SPECTRAL ANALYSIS OF MULTIVARIATE TIME SERIES

R. von Sachs

DISCUSSION PAPER | 2019 / 08

UCLouvain

ISBA | LIDAM Louvain Institute of Data Analysis and Modeling in economics and statistics

Spectral Analysis of Multivariate Time Series

Rainer von Sachs

Institute of Statistics, Biostatistics and Actuarial Sciences,
LIDAM, Université catholique de Louvain
email: rvs@uclouvain.be

April 5, 2019

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 a lot in understanding the dynamics in many serially correlated data without necessarily needing to develop complex parametric models. In this work, we will give a non-exhaustive review on, mostly recent, nonparametric methods of spectral analysis of multivariate time series, with a certain emphasis on model-based approaches. We try to give insights into a variety of complimentary approaches, for standard and less standard situations (such a non-stationary, replicated or high-dimensional time series), discuss estimation aspects (such as smoothing over frequency) and include some examples stemming from life science applications (brain data).

Contents

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 and Tukey, 1965), the analysis of serial auto- and cross correlation in the frequency domain has helped a lot in understanding the dynamics in many serially correlated data without necessarily needing to develop complex parametric models. By studying the (linear) dependence structure of the components of a multivariate time series over time one can get already quite a clear, although not always comprehensive, idea about, on the one hand, aspects such as causality, i.e. the lead and lag between two components of the observed series, for e.g. for economical indeces. 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 those as for the connectivity in the brain measured via multi-channel EEG or for temperature (and even climate) observations showing pronounced temporal cycles of random nature.

Whereas 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, it turned out that they are not sufficient to realistically describe many real life phenomena which show transients, regime changes over time etc. Hence methodological research started about twenty years ago to generalize spectral analysis in a way that would still allow for the use of the FFT (at least) locally over time. While these ideas had been around on the empirical side in the engineering literature already for quite a while, statisticians started only later to develop models that would accompany proposed modern methodologies (such as local FFT, spectrogram, Wigner-Ville or wavelet transform) allowing to derive their theoretical properties.

In this overview we like to concentrate on discussing a few of these recent mostly model-based approaches in nonparametric spectral analysis of uni- and multivariate time series. We will 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 will discuss in some more detail also time-varying spectral analysis. More specifically, we intend to mostly address the following aspects of modern spectral analysis:

1. How to 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 and Stoffer (2007) on spline smoothing, and Yuan et al. (2012) as well as Chau and von Sachs (2017) 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 and von Sachs (1997), Nason et al. (2000))?

3. How to perform spectral analysis in the presence of replicated time series in more complex situations of, e.g., supervised brain signal experiments (Gorrostieta et al. (2019), Fiecas and Ombao (2016), Chau and von Sachs (2016))?

4. What are the challenges of spectral analysis of high-dimensional time series (Böhm and von Sachs (2009), Fiecas and Ombao (2011), Fiecas and von Sachs (2014))?

5. What approaches exist that go beyond the classical Fourier based spectral analyses (Dette et al. (2015), Davis et al. (2013)), how to do spectral analysis for point processes (Roueff and von Sachs, 2019)?

Our choice is certainly biased and cannot be exhaustive - we somewhat opted for discussing almost exclusively nonparametric approaches which borrow strength from other fields of research (curve estimation, quantile regression, regularization methods, random effects modelling, .) and which allow for some theoretical development of the, mostly asymptotic, properties of the proposed estimation schemes. We will 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, some others likely to be more suited to find its usefulness in concrete applications.

With this choice we need to apologize beforehand that we are not able to include in our discussion a lot of other interesting topics, such as semi-parametric spectral analysis, goodness of fit of spectral distributions, testing, graphical time series models for partial correlation/coherence analysis, and many more (we refer to our last Section 7 where we mention some references).

2 Some Background on Spectral Analysis

There exist many good books (Brillinger (1981), Brockwell and Davis (1991), Shumway and Stoffer (2006) or Koopmans (1974), to name but a few) which give a detailed introduction into spectral analysis of multivariate times series. Hence we restrict ourselves here to recal some background, both formal and by pointing to some illustrative examples - in as much as we need it to prepare the ground for the following review of modern nonparametric methods applicable to spectral analysis. We start with setting up notations 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)

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

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.

$$\mathbf{f}(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \mathbb{E}(\mathbf{X}(t)\mathbf{X}(t+h)^\top) \exp(-i\omega h), \tag{2.1}$$

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

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 sinus and cosine waves with random amplitudes (and phase)

$$X_i(t) \approx \sum_k \left[U_{ik}\cos(\omega_k t) + V_{ik}\sin(\omega_k t)\right] \ , \ \omega_k = \frac{2\pi k}{T} \ , \tag{2.2}$$

where U_{ik}, V_{ik} are i.i.d. $(0, \sigma^2_{ik})$ such that

$$\mathrm{Var}\left(X_i(t)\right) \approx \sum_k \sigma^2_{ik} \ , \tag{2.3}$$

and

$$c_{ii}(h) \approx \sum_k \sigma^2_{ik} \ \cos(\omega_k h) \ . \tag{2.4}$$

Rigorously written, and generalised 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

$$\mathbf{c}_X(h) = \int_{-\pi}^{\pi} \mathbf{f}(\omega)\exp(i\omega h) \ d\omega \ . \tag{2.5}$$

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}, normalised by the autocovariances c_{ii} and c_{jj}) has a pronounced peak (only) at $h = \tau$, and 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 VARMA (vector autoregressive-moving average) process of order (2,2) based on Gaussian white noise Z. 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 also 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:

$$\mathbf{d}_X(\omega) = \sum_{t=1}^{T} \mathbf{X}(t)\exp(-i\omega t). \tag{2.6}$$

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

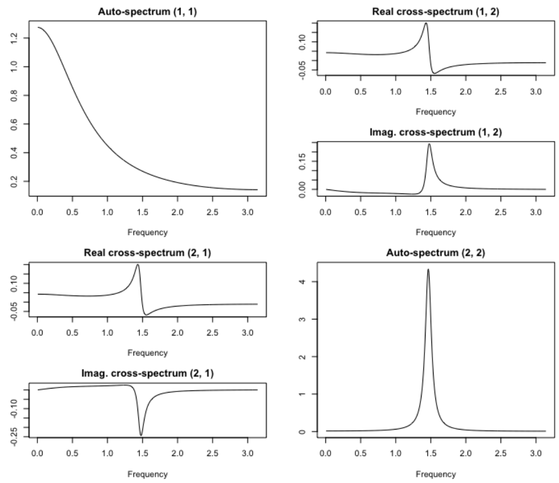

Figure 1: Population spectral matrix of a simulated VARMA(2,2)-process

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

$$\mathbf{I}_T(\omega) = (2\pi)^{-1} T^{-1} \mathbf{d}_X(\omega) \mathbf{d}_X(\omega)^*, \tag{2.7}$$

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 $T \to \infty$,

$$E\left(\mathbf{I}_T(\omega)\right) \ \to \ \mathbf{f}(\omega) \tag{2.8}$$

$$\operatorname{Var}\left(\mathbf{I}_T(\omega)\right) \ \to \ \mathbf{f}^2(\omega) \quad \omega \neq 0, \pm\pi \ . \tag{2.9}$$

In the next **Figure 2** we show a periodogram for a simulated stretch of length $T = 4096$ of the same VARMA-process as before.

