

Big-data Feature Analysis in a Cyclostationary Model Framework

François A. Marshall*

February 27, 2023

1. Abstract

When a harmonic signal appears in noise and the residual noise is red, this may be evidence for residual periodicity, and a simple model for this is to assume the presence of normal modes in the power spectrum. As long as spectral power does not diverge too rapidly approaching zero frequency, it may also be true that the finite-dimensional distribution (FDD) exhibits normality. Here, two visual diagnostics are introduced that make use only of the power spectrum to characterize the FDD. The first diagnostic specifies the quality of harmonic periodicity in both the mean and spectral-power signals, while the second reveals the extent to which the joint FDD of the multitaper Fourier-transform processes exhibits spherical, proper, Gaussian behaviour. It is shown for a two-sample survey of epileptic-seizure, microelectrode-voltage time series how these visualization techniques explain time-evolutionary periodicity for long records in a manner more efficient than a typical spectrogram.

2. Introduction and theory

Often in spectral analysis, a good diagnostic ought to produce an intuitive picture of the generative dynamics underlying a physical time series. e.g., when neuroscientists stare at a spectrogram the detailed mathematics are unnecessary when it comes to identifying stereotyped periodicity - and even constructing biophysical models. Here in this manuscript is introduced a low-complexity diagnostic tool akin to the spectrogram but more amenable to large complex surveys. Two time series are considered from clinical experiments conducted on epileptic-seizure patients - spanning two patient strata (designated Patients C5 and MG49). Microelectrode arrays (MEA) were used to obtain voltage data from the neocortex (for details of the experiment and data, refer to [8, 14, 15]). One electrode recording per patient is here considered (labeled “Electrode 1” from a common MEA), and the j 'th time series is divided into H_j strata of size, N ($N = O(10^3)$).

The simplest way to model a cyclostationary process is using a signal-plus-noise model whose mean signal contains the interesting features (a low-variance approximation). This way, the power spectrum explains much of the cyclostationarity present, in spite of several support curves from the bifrequency plane having been missed out [10]. For a multitaper spectral analysis and associated hypothesis tests for line components, reliability is determined both by record size and the extent to which is joint-spherical and proper the FDD

*2022 Visiting Scholar in the Department of Mathematics and Statistics of Boston University, 111 Cummington Mall #140C, Boston, MA 02215, United States

of the eigencoefficient processes (special low-bias tapered Fourier-transform processes). Too short a record size, and the resolution bandwidth of spectral estimates is too high - not to mention the non-Gaussian effects that would affect test power [6].

Consider the h 'th temporal stratum. Denote the stratum time series by $\mathbf{x} \in \mathbb{R}^N$. Assume that it has been realized from the random vector, \mathbf{X} , where for $t_n \in [0, t_{N-1}]$, $\mathbf{X}[n+1] = X(t_n)$. Here, X is an L^p -process ($p \in \mathbb{N}_{>1}$, [3]), while $t_n = n\Delta t$ for Δt the sampling period. Next, assume that the baseband- B demodulate, X_B , of X has the decomposition [5, 20],

$$X_B(t) \stackrel{w.p.1}{=} \sum_{m \in \mathbb{Z}_{\geq 0}} \chi_B(f_m) \cdot |X_m| \cdot \cos \left(2\pi f_m \frac{t}{\Delta t} + \text{Arg}[X_m] \right) + X^{(NSS)}(t), \quad (1)$$

where the terms are given as follows.

1. X_m is the frequency- f_m normalized spectral jump of X [5], while f_m is the corresponding normalized frequency and given by $f_m = \frac{m}{M}$ [20].
2. $|\cdot|$ and Arg respectively denote the modulus and principal-argument operations from complex analysis [11]. The χ_B function is the indicator function given B [11]. The notation, *w.p.1*, stands for “with probability one”, referring to the series limit in the representation of the right-hand side [1, 2, 11].
3. $X^{(NSS)}$ is a zero-mean, narrow-sense-stationary (NSS) stochastic process.

