Wavelets and Applications*. London: Springer

Jentsch C, Kreiss JP. 2010. The multiple hybrid bootstrap resampling multivariate linear processes. *J. Multivar. Anal.* 101:2320–45

Kakizawa Y, Shumway R, Taniguchi M. 1998. Discrimination and clustering for multivariate time series. *J. Am. Stat. Assoc.* 441:328–40

Kirch C, Politis DN. 2011. TFT-bootstrap: resampling time series in the frequency domain to obtain replicates in the time domain. *Ann. Stat.* 39:1427–70

Koopmans J. 1974. *The Spectral Analysis of Time Series*. Amsterdam: Elsevier

Krafty R, Collinge W. 2013. Penalized multivariate Whittle likelihood for power spectrum estimation. *Biometrika* 100:447–58

Krafty R, Hall M, Guo W. 2011. Functional mixed effects spectral analysis. *Biometrika* 98:583–98

Krafty R, Xiong S, Stoffer D, Buysse D, Hall M. 2012. Enveloping spectral surfaces: covariate dependent spectral analysis of categorical time series. *J. Time Ser. Anal.* 33:797–806

Kreiss JP, Paparoditis E. 2020. *Bootstrap for Time Series: Theory and Applications*. New York: Springer. In press

Ledoit O, Wolf M. 2004. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* 88:365–411

Ledoit O, Wolf M. 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Stat.* 40:1024–60

Ledoit O, Wolf M. 2017. *Analytical nonlinear shrinkage of large-dimensional covariance matrices*. Work. Pap. No. 264, Dep. Econ., Univ. Zurich

Lii K, Rosenblatt M. 2002. Spectral analysis for harmonizable processes. *Ann. Stat.* 30:258–97

Martin W, Flandrin P. 1995. Wigner–Ville spectral analysis of nonstationary processes. *IEEE Trans. Signal Process.* 33:1461–70

Moulin P. 1994. Wavelet thresholding techniques for power spectrum estimation. *IEEE Trans. Signal Process.* 42:3126–36

Nason G. 2008. *Wavelet Methods in Statistics with R*. New York: Springer

Nason G, Silverman B. 1995. The stationary wavelet transform and some statistical applications. In *Wavelets and Statistics*, ed. A Antoniadis, G Oppenheim, pp. 281–300. New York: Springer

Nason G, von Sachs R, Kroisandt G. 2000. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc. B* 62:271–92

Neumann M. 1996. Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *J. Time Ser. Anal.* 17:601–33

Neumann M, von Sachs R. 1997. Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Stat.* 25:38–76

Ombao H, Raz J, von Sachs R, Guo W. 2002. The SLEX model of a non-stationary random process. *Ann. Inst. Stat. Math.* 54:171–200

Ombao H, Raz J, von Sachs R, Mallow B. 2001. Automatic analysis of bivariate non-stationary time series. *J. Am. Stat. Assoc.* 96:543–60

Ombao H, von Sachs R, Guo W. 2005. SLEX analysis of multivariate non-stationary time series. *J. Am. Stat. Assoc.* 100:519–31

Panaretos V, Tavakoli S. 2013. Fourier analysis of stationary time series in function space. *Ann. Stat.* 41:568–603

Park T, Eckley I, Ombao H. 2014. Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Trans. Signal Process.* 62:5240–50

Parzen E. 1985. Time series model identification and quantile spectral analysis. *IFAC Proc. Vol.* 18:731–36

Pourahmadi M. 2011. Covariance estimation: the GLM and regularization perspectives. *Stat. Sci.* 26:369–87

Priestley M. 1965. Evolutionary spectra and non-stationary processes. *J. R. Stat. Soc. B* 27:204–37

Rahman I, Drori I, Stodden V, Donoho D, Schröder P. 2005. Multiscale representations for manifold-valued data. *Multiscale Model. Simul.* 4:1201–32

Rosen O, Stoffer D. 2007. Automatic estimation of multivariate spectra via smoothing splines. *Biometrika* 94:335–45

Rosen O, Stoffer D, Wood S. 2009. Local spectral analysis via a Bayesian mixture of smoothing splines. *J. Am. Stat. Assoc.* 104:249–62

Rosen O, Stoffer D, Wood S. 2012. Adaptspec: adaptive spectral estimation for nonstationary time series. *J. Am. Stat. Assoc.* 107:1575–89

Roueff F, von Sachs R. 2019. Time-frequency analysis of locally stationary Hawkes processes. *Bernoulli* 25:1355–85

Roueff F, von Sachs R, Sansonnet L. 2016. Locally stationary Hawkes processes. *Stoch. Process. Appl.* 126:1710–43

Sakiyama K, Taniguchi M. 2004. Discriminant analysis for locally stationary processes. *J. Multivar. Anal.* 90:282–300

Sanderson J, Fryzlewicz P, Jones M. 2010. Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika* 97:435–46

Shumway R. 2003. Time-frequency clustering and discriminant analysis. *Stat. Probab. Lett.* 63:307–14

Shumway R, Stoffer D. 2006. *Time Series Analysis and Its Applications*. New York: Springer

Stoffer D. 2015. Spectral analysis of qualitative time series. In *Handbook of Discrete-Valued Time Series*, ed. RA Davis, SH Holan, R Lund, N Ravishanker, pp. 287–309. London: CRC Press

Subba Rao S. 2018. Orthogonal samples for estimators in time series. *J. Time Ser. Anal.* 39:313–37

Thomson D. 1982. Spectrum estimation and harmonic analysis. *Proc. IEEE* 70:1055–96

van Delft A, Eichler M. 2018. Locally stationary functional time series. *Electron. J. Stat.* 12:107–70

von Sachs R. 1994. Peak-insensitive nonparametric spectrum estimation. *J. Time Ser. Anal.* 15:429–52

Wahba G. 1980. Automatic smoothing of the log periodogram. *J. Am. Stat. Assoc.* 75:122–32

Walden AT. 2000. A unified view of multitaper multivariate spectral estimation. *Biometrika* 87:767–88

Wang Y, Ting CM, Ombao H. 2017. Estimating dynamic connectivity states in fMRI using regime-switching factor models. *IEEE Trans. Med. Imaging* 37:1011–23

Welch P. 1967. The use of Fast Fourier Transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* 15:70–73

Wu WB, Zaffaroni P. 2018. Asymptotic theory for spectral density estimates of general multivariate time series. *Econometr. Theory* 34:1–22

Wu WB, Zhou Z. 2011. Gaussian approximations for non-stationary multiple time series. *Stat. Sin.* 21:1397–413

Yuan Y, Zhu H, Lin W, Marron J. 2012. Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. B* 74:697–719