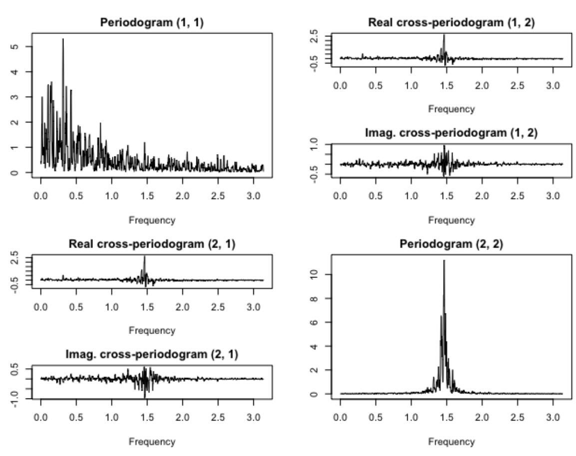

Figure 2: Periodogram matrix of the simulated VARMA(2,2)-process with spectrum shown in Figure 1

As can be seen from the wildly erratic behaviour 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 $M_T^{(jk)}$:

$$\widetilde{f}_T^{(jk)}(\omega) = \int_{-\pi}^{\pi} K_T^{(jk)}(\omega - \theta)\ \mathbf{I}_T^{(jk)}(\theta)\ d\theta\ , \tag{2.10}$$

where the integration (expressing a convolution of the periodogram with the kernel weight which asymptotically concentrates on frequency ω) is replaced in practice by a summation (over the grid of Fourier frequencies). If the smoothing span $M_T^{(jk)}$, essentially the number of periodogram ordinates over which we smooth, tends asymptotically to infinity, however by a slower rate than sample size T, and 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 one (Brillinger, 1981). The form of its (asymptotic) variance, as a consequence of (2.9), presents another certain challenge,

$$\mathrm{Var}\left(\widetilde{f}_T^{(jk)}(\omega)\right) \sim \frac{1}{M_T^{(jk)}} \int_{-\pi}^{\pi} [K_T^{(jk)}(\theta)]^2 d\theta\ [f^{(jk)}]^2(\omega)\ . \tag{2.11}$$

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 minimisation, for which we refer to the discussion in Section 2.2, and also for subsequent inference (confidence intervals or tests).

The following **Figure 3** shows an example for smoothing the periodogram matrix of **Figure 2** over frequency, here we actually used the more sophisticated wavelet method of Section 3.2 which automatically adapts to the different degree of smoothness over frequency of the different components of the underlying population spectral matrix (displayed in **Figure 1**) by using only one global smoothing parameter (for the choice of which as well as for details on the actual wavelet smoother we refer to Sections 3.1 and 3.2 below).

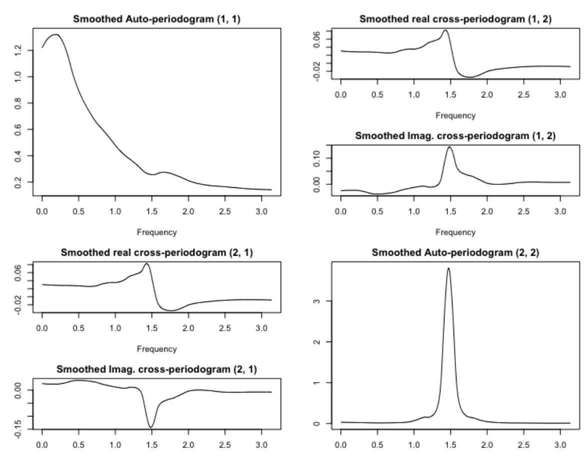

Figure 3: Smoothed Periodogram matrix of the simulated VARMA(2,2)-process with spectrum shown in Figure 1

Furthermore, in order to do inference, asymptotic normality of the smoothed periodogram can be used (we also refer to Section 2.3.2 for more on that).

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 and Davis (1991), Section 11.6). A yet more direct measure of the amount of cross-correlation of two components of $\mathbf{X}(t)$ in the frequency domain is, however, 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 (2.7). 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 (a value near one indicating 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 (2.13) given below). The reader is referred to Section 2.3.2 for an example on (time-varying) coherence analysis of multivariate EEG (**Figure 6**), but at this point, for a first illustration, we rather refer to a simple bivariate example, taken from Brockwell and Davis (1991), Example 11.6.1, where $X_{t1} = Z_t$, $X_{t2} = Z_t + .75Z_{t-10}$, 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 on another (econometrics) example about a price time series lead by a market indicator series by several units in time we refer again to the book by Brockwell and 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. around Figure 7.8.9). The general interest is to identify the frequencies (or frequency bands) which 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 and Davis (1991), Figure 11.7, and Figure 11.8 for the estimated phase).

And finally, it is also common to look on *partial coherence* (as in Park et al. (2014)) as the frequency domain analogue of *partial correlation* (for this concept, widely used for studying brain connectivity, e.g., we simply refer to Brillinger (1981), Sections 8.3 and 8.4).

2.2 The challenges in smoothing the periodogram matrix

Classical text books on modern spectral analysis suggest to use one of the standard nonparametric curve smoothing techniques, such as kernel, nearest neighbour, spline or, more recently, wavelet techniques. It is not the goal of this review to furnish any details on these well-documented methods, however, some preliminary comments are in order to prepare the further ground for treating the non-standard but often very realistic situation of, possibly also time-varying, spectra showing spatially localised features (over frequency and/or over time). According to the author's long year experience with smoothing periodograms, unless we are in the much easier to handle the situation of smoothing the log-periodogram of a univariate time series based on the well-known "Wahba" approximation (Wahba, 1980)

$$\log(I_T(\omega)) \;\approx\; \log(f(\omega)) - \gamma + e(\omega) \;, \tag{2.12}$$

with γ denoting the Euler-Mascheroni constant (= 0.57722...) and with mean centered, constant variance i.i.d. errors $e(\omega)$, it is considerably more difficult to smooth multivariate periodograms over frequency; cf. the discussion around equation (2.11):

1. The first reason is that, as the univariate periodogram, the elements of the $\mathbf{I}_T(\omega)$ are known to be somewhat non-normally distributed, i.e. Wishart in the multivariate case as a generalisation of χ^2 in the univariate case, and this only approximately so, for T sufficiently large. Hence, any (linear) smoothing technique which essentially averages periodograms either over neighbouring Fourier frequencies or, in the stationary case, over neighboring blocks over time (Welch (1967)), works under the paradigm of an asymptotically (not too fastly) growing window (with smoothing span $M_T \to \infty$ but $M_T/T \to 0$ as $T \to \infty$) to achieve consistency, and ideally asymptotic normality. In theory, these results being completely analogous to classical curve estimation under an additive signal-plus-noise model (such as the one of equation (2.12)) are achieved from the (asymptotic) uncorrelatedness of the $\mathbf{I}_T(\omega_\ell)$ over the grid of Fourier frequencies. Nevertheless in practice it might be interesting, instead of basing inference on asymptotic normality of the smoothed periodogram, to work with an "intermediate" approximation based on weighted sums of χ^2-ordinates (the "Satterthwaite" approximation, see, e.g. Brockwell and 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 (cf. equations (2.9) and (2.11)). This poses problem for choosing the smoothing parameter M_T (or the bandwidth) via mean-squared-error minimisation, be it 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 will be revisited 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 which is widely used in time series analysis for exploiting the link between autocovariance and its Fourier transform (cf. equation (2.5)). 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 (2.7)). We come back to this challenge in Section 3.2 on PD-preserving wavelet estimation.