Assume that the mean signal of X_B has the decomposition [20],

$$E \circ X_B(t) = \mu_0 + \sum_{m=1}^{\frac{M}{2}-1} \chi_B(f_m) \cdot \left\{ |\mu_m| \cdot \cos \left(2\pi f_m \frac{t}{\Delta t} + \text{Arg}[\mu_m] \right) + \sum_{u=-\frac{M}{2}+1}^{\frac{M}{2}-1} g_m \left(\frac{t}{\Delta t} - \frac{u}{f_m} \right) \right\}, \quad (2)$$

where: M is constant; $\mu_m \in \mathbb{C}$ is the m 'th Fourier coefficient; and g_m is a $\left\{ \frac{1}{2f_m} \cdot [-1, 1] \right\}$ -timelimited cosine-series signal [17]. In Equation 2: the first part of the m 'th term corresponds to a normal mode of X (it is said that the sum, $\mu^{(HMC)}$, of the normal modes yields the harmonic (HMC) mean signal of X); and the second part corresponds to the m 'th echo signal (it is said that the sum of the echo signals yields the echo mean of X). The noise (NSE), $X^{(NSE)}$, of X is defined as $X^{(NSE)} = X - \mu^{(HMC)}$. The Fourier-transform signal of log spectral X -power is called the cepstrum of X , and the abscissa of the cepstrum are called quefrequencies [20]. For the model in Equation 2, the m 'th quefrequency is equal to $\frac{1}{f_m}$.

3. Methods

In the multitaper analyses this manuscript describes, M was chosen using the methods discussed in [7, 8]. The multitaper bandwidth parameter, W , was set as $NW = 5$. To identify and infer the nonzero μ_m from Equation 2, conduct a multitaper harmonic F-test at each point on the fast-Fourier-transform (FFT)-grid at the $\approx \left(1 - \frac{1}{N} \right) \cdot 100\%$ level¹. Here, the zero-padded version of \mathbf{x} has size M . The same steps are

¹All F-tests discussed in this manuscript have been performed using precisely the same methods as detailed in [8] - where is explained the presence of the \approx symbol. Remark: relying on a ratio test statistic, the F-test is invariant to linear-filter effects.

used to identify the nonzero g_m , except now conducting F-tests for the time series that comprises estimates of log spectral $X^{(NSE)}$ power. Here, the frequency- f_m log-power estimate has been computed using the natural logarithm of the corresponding multitaper spectral X power estimate. Remark how: for the F-tests associated with X , the approximation has been made that $X^{(NSE)}$ is NSS. The approximation holds valid under a regime of high signal-to-noise ratio. The same kind of approximation has been made for log spectral power, since its m 'th mean signal element is the sum of: a period- f_m cosine series associated with log spectral g_m direct-current power; and the sum of log spectral g_m alternating-current power and log spectral $\mu^{(HMC)}$ power. The approximation has been made of treating as NSS the combination of log spectral $X^{(NSS)}$ power with the latter of these two summed signals.

Moving on to the quality assessment for the multitaper spectral analysis, the problem for each eigencoefficient vector is to conduct tests for both aspherical and improper behaviours. Denote by $T^{(SPH)}$ and $T^{(IMP)}$ the multitaper sphericity (SPH) and impropriety (IMP) test statistics, respectively (the asphericity and impropriety tests respectively are those from: [4, 19, 20]; and [16, 21]). Corresponding to these statistics are the respective cumulative distribution functions, $F^{(SPH)}$ and $F^{(IMP)}$ - i.e., the probability integral transforms (PIT, [6]). Under the null hypotheses of considered multitaper hypothesis tests: the empirical spectral-distribution functions (SDF) associated with these functions together evince spherical, proper behaviour of the eigencoefficient vectors. The following explains how to construct a novel diagnostic plot that illustrates the co-occurrence of these distributional behaviours.

1. Choose an FFT-index bin width of $\lfloor 2MW \rfloor^2$, and set a frame of $\left[0, \frac{M}{2}\right]^2$. In each bin, compute the average of $(F^{(SPH)}, F^{(IMP)})$ -points.
2. Set a frame of $[0, 1]^2$, and situate each of the $(F^{(SPH)}, F^{(IMP)})$ -averages in it. For the resulting scatter plot of these averages, infer the first two principal-component (PC) axes. Translate the intersection of these axes to the mean $(F^{(SPH)}, F^{(IMP)})$ -average coordinate, and then truncate each axis at one PC standard deviation from the mean of the average vectors.