In the next subsection on time-varying spectral analysis these aspects become even more challenging when, e.g. trying to smooth a time-varying periodogram, such as in equation (2.14), not only over frequency ω but also over time - which can be done, e.g., by a two-fold kernel smoother as sketched out below in (2.15).

2.3 Time-varying spectral analysis

The above-discussed spectral analysis is shift-invariant with respect to time, however many time series data sets (such as speed and sound, geophysical or climate data, EEG or ECG, ...) show a time-varying second order behaviour, possibly even with abrupt transients. Many approaches of generalising to time-dependent spectral analysis can be found in the literature, among the first ones in statistical time series analysis the most prominent by Priestley (1965). For our review we focus on two more recent and complimentary representatives which both are statistical - and somewhat model based - approaches built on the idea of a short-time Fourier transform ("windowed" or "segmented" periodogram). Other important and statistically interesting approaches do exist, such as those based on the Wigner-Ville-spectrum, a very localised time-frequency spectrum which 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 to Martin and Flandrin (1995) and Neumann and von Sachs (1997), and to a parametric approach in Dahlhaus (2000a) which is based on the Whittle likelihood.

2.3.1 Locally stationary time series

We start from the rigorous analog of the motivating equation (2.2). The so-called "Cramér spectral representation" of a stationary time series $X(t)$ with spectral density $f(\omega)$ reads in its univariate version as

$$X(t) = \int_{-\pi}^{\pi} A(\omega) \exp(i\omega t) \; dZ(\omega) \; . \tag{2.13}$$

In this equation, the "transfer" function A is related to the spectrum $f(\omega)$ by the relation $f(\omega) = |A(\omega)|^2$, whereas $dZ(\omega)$ denotes an *orthonormal increment process*, i.e. a process with variance one and uncorrelated increments: $\mathrm{Cov}\,(dZ(\omega), dZ^*(\omega')) = \delta(\omega, \omega')$. (Note that often in the literature we find this representation given in a form based on *orthogonal* increments, i.e. normalised such that $\mathrm{Var}\,(dZ(\omega)) = f(\omega)$.) The multivariate spectral representation of $\mathbf{X}(t)$ is analogous (see Brockwell and Davis (1991), Chapter 11.8) and not shown here to keep the exposition not too technical.

R. Dahlhaus wrote a series of seminal papers on locally stationary time series for which we only cite the first two related to the frequency domain, i.e. Dahlhaus (1997) for the univariate approach and Dahlhaus (2000a), including also the case of multivariate processes. An excellent overview of his and all related historical work to be found 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 into 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 $\{X_T(t)\}$ the time-changing spectral content of which is now characterised 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 which 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-localised periodograms along segments of length N (smaller than T):

$$I_{N,T}(u, \omega) = (2\pi)^{-1} N^{-1} \left| \sum_{s=-N/2+1}^{N/2} X_T([uT] + s) \exp(-i\omega s) \right|^2 , \; t = [uT] \in [N/2, T - N/2] \; . \tag{2.14}$$

The direct generalisation of smoothing periodograms of stationary time series given by equation (2.10) 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

$$\widehat{f}(u, \omega) = \sum_i K_b^t(u - u_i) \int_{-\pi}^{\pi} K_h^f(\omega - \theta) \; I_{N,T}(u_i, \theta) \; d\theta \; , \tag{2.15}$$

where K^t and K^f denote some classical kernel weights in time and frequency, with bandwidths b and h respectively (e.g. $h = h_T = M_T^{-1}$, i.e. the reciprocal of the aforementioned smoothing span

M_T) and using the common notation $K_h(\cdot) := h^{-1}K(\cdot/h)$. Subsequent theory 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 has thus been made possible.

2.3.2 SLEX Analysis - a particular localised periodogram approach

A different approach which is less motivated from nonparametric estimation theory is directly built on a computationally efficient modification of the time-localised periodogram (2.14). The idea of a series of papers by H. Ombao and co-workers (Ombao et al. (2001), Ombao et al. (2002), Ombao et al. (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, among others, by 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, without offering a cross-spectral analysis, though.

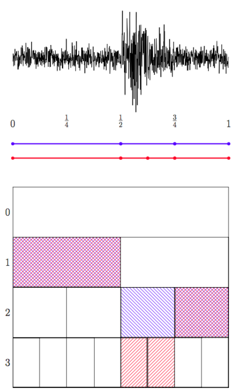

Figure 4: Illustration of adaptively segmenting a non-stationary (EEG) time series signal

Similarly 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 non-redundant information in this collection of SLEX-periodograms and allowing for fast computation along the dyadic blocks of resulting "SLEX library", a specific window function (the SLEX-window) is used which acts like a taper (to reduce the spectral bias occurring with rectangular windows) but preserves at the same time orthogonality (uncorrelatedness) between the segmented periodograms at each Fourier frequency. See the illustrations in **Figure 4** and **Figure 5**.

The key ingredient of this method is the use of a entropy-related cost function in the time-frequency domain the optimum of which 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 for a truly multivariate extension, based on PCA (Principal Component Analysis) in the SLEX frequency domain (Ombao et al., 2005). With the following **Figure 6** 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.

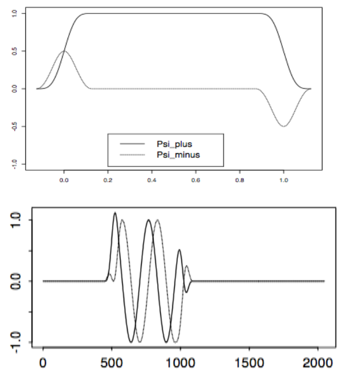

Figure 5: SLEX-window function and a time-localised waveform of the SLEX basis

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 which had previously been in coherent states along a low frequency band (about roughly 10 Hz) become less uniformly coherent, with spectral power shifting to higher frequency. 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, we refer again to Ombao et al. (2005).

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 its appealing and intuitive for practitioners (such as medical doctors in brain data treatment).

3 Modern Nonparametric Methods for Smoothing the Periodogram

Having introduced the challenges in smoothing the periodogram matrix in Section 2.2, we now discuss in this section some "solutions" based on modern nonparametric smoothing methods (kernel, spline, wavelet, ...) adapted to the situation at hand. Whereas the first short part of this section aims at providing the reader with a (non-exhaustive) overview on 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, we devote its second part to illustrating a recent potential solution to the third (and most challenging) problem of preserving positive-definiteness of the resulting spectral estimator.

3.1 Using wavelet thresholding compared to linear smoothers (kernel, spline,...)

Classical nonparametric curve smoothers, such as kernels, with a global or local bandwidth, nearest-neighbour methods, smoothing splines or local polynomials (see Fan and Gijbels (1992) for an excellent overview), in a frequentist or in a Bayesian fashion, have been successfully used to address the problem of smoothing the periodogram matrix. It is not the goal of this review article to compare them among others or discuss how to optimally choose their smoothing parameter - 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 (2.12)) if one takes the additional complications discussed in Section 2.2 into account. We recall however that the heteroscedasticity

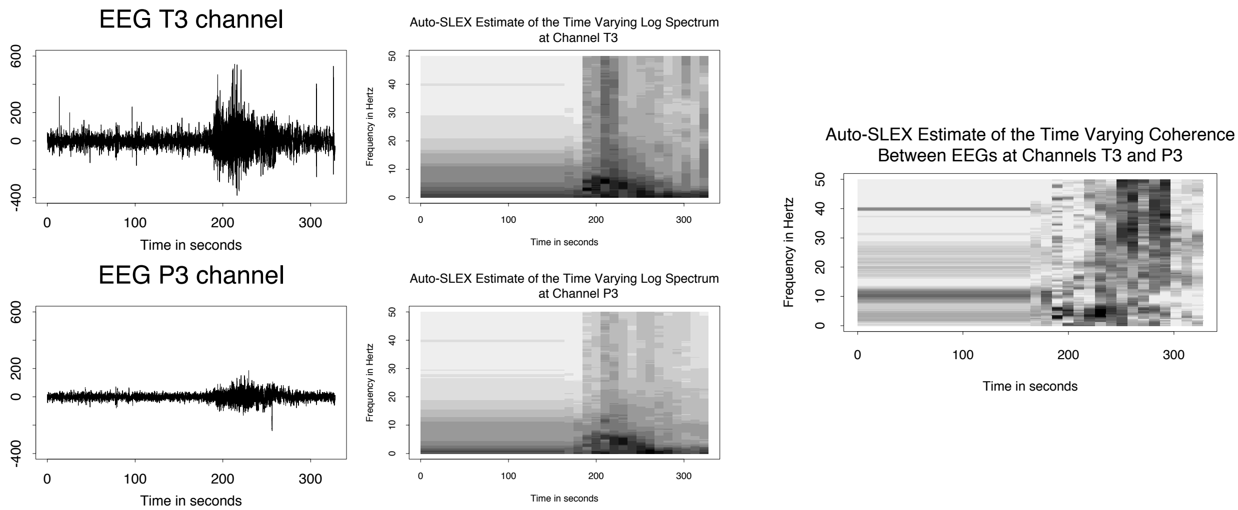

Figure 6: Bivariate SLEX analysis of EEG T3/P3 channels: Auto-spectra and coherency estimators (gray scale amplitudes vary from white=low to black=high values of the time-varying estimators)

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 working with a local smoothing parameter (Bühlmann, 1996), which however amounts to a rather high computational effort in case of a multivariate time series with high dimensionality P, also it makes the data-driven smoothing method highly non-linear in the data.

That is why working with a different, a non-linear, smoother based on wavelet technology attracted its interest about 20 years ago already for univariate periodograms (see, e.g., Neumann (1996), Gao (1997) as well as Moulin (1994) for log-periodograms under the model (2.12). For the details on wavelet smoothing we refer to the cited literature, let us just summarize here what is necessary for grasping the ideas for 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 allow to 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 non-zero coefficients), this signal is compressed in 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 is an additional challenge due to its (finite sample) non-normality and also its heteroscedasticity (as discussed in Section 2.2). This essentially calls for level-and location dependent thresholds: each wavelet coefficients will be thresholded adaptively by its own rule. Unfortunately, and similarly to equation (2.11), 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 non-decimated wavelets (Nason and Silverman, 1995), tree-thresholding (Freyermuth et al., 2010) or the wavelet-Fisz approach (Fryzlewicz et al., 2008)) have been proposed to cope with this. Looking again on **Figure** 3 one can get an impression of the potential of this methodology to produce smoothed version of potentially very irregular curves by choosing, essentially, only one smoothing parameter (the threshold).

3.2 Spectral estimators preserving positive-definiteness (PD)

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 PD*. Remember that PD, similarly to positivity of the spectral density for univariate time series, is important both for computation and interpretation (after application of a PCA, e.g., the spectrum of each component should remain positive). Having motivated the potential of wavelet smoothers to remain spatially adaptive to local structure without requiring a local smoothing parameter, this method directly proposes its investigation for the problem discussed now. In a series of papers, Chau and von Sachs (2017), Chau and von Sachs (2018) and Chau et al. (2019)) suggested to embed the problem more generally into curve estimation on (Riemannian) manifolds - an elegant approach which intrinsically preserves, among other interesting properties, PD of the resulting estimators. The authors successfully implemented their idea by constructing a wavelet-based algorithm that fulfils this program to operate directly on *matrix*-valued curve denoising (or curve-valued matrix estimation). In other words, their constructed wavelet transform is no more 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 actually a Riemannian manifold (this property being the key to success, as discussed below). Distances in this space are no more Euclidean as for classical wavelet-based algorithms, and hence "second-generation"-type wavelet algorithms (Jansen and Ooninckx, 2005) need to be used: Chau and von Sachs found out that using Average-Interpolation algorithms allow to mimic and transfer many of the advantageous properties of scalar to matrix-valued wavelet thresholding. Their generic approach which is neither only taylored to the spectral estimation problem nor to the use of wavelets allows for use of other distances such as the log-Euclidean or the Cholesky (cf also Yuan et al. (2012)). However, it is only the Riemannian distance that shares all desirables properties to provide, beyond PD, for *equivariant* estimators with respect to affine-transformation 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 Krafty and Collinge (2013), Rosen and Stoffer (2007) or Dai and Guo (2004) and Guo and Dai (2006) (the latter two based on smoothing the Cholesky square root of the periodogram matrix).

To illustrate the properties of the aforementioned method (Chau and von Sachs, 2017), we revisit the bivariate brain data set (Local Field Potentials, see **Figure 8**) that will be used also in Section 4 in the context of replicated time series analysis. The following **Figure 7** shows both the auto- and cross-spectral elements of the noisy periodograms and of the denoising wavelet threshold estimator, averaged over each of three different subsets (start, middle and end trials) which are composed of actually 10 trials (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.

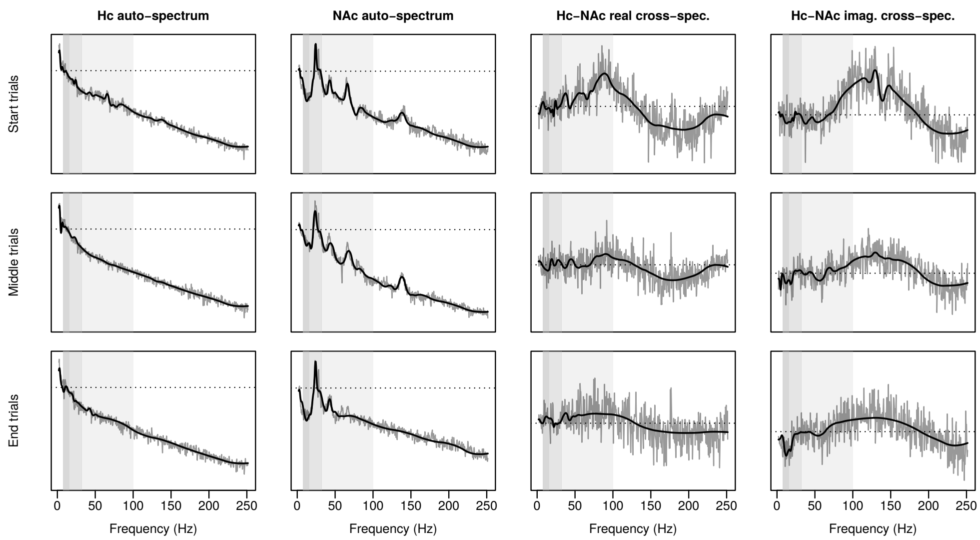

Figure 7: Auto- and cross spectral periodograms and PD threshold wavelet estimators of three different trials of the bivariate Hc/NAc LFP series shown in Figure 8