4. Results and Discussion

For the Patient C5 stratum, Figure 1 displays diagnostic plots. Both upper plots are typical of what neuroscientists explore to explain seizures [12, 13, 18]. The spectrogram in B) tracks the voltage trace of A): starting with minimal autocorrelation structure < 10 s into record (low spectral activity), the seizure disturbs voltage [10, 70] s (ripples along the frequency axis) before dying down > 70 s (low activity). The bottom two plots display the novel contributions of this manuscript. In A), two over-frequency phasor-boxplot spectrograms are plotted: one for the μ_m ; and one for the corresponding Fourier coefficients in the log-power representation. Red cyclic phasor clocks are scattered across time; unsurprising given that $X^{(NSE)}$ contributes during most intervals. Black harmonic phasor clocks are rarer, only standing out where the voltage trace shows outstanding periodicity.

²Bands spaced $2MW$ apart contain independent spectral statistics. Refer to [6] for details.

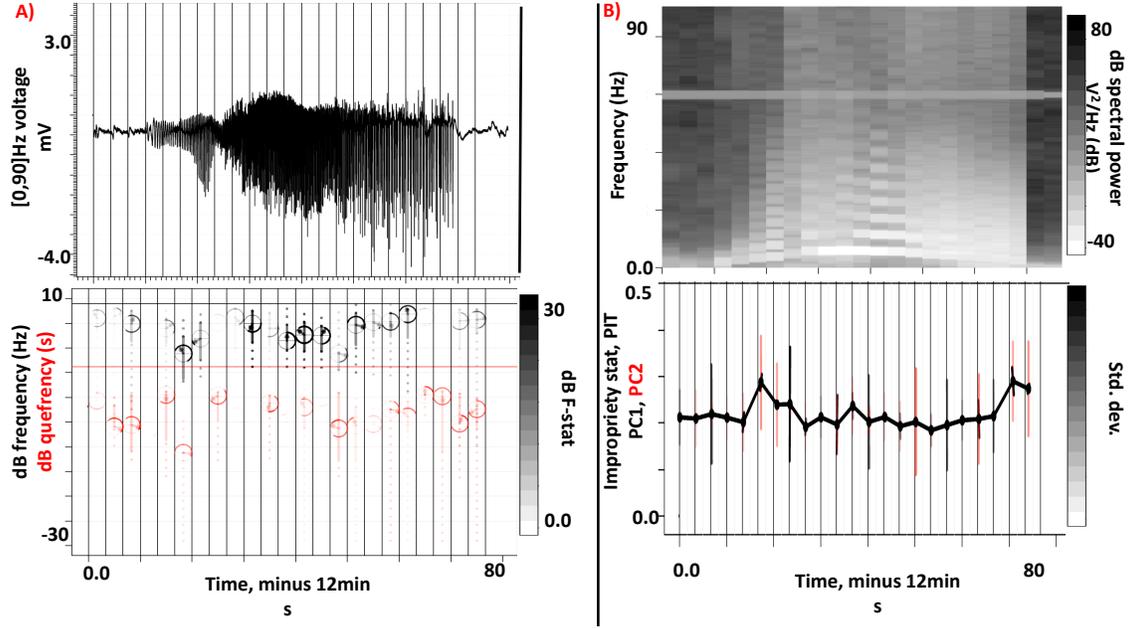


Figure 1: A low-dimensional representation of the X SDF explains fixed effects over the temporal strata. A) Top - reconstruction of the $[0, 90]$ Hz X demodulate (lowpass filters: early/late seizure - Slepian, same as used in [9]; central interval - Chebyshev Type I); bottom - spectrograms of μ_m (black) and cepstral-coefficient (red) phasors (centre horizontal line - 0; central solid vertical line - fourth spread; dotted vertical lines - min/max; opacity - F-stat heat bar) and phase clocks (right horizontal line - 0 rad; arrow - median; line pair - quartiles). B) Top - spectrogram; bottom - spectrogram of $(F^{(SPH)}, F^{(IMP)})$ PRC-plots (solid disc - average $(F^{(SPH)}, F^{(IMP)})$ -coordinate; black and red disc-crossing lines - respectively, PC-axis of first and second PC's extended to 1-sigma of mean coordinate).