A more comprehensive study developed in Chau and von Sachs (2017) shows in addition how building on the Riemann distance allows to construct 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 PD-preserving estimation via non-linear wavelet methodology, partly inspired by Rahman et al. (2005), is the proposition of a global smoothing parameter, a kind of universal threshold based on the information in the trace of the spectral (and periodogram) matrix: for this, the authors were able to develop an additive signal-plus-noise model akin traditional scalar curve estimation (and equation (2.12)) which respects the theoretical justification for successful denoising. However, more research in this recent approach would be necessary to allow for a more flexible threshold choice here, in the direction of potentially smoothing each element of the periodogram matrix by its own appropriate threshold. Interestingly enough, the off-diagonal cross-spectral structure often does not show less smoothness over frequency than the autospectra (cf **Figure 1**), which saves us from needing to be too ambitious here. Note finally that Chau and von Sachs (2018) provides for 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 aspect with wide medical and neuroscience applications (but not only so). Think, e.g., about medical studies where subjects (animals, or human patients) are exposed to some supervised experiments ("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 disposes of only one time series set. However, assuming independent replicates is not always very realistic in practice. Hence, Chau and 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 chapter we discuss now only a few approaches on replicated time series analysis which should reflect the recent evolution of these kind of models specifically in the frequency domain.

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 to apply less stringent smoothing over frequency (e.g. just enough to make the periodogram matrix full rank) and hence gain frequency resolution, or also to treat data with a time-varying spectral content without needing to smooth a lot 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. Widely used in the time domain, there are far less contributions 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 $s-$th subject specific (log-) periodogram $Y^s(\omega) := \log I_T(\omega)$, at a given frequency ω, is modelled 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 non-systematic (i.i.d.) error $e^s(\omega)$ (which for each $s = 1, \ldots, S$ follows the same approximative $\log(\chi_2^2/2)-$distribution as the error in the Wahba approximation of equation (2.12)):

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

Here the errors $e^s(\omega)$ are independent between different replicates s and independent of the functional random effects $U^s(\omega)$.

In the context of wavelet smoothing of the observed (log-) periodograms, Freyermuth et al. (2010), and Chau and von Sachs (2016) allowing for a correlation structure within the subject population (i.e. the $U^s(\omega)$), made some progress on transferring the advantages of non-linear wavelet thresholding (cf Section 3.1) to this more refined set-up, where they develop some proper theory of including the mixed-effects situation. It is yet an open problem how to render this promising approach fully multivariate, 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, e.g., in Gorrostieta et al. (2012) and Fiecas and Ombao (2016). As an alternative to treating time-varying spectral analysis as exposed in Section 2.3, in Fiecas and Ombao (2016) the replicates are provided by cutting the time-line of an associative learning experiment of an animal in a brain study up into a series of (time-ordered) trials, and then develop a model of these "trial-replicated" spectral analysis. The authors can clearly trace the evolution of the animal acquiring learnt patterns over the time-ordered trials which shows up in the evolution of estimated spectral structure (changing linear dependencies) of the local field potential (LFP) data of the animals brain. (A different though related analysis has been provided by the estimators in **Figure 7**.) To give an idea about the nature of these data, in **Figure 8** on a bivariate subset we show, just for illustration, three different trials, of 2s duration each, recorded in the hippocampus (Hc) and the nucleus accumbens (NAc) 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 sufficiently well model many biological signals with complex dependence structures such as cross-oscillatory interactions between a low-frequency component in one signal and a high-frequency

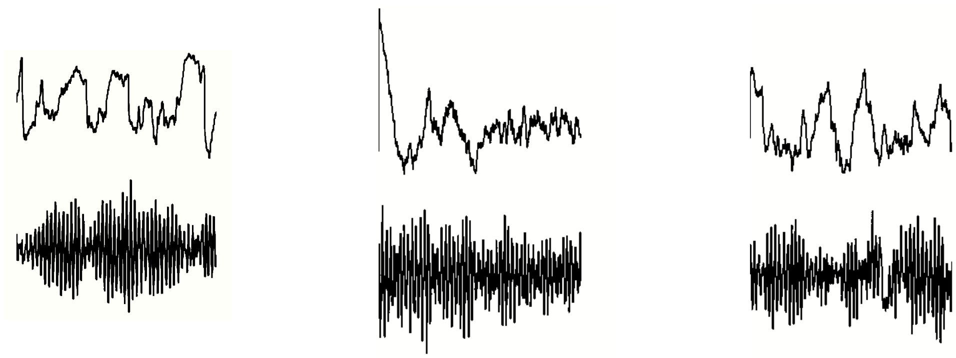

Figure 8: A subset of three different trials of a recorded LFP bivariate brain data set (HC top, NAc bottom plots)

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

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 non-regular (and hence non-invertible). To face this problem different *regularisation* approaches have been proposed in the literature, mostly in the context of the related problem of regularisation of high-dimensional covariance matrices. Whereas a popular method is based on banding, tapering or thresholding (discussed among many other methods in the excellent overview article by Pourahmadi (2011)), we focus on the complimentary approach based on *shrinkage*. This is supposed to work without any assumption of sparsity (typically rarely fulfilled for spectral density matrices, opposed to covariance matrices for which the first type of methods work seemingly well). This method which essentially amounts to adding a "ridge" (a multiple of the identity matrix) to the covariance matrix has been first developed by the seminal work of Ledoit and Wolf (2004). While the linear shrinkage" approach, being particularly simple, has entered into spectral analysis in a series of papers which we will briefly discuss now, its non-linear version, developed in Ledoit and Wolf (2012), is an interesting but rather impractical generalisation, related to random matrix theory (a practical improvement has been proposed, however, recently by the authors in Ledoit and Wolf (2017)). Note that this research has triggered a variety of related alternative regularisation approaches which to discuss would however go beyond the purpose of this review.

Linear shrinkage of covariance or spectral density matrices amounts to finding the (asymptotic) mean-squared-error (MSE) optimal linear combination between the given empirical covariance or periodogram matrix and a prespecified target which regularises the former. The solution enjoys good statistical and numerical properties having in general a far less spread of its eigenvalues, i.e improving on its condition number (the ratio of largest to smallest eigenvalue), and in particular lifting the smallest eigenvalues away from zero. Böhm and 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 to be applied in absence of any a priori information. In structural shrinkage, as in Böhm and 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 modelling 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) combines 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, e.g., in several papers by M. Fiecas. Fiecas and von Sachs (2014) develop a "time-frequency-toggle" bootstrap method based on Kirch and Politis (2011), to choose the optimal shrinkage parameter in the mixture of a kernel smoothed high-dimensional periodogram with any suitable regularisation target, whereas Fiecas and Ombao (2011), following the idea of Böhm and von Sachs (2008), develop a generalised 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 (non-exhaustive) series of approaches to spectral analysis which are not based on classical Fourier analysis.