For each temporal stratum, the lower B) plot displays the diagnostic PC-map (the $F^{(SPH)}$ plotting range is contained inside the tight temporal-stratum window, while the $F^{(IMP)}$ range extends the entire y-axis). Most of the time, black lines (PC1) stand out far more than red lines (PC2), indicating linear correlation in the $(F^{(SPH)}, F^{(IMP)})$ scatter. Together with all mean coordinates lying leftward and downward of the 50% percentiles: this correlation suggests that the eigencoefficient vectors tend to exhibit spherical, proper behaviour. This suggests high test power that consolidates the spectrogram results in A).

Conclude by illustrating how those efficiency gains of the A) spectrogram over the B) spectrogram hold across patient strata - motivating its widespread use in the physical sciences as a diagnostic tool when it is inefficient examine every spectrogram of the B) form (e.g., many patients as opposed to just two - and for preliminary dynamical analyses). Figure 2 displays the same content as Figure 1, except now for the Patient MG49 stratum. Again, red phasor clocks remain conspicuous throughout the record; therefore, the echo mean is robust to perturbations in noise structure. Consistent with the improved quality of harmonic periodicity over that for Patient C5 that one sees just looking at the voltage trace: more prominent black clocks stand out in the latter record stages. The Gaussian-diagnostic spectrogram shows similar distributional behaviour as seen for Patient C5.

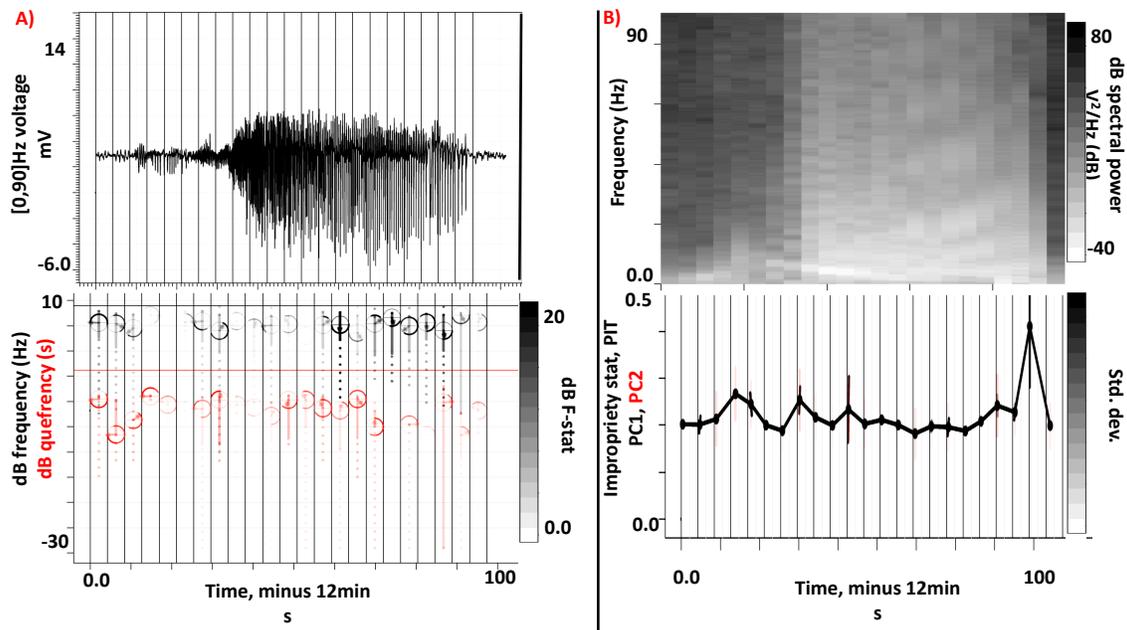


Figure 2: Efficiency of the diagnostic boxplots is reproduced in the Patient MG49 stratum. The same as in Figure 1, except now for the Patient MG49 stratum.