6.1 Dual-frequency spectral analysis for harmonizable processes

Motivated from a refined analysis of electroencephalogram data recorded in a motor intention experiment, the work of Gorrostieta et al. (2012), and of Gorrostieta et al. (2019) for the more realistic case of time-varying spectral structure, allows for modelling 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 (2.13) of stationary processes by the one of of *harmonizable processes*. For those, their increment processes are no more orthogonal over frequency but rather fulfil

$$\mathrm{Cov}(dZ(\omega_1), dZ(\omega_2)) = f(\omega_1, \omega_2)\ d\omega_1 d\omega_2\ , \tag{6.1}$$

where $f(\omega_1, \omega_2)$ is the so-called "Loève"-spectrum (Lii and 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 are 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 to the discussion of the aspect of replicated time series analysis in Section 4 above.

6.2 Wavelet spectra

It seems to be an obvious idea trying to replace Fourier analysis of time series by something which is similar in spirit though more localized in time. Having learnt about the success story of wavelets in statistics (primarily for curve estimation, cf. also Section 3.1), several researches, at the end of the last century, thought about the idea to use them as an alternative to a (local) Fourier analysis. In their work 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 (2.13), replacing the Fourier by a (particular) wavelet basis $\{\psi_{jk}(t)\}$, localised at (dyadic) scale $j \geq 0$ and locations $k = 0, \ldots, 2^j - 1$,

$$X(t) = \sum_j \sum_k w_{jk} \ \psi_{jk}(t) \ \xi_{jk} \ . \tag{6.2}$$

Here the amplitudes w_{jk} take the role of the localised amplitude functions $A_t(\omega)$ discussed in Section 2.3.1, whereas the uncorrelated sequence (over j and k) ξ_{jk} can be seen as discrete version of the increment process $dZ(\omega)$, of a time-dependent Cramér spectral representation. Here, the idea is that scale j replaces (the reciprocal) of (a discretized) frequency ω whereas local time t would correspond to a localisation 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 a 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 (6.2) Nason and von Sachs had to use, as building blocks, "non-decimated" wavelets (Nason and Silverman, 1995) which 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 more orthogonal basis functions, and 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 $u \in [0, 1]$ (cf. Section 2.3.1) amounts to w_{jk}^2 at localisation $u = k/T$. However, similarly to Dahlhaus' theory of evolutionary spectra (Dahlhaus, 2012), they work with some doubly-indexed array of processes arising from equation (6.2) to allow for uniquely defined limiting spectra, accompanied by a theory of its 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 $X(t)$, or rather (as for the classical Fourier periodogram) a smoothed version of it (smoothing 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 is 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 its flavor, triggered a long series of subsequent work, both on modelling and on practical application of this 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 generalisations of (6.2) 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 interesting in itself, it remains questionable whether the benefit of working with a truly local (compared to the non-local Fourier) basis is not somewhat overruled by the fact that phase interpretation is much more difficult in the wavelet domain (compared with the elegant property of Fourier functions, with linear phase, to give direct information about lead and leg of two components of a multivariate time series).

6.3 Spectral analysis for locally stationary Hawkes processes

Locally stationary (multivariate) *Hawkes processes* are a class of time-inhomogeneous (self-exiciting) point processes, spectral analysis for which has been developed in Roueff et al. (2016) and Roueff and 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

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

where the integral of p with respect to the counting process N is to be understood as 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 modelling phenomena in, e.g., seismology, 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 here the analog of equation (2.13), i.e. the spectral decomposition of the variance of a point process N, allowing to subsequently define its "Bartlett" spectrum:

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

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. $\widehat{g}(\omega) = \int g(t)\exp(-it\omega)dt$. For stationary Hawkes processes with immigrant intensity λ_c and fertility function p, the Bartlett spectral density is then given by

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

Motivated by the concept of locally stationary autoregressive processes as developed by R. Dahlhaus (Dahlhaus, 2012), inherently different techniques are applied in the work by Roueff and von Sachs to describe the time-varying dynamics (i.e. a time-dependent Bartlett spectrum $f(t,\omega)$) of self-exciting point processes, modelled via now time-depending 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 algorithm for estimating these objects (including rates of convergence) were delivered by the second work of Roueff and von Sachs (2019) where we also find an application to order-book tick by tick data modelled 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, does only take structure of *linear* dependencies in a given time series into account. A variety of approaches exist trying to remedy this drawback, to allow for more general than 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 arising in climatology, and more generally also for spatial-temporal data, e.g. First ideas of quantile spectral analysis can already be found in Parzen (1985) but it is the recent series of articles by T. Kley, S. Vogulshev, S. Birr and their co-workers (starting from Dette et al. (2015)) that has developed rigorous modern statistical concepts in this field, including cross-spectral analysis, copula spectral densities and embedding of those into locally stationary process theory (Birr et al., 2017). In a nutshell, this type of spectral analysis is based on Fourier transforms of the following two types of generalised covariance, here given only for the case of auto-covariance of a univariate strictly stationary process $\{X_t\}_t$, but with obvious modifications for the multivariate case, and what is more, also time-varying versions of it:

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

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

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

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 times series has, however, been developed in parallel by R. Davis, T. Mikosch and various co-authors, who invented the concept of *extremograms* (of uni- and multivariate time series) - we refer to the overview in Davis et al. (2013) on recent measures of serial extremal dependence in a strictly stationary time series as well as their estimation. For completeness, here we just recall the definition of the extremogram in the univariate case, essentially akin a (asymptotic) correlogram for extreme events; its generalisation to the cross-extremogram can be found in the cited literature. In fact, the extremogram is defined as a limiting sequence given by

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

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 $T\ P(|X| > a_T) \sim 1$, which leads to $\gamma_{AB}(h) = \lim_{T\to\infty} T\ P(X_0 \in a_T A, X_h \in a_T B)$. Motivating examples of extremograms are the limiting conditional probabilities $\lim_{T\to\infty} T\ P(X_h \in a_T B | X_0 \in a_T A)$. 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 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 non exhaustive review on (multivariate) spectral analysis we tried to alert the reader to the existing challenges in this area and how to (partially) address them. We somehow focused on the author's experience with various approaches to smoothing the (uni-or multivariate) periodogram statistics, over frequency, possibly avoiding to lose the important property of a positive definite 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 just mentioned a few approaches which 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 just want to mention a number of important issues not being addressed here for which we need to apologize beforehand: Parametric spectral analysis based on VARMA-modelling (still very much used in fields such as econometrics), semi-parametric approaches, based e.g. on the Whittle Likelihood (see again Dahlhaus (2012), or also in particular Dahlhaus and Neumann (2001) on fitting semi-parametric models to nonstationary time series and Dahlhaus and Polonik (2006) on nonparametric quasi maximum likelihood estimation for locally stationary time series). A further important field being under full development and having become important for high-dimensional data modelling are *graphical models* for time series (Dahlhaus (2000b), Eichler (2012)). We did not touch either on the very active field of *functional time series* for which most of the ideas of spectral analysis can be transferred to analysing curve-valued time series, the necessary developments for this though being highly non-trivial (Panaretos and Tavakoli (2013), Hörmann et al. (2015) and van Delft and Eichler (2018), e.g.) Also did we neither discuss aspects of *outlier-robust* spectral analysis (von Sachs, 1994) nor what to do with missing values - both problems being extremely important for practical time series analysts. Another important field of application of spectral analysis is *discrimination* of time series, in the frequency domain (Shumway and Stoffer (2006), Huang et al. (2004)). And finally, again from a more econometric point of view, we should mention the use of spectral estimation for *dynamic factor analysis* - such as in a 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 ideosynchratic structure - possibly as general as a locally stationary process (Eichler et al., 2011). *Bayesian* spectral analysis (Rosen et al. (2009), Rosen et al. (2012) including both Matlab and R code) is an important field that has not been discussed explicitly either, nor spectral analysis of qualitative or *categorical* time series (Stoffer (2015), Krafty et al. (2012)).