References

- [1] P. Billingsley. *Probability and measure*. John Wiley & Sons, 2008.
- [2] P. J. Brockwell and R. A. Davis. *Time series: theory and methods, Second Edition*. Springer Science & Business Media, 1991.
- [3] V. Krishnan. *Nonlinear filtering and smoothing: An introduction to martingales, stochastic integrals and estimation*. Wiley Interscience, 1984.
- [4] K. Q. Lepage and D. J. Thomson. Reduced mean-square error quadratic inverse spectrum estimator. *IEEE Transactions on Signal Processing*, 62(11):2958–2972, 2014.
- [5] M. Loève. *Probability theory*. D. Van Nostrand Co., Inc., Princeton, NJ-Toronto-New York-London, 1963.
- [6] F. A. Marshall. *Advances in the detection and characterization of nonstationary processes: An application to riometers*. PhD thesis, Queen’s University, Kingston, Ontario, 2020.
- [7] F. A. Marshall. Code Algorithms, 2023. GitHub. Last updated 2023. <https://github.com/fmarshall1/Multitaper-Harmonic-Analysis>.
- [8] F. A. Marshall. Temporal-lobe epilepsy: Harmonic and anharmonic periodicity in microelectrode voltage, 2023.

- [9] F. A. Marshall, D. J. Thomson, G. Takahara, R. A. Fiori, and D. W. Danskin. A characterization of periodicity in the voltage time series of a riometer. *Journal of Geophysical Research: Space Physics*, 2020.
- [10] A. Napolitano. *Generalizations of cyclostationary signal processing: spectral analysis and applications*. John Wiley & Sons, 2012.
- [11] W. Rudin. *Real and complex analysis, Third Edition*. WCB McGraw-Hill, 1987.
- [12] C. A. Schevon, S. Tobochnik, T. Eissa, E. Merricks, B. Gill, R. R. Parrish, L. M. Bateman, G. M. McKhann Jr, R. G. Emerson, and A. J. Trevelyan. Multiscale recordings reveal the dynamic spatial structure of human seizures. *Neurobiology of disease*, 127:303–311, 2019.
- [13] C. A. Schevon, S. A. Weiss, G. McKhann, R. R. Goodman, R. Yuste, R. G. Emerson, and A. J. Trevelyan. Evidence of an inhibitory restraint of seizure activity in humans. *Nature communications*, 3(1):1–11, 2012.
- [14] E. D. Schlaflly and M. A. Kramer. Human seizure MEA, 2022. Boston College, Accessed: 2022, Last updated 2022-05-25.
- [15] E. D. Schlaflly, F. A. Marshall, E. M. Merricks, U. T. Eden, S. S. Cash, C. A. Schevon, and M. A. Kramer. Multiple sources of fast traveling waves during human seizures: Resolving a controversy. *Journal of Neuroscience*, 42(36):6966–6982, 2022.
- [16] P. J. Schreier and L. L. Scharf. *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*. Cambridge University Press, 2010.
- [17] D. Slepian. On bandwidth. *Proceedings of the IEEE*, 64(3):292–300, 1976.
- [18] E. H. Smith, J.-Y. Liou, T. S Davis, E. M. Merricks, S. S. Kellis, S. A. Weiss, B. Greger, P. A. House, G. M. McKhann II, R. R. Goodman, et al. The ictal wavefront is the spatiotemporal source of discharges during spontaneous human seizures. *Nature communications*, 7(1):1–12, 2016.
- [19] D. J. Thomson. Quadratic-inverse spectrum estimates: Applications to palaeoclimatology. *Phil. Trans. R. Soc. Lond. A*, 332(1627):539–597, 1990.
- [20] D. J. Thomson. Multitaper analysis of nonstationary and nonlinear time series data. *Nonlinear and nonstationary signal processing*, pages 317–394, 2000.
- [21] A. T. Walden and P. Rubin-Delanchy. On testing for impropriety of complex-valued Gaussian vectors. *IEEE Transactions on Signal Processing*, 57(3):825–834, 2008.