Clearly, this review put a lot more emphasis on existing model approaches rather than talking about statistical estimation (apart from the specific aspect of periodogram smoothing) and inference. Of course, the latter one is equally important, and space permitting, a more comprehensive overview about different inference methods would be rewarding, too (and possibly left for a different occasion), including a variety of recent mathematical-statistical work on testing problems related to spectral analysis (see, e.g., Dette and Paparoditis (2009)). Finally, researchers who are experts in resampling methods such as the *bootstrap* (see Franke and Härdle (1992), Jentsch and Kreiss (2010), Kirch and Politis (2011), and the excellent overview in Kreiss and Paparoditis (2019)) have contributed important methodology to a field where due to the (even asymptotically) challenging variance-covariance structure of spectral estimators (compare equation (2.11) 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 practically working algorithms, with precise guidelines how and in which situation they can safely be applied. This review tried to give some examples where nowadays more and more frequently arising non-standard situations - such as replicated and non-stationary multivariate time series data and/or high-dimensional serially correlated data, possibly with outliers and/or missing values - present challenges to which researches in this field need to give practical solutions. There is still a need for better understanding the developed algorithms and their properties (such as preserving positive-definiteness or their regularisation capacity for high dimensional data), and for increased interaction with practitioners. Whereas time series analysis in general has found its way for some time now 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 modelling and analysis in neuroscience, where however a new challenge has been arisen by combining more classical spectral analysis of EEG data, known to give high frequency resolution, by incorporation of fMRI (functional magnetic resonance imaging) data, being complementary with their better spatial-temporal resolution (Wang et al., 2017). It remains a challenge to further explore this combination from the statistical side.

ACKNOWLEDGMENTS

The author wants to thank J. Chau and M. Fiecas for some help preparing this review. At the same time he apologizes for any missing reference he should have cherished.

References

Adak, S. (1998). Time dependent spectral analysis of non-stationary time series. *Journal of the American Statistical Association 93*(446), 1488–1501.

Birr, S., S. Volgushev, M. Hallin, T. Kley, and H. Dette (2017). Quantile spectral analysis for locally stationary time series. *Journal of the Royal Statistical Society, Series B 79*, 1619–1643.

Böhm, H., H. Ombao, R. von Sachs, and J. Sanes (2010). Classification of multivariate non-stationary signals: The SLEX-shrinkage approach. *Journal of Statistical Planning and Inference 140*, 3754–3763.

Böhm, H. and R. von Sachs (2008). Structural shrinkage of nonparametric spectral estimators for multivariate time series. *Electronic Journal of Statistics 2*, 696–721.

Böhm, H. and R. von Sachs (2009). Shrinkage estimation in the frequency domain of multivariate time series. *Journal of Multivariate Analysis 100*, 913–935.

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

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

Bühlmann, P. (1996). Locally adaptive lag-window spectral estimation. *Journal of Time Series Analysis 17*, 247–270.

Chau, J. (2018). *pdSpecEst: Positive-Definite Wavelet-Based Multivariate Spectral Analysis*. R package version 1.2.3.

Chau, J., H. Ombao, and R. von Sachs (2019). Intrinsic data depth for Hermitian positive definite matrices. *Journal of Computational and Graphical Statistics*, to appear.

Chau, J. and R. von Sachs (2016). Functional mixed effects wavelet estimation for spectra of replicated time series. *Electronic Journal of Statistics 10*(2), 2461–2510.

Chau, J. and R. von Sachs (2017). Intrinsic wavelet regression for curves of Hermitian positive definite matrices. *ArXiv preprint 1701.03314; in revision*.

Chau, J. and R. von Sachs (2018). Intrinsic wavelet regression for surfaces of Hermitian positive definite matrices. *ArXiv preprint 1808.08764*.

Cooley, J. and J. Tukey (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation 90*, 297–301.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics 25*(1), 1–37.

Dahlhaus, R. (2000a). A likelihood approximation for locally stationary processes. *The Annals of Statistics 28*, 1762–1794.

Dahlhaus, R. (2000b). Graphical interaction models for multivariate time series. *Metrika 51*, 157–172.

Dahlhaus, R. (2012). *Locally stationary processes*, Chapter in Time Series Analysis: Methods and Applications, Vol. 30, pp. 351–413. Amsterdam: Elsevier B.V.

Dahlhaus, R. and M. Neumann (2001). Locally adaptive fitting of semiparametric models to nonstationary time series. *Stochastic Processes and their Applications 91*, 277–308.

Dahlhaus, R. and W. Polonik (2006). Nonparametric quasi maximum likelihood estimation for Gaussian locally stationary processes. *The Annals of Statistics 34*, 2790–2824.

Dai, M. and W. Guo (2004). Multivariate spectral analysis using Cholesky decomposition. *Biometrika 91*, 629–643.

Davis, R., T. Mikosch, and Y. Zhao (2013). Measures of serial extremal dependence and their estimation. *Stochastic Processes and their Applications 123*, 2575–2602.

Davis, R., L. T, and G. Rodriguez-Yam (2006). Structural break estimation for non-stationary time series models. *Journal of the American Statistical Association 101*(473), 223–239.

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

Dette, H. and E. Paparoditis (2009). Bootstrapping frequency domain tests in multivariate time series with an application to comparing spectral densities. *Journal of the Royal Statistical Society B 71*, 831–857.

Eichler, M. (2012). Graphical modelling of multivariate time series. *Probability Theory and Related Fields 153*, 233–268.

Eichler, M., G. Motta, and R. von Sachs (2011). Fitting dynamic factor models to non-stationary time series. *Journal of Econometrics 163*, 51–70.

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

Fiecas, M., J. Franke, R. von Sachs, and J. Tadjuidje (2017). Shrinkage estimation for multivariate hidden Markov mixture models. *Journal of the American Statistical Association 112*, 424–435.

Fiecas, M. and H. Ombao (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *The Annals of Applied Statistics 5*, 1102–1125.

Fiecas, M. and H. Ombao (2016). Modeling the evolution of dynamic brain processes during an associative learning experiment. *Journal of the American Statistical Association 111*(516), 1440–1453.

Fiecas, M. and R. von Sachs (2014). Data-driven shrinkage of the spectral density matrix of a high-dimensional time series. *Electronic Journal of Statistics 8*(2), 2975–3003.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). Fitting dynamic factor models to non-stationary time series. *The Review of Economics and Statistics 82*, 540–554.

Franke, J. and W. Härdle (1992). On bootstrapping kernel spectral estimates. *Annals of Statistics 20*, 121–145.

Freyermuth, J.-M., H. Ombao, and R. von Sachs (2010). Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series. *Journal of the American Statistical Association 105*, 634–646.

Fryzlewicz, P., G. Nason, and R. von Sachs (2008). A wavelet-Fisz approach to spectrum estimation. *Journal of Time Series Analysis 29*, 868–880.

Gao, H. (1997). Choice of thresholds for wavelet shrinkage estimate of the spectrum. *Journal of Time Series Analysis 18*, 231–251.

Gorrostieta, C., H. Ombao, R. Prado, S. Patel, and E. Eskandar (2012). Exploring dependence between brain signals in a monkey during learning. *Journal of Time Series Analysis 33*(5), 771–778.

Gorrostieta, C., H. Ombao, and R. von Sachs (2019). Time-dependent dual-frequency coherence in multivariate non-stationary time series. *Journal of Time Series Analysis 40*(1), 3–22.

Guo, W. and M. Dai (2006). Multivariate time-dependent spectral analysis using Cholesky decomposition. *Statistica Sinica 16*, 825–845.

Hörmann, S., L. Kidzin'ski, and M. Hallin (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society, Series B 77*, 319–348.

Huang, H.-Y., H. Ombao, and D. Stoffer (2004). Discrimination and classification of nonstationary time series using the SLEX model. *Journal of the American Statistical Association 99*, 763–774.

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

Jentsch, C. and J.-P. Kreiss (2010). The multiple hybrid bootstrap resampling multivariate linear processes. *Journal of Multivariate Analysis 101*, 2320–2345.

Kirch, C. and D. N. Politis (2011). TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *Annals of Statistics 39*, 1427–1470.

Koopmans, J. (1974). *The Spectral Analysis of Time Series*. Academic Press: Elsevier.

Krafty, R. and W. Collinge (2013). Penalized multivariate Whittle likelihood for power spectrum estimation. *Biometrika 100*, 447–458.

Krafty, R., M. Hall, and W. Guo (2011). Functional mixed effects spectral analysis. *Biometrika 98*(3), 583–598.

Krafty, R., S. Xiong, D. Stoffer, D. Buysse, and M. Hall (2012). Enveloping spectral surfaces: covariate dependent spectral analysis of categorical time series. *Journal of Time Series Analysis 33*(5), 797–806.

Kreiss, J.-P. and E. Paparoditis (2019). *Bootstrap for Time Series: Theory and Applications*. New York: Springer; to appear.

Ledoit, O. and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*, 365–411.

Ledoit, O. and M. Wolf (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics 40*, 1024–1060.

Ledoit, O. and M. Wolf (2017). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *Economics Working Paper No. 264, University of Zurich*.

Lii, K. and M. Rosenblatt (2002). Spectral analysis for harmonizable processes. *Annals of Statistics 30*(1), 258–297.

Martin, W. and P. Flandrin (1995). Wigner Ville spectral analysis of nonstationary processes. *IEEE Trans. Signal Processing 33*, 1461–1470.

Moulin, P. (1994). Wavelet thresholding techniques for power spectrum estimation. *IEEE Trans. Signal Processing 42*(11), 3126–3136.

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

Nason, G. and B. Silverman (1995). *The stationary wavelet transform and some statistical applications*, Chapter in Lect. Notes Statist., Vol. 103, pp. 281–300. New York: Springer.

Nason, G., R. von Sachs, and G. Kroisandt (2000). Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum. *Journal of the Royal Statistical Society, Series B 62*(2), 271–292.

Neumann, M. (1996). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *Journal of Time Series Analysis 17*(6), 601–633.

Neumann, M. and R. von Sachs (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Annals of Statistics 25*, 38–76.

Ombao, H., J. Raz, R. von Sachs, and W. Guo (2002). The SLEX model of a non-stationary random process. *Annals Institute Statistical Mathematics 54*(1), 171–200.

Ombao, H., J. Raz, R. von Sachs, and B. Mallow (2001). Automatic Analysis of Bivariate Non-Stationary Time Series. *Journal of the American Statistical Association 96*(454), 543–560.

Ombao, H., R. von Sachs, and W. Guo (2005). SLEX Analysis of Multivariate Non-Stationary Time Series. *Journal of the American Statistical Association 100*(470), 519–531.

Panaretos, V. and S. Tavakoli (2013). Fourier analysis of stationary time series in function space. *Annals of Statistics 41*, 568–603.

Park, T., I. Eckley, and H. Ombao (2014). Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Trans. Signal Processing 62*, 5240–5250.

Parzen, E. (1985). Time series model identification and quantile spectral analysis. *IFAC Proceedings Volumes 18*(5), 731–736.

Pourahmadi, M. (2011). Covariance estimation: The GLM and regularization perspectives. *Statistical Science 26*, 369–387.

Priestley, M. (1965). Evolutionary spectra and nonstationary processes. *Journal of the Royal Statistical Society, Series B 27*(2), 204–237.

Rahman, I., I. Drori, V. Stodden, D. Donoho, and P. Schröder (2005). Multiscale representations for manifold-valued data. *Multiscale Modeling & Simulation 4*(4), 1201–1232.

Rosen, O. and D. Stoffer (2007). Automatic estimation of multivariate spectra via smoothing splines. *Biometrika 94*(2), 335–345.

Rosen, O., D. Stoffer, and S. Wood (2009). Local spectral analysis via a bayesian mixture of smoothing splines. *Journal of the American Statistical Association 104*, 249–262.

Rosen, O., D. Stoffer, and S. Wood (2012). Adaptspec: Adaptive spectral estimation for nonstationary time series. *Journal of the American Statistical Association 107*, 1575–1589.

Roueff, F. and R. von Sachs (2019). Time-frequency analysis of locally stationary Hawkes processes. *Bernoulli 25*(2), 1355–1385.

Roueff, F., R. von Sachs, and L. Sansonnet (2016). Locally stationary Hawkes processes. *Stochastic Processes and their Applications 126*, 1710–1743.

Sanderson, J., P. Fryzlewicz, and M. Jones (2010). Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika 97*, 435–446.

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

Stoffer, D. (2015). *Spectral analysis of qualitative time series*, Chapter in Handbook of Discrete-valued Time Series, pp. 287–309. London: CRC Press.

Subba Rao, S. (2018). Orthogonal samples for estimators in time series. *Journal of Time Series Analysis 39*, 313–337.

van Delft, A. and M. Eichler (2018). Locally stationary functional time series. *Electronic Journal of Statistics 12*, 107–170.

von Sachs, R. (1994). Peak-insensitive nonparametric spectrum estimation. *Journal of Time Series Analysis 15*, 429–452.

Wahba, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association 75*(369), 122–132.

Wang, Y., C.-M. Ting, and H. Ombao (2017). Estimating dynamic connectivity states in fmri using regime-switching factor models. *IEEE Transactions on Medical Imaging 37*(4), 1011–1023.

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 and Electroacoustics 15*(2), 70–73.

Yuan, Y., H. Zhu, W. Lin, and J. Marron (2012). Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society, Series B 74*(4), 697–719.