Spectral Analysis Using Multitaper Whittle Methods with a Lasso Penalty

Peter F. Craigmile

The Ohio State University department of STATISTICS

http://www.stat.osu.edu/~pfc/

Multitaper Spectral Analysis, Banff International Research Station online workshop, Banff, Canada, Jun 24-26, 2022.

Joint research with Shuhan Tang and Yunzhang Zhu,

The Ohio State University.

Spectral analysis

Spectral analysis provides key insights into the **frequency domain** characteristics of a time series [e.g., Priestley, 1981, Percival and Walden, 1993].

Analyzing the **spectral density function (SDF)**:

- allows us to explore **periodicities** in the data;
- provides an alternative way to analyze and estimate the **covariance structure** of stationary time series;
- can be used to understand the **effect of preprocessing** a time series

Spectral density estimation

- **Nonparametric** estimators provide an adequate tradeoff between bias and variance, but often such estimates are still **too noisy** when a stable SDF estimate is required.

  e.g., periodogram, direct spectral estimators, lag window and overlapping segment averaging spectral estimators, and multitaper (MT) spectral estimators.

- Using a **parametric** approach, model misspecification induced by considering a limited class of models for the SDF, can compromise estimation.

- We use a **semiparametric** model for the SDF, in which the log SDF is expressed in terms of a truncated basis expansion, where the number of basis functions are allowed to increase with the sample size.

The statistical problem

How to enforce **sparsity** by selecting the basis functions and estimating the model parameters to adequately estimate the SDF, but also have **computational efficiency** as the sample size increases.

- Gao [1993], Gao [1997], Moulin [1994] and Walden et al. [1998] enforce sparsity using a penalized least squaress (LS) for the log SDF with wavelet soft thresholding.

  Computational complexity: $O(N)$, for a time series of N regularly sampled values.

- A number of approaches enforce smoothness of the SDF via an L_2 penalty: Cogburn and Davis [1974], Wahba and Wold [1975] and Wahba [1980] use penalized LS, and Pawitan and O'Sullivan [1994] uses a penalized Whittle method.

  To enforce sparsity, some L_2 methods use model selection, often in combination with cross-validation, to select the basis functions.

Our approach: Tang et al. [2019]

- Use a quasi-likelihood method for estimating SDFs with a Whittle likelihood [Whittle, 1953] based on multitaper (MT) spectral estimates while enforcing sparsity.
  - MT estimates provide good bias–variance tradeoff and can yield more efficient estimates of the SDF [Thomson, 1982, Percival and Walden, 1993, Walden et al., 1998].
  - The Whittle likelihood method improves estimation over traditional LS approaches.

Spectral representation of a time series

- For univariate stationary time series $\{X_t : t \in \mathbb{Z}\}$ collected at sampling interval $\Delta = 1$ let

$$\gamma(\tau) = \mathrm{cov}(X_t, X_{t+\tau}), \quad \tau \in \mathbb{Z}$$

be the **autocovariance function** (ACVF).

- For an absolutely summable ACVF, the **spectral density function (SDF)** is

$$S(f) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-i2\pi f\tau}, \text{ for frequency } |f| \leq 1/2,$$

with

$$\gamma_X(\tau) = \int_{-1/2}^{1/2} e^{i2\pi f\tau} S(f) df, \quad \text{for all integers } \tau,$$

Thus, $\{\gamma_X(\tau)\}$ and $\{S(f)\}$ are Fourier transform pairs.

Multitaper spectral estimation

- Many estimates of the SDF can be written as a **multitaper (MT) spectral estimate**, an average of a number of tapered spectral estimates [Walden, 2000].
- Suppose we observe N observations, $\boldsymbol{X} = (X_1, \ldots, X_N)^T$, of process $\{X_t\}$.

  Let $\{h_{k,t} : k = 1, \ldots, K,\ t = 1, \ldots, N\}$ denote K orthonormal data tapers.
- The standard MT spectral estimator of the SDF is

$$\widehat{S}^{(\mathrm{mt})}(f) \;=\; \frac{1}{K} \sum_{k=1}^{K} \widehat{S}_k^{(\mathrm{mt})}(f), \tag{1}$$

  where the kth ($k = 1, \ldots, K$) tapered spectral estimator (eigenspectrum) is

$$\widehat{S}_k^{(\mathrm{mt})}(f) = \left| \sum_{t=1}^{N} h_{k,t} X_t \exp(-i2\pi f t) \right|^2 .$$

Basis models for SDFs

- With basis functions $\boldsymbol{\phi}(f) = (\phi_1(f), \ldots, \phi_p(f))^T$ and letting $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, for each frequency f let

$$\log S(f) = \sum_{l=1}^{p} \phi_l(f)\beta_l = \boldsymbol{\phi}^T(f)\boldsymbol{\beta}, \tag{2}$$

  where number of basis functions p is allowed to increase with the sample size.

- Many choices:

  Orthogonal polynomial bases, Fourier bases, B-spline bases, wavelet bases, as well as some mixed dictionary bases.

  - A wavelet basis based on the discrete wavelet transform with the Daubechies least asymmetric wavelet filter of width 8 had good performance across our simulations.

MT-Whittle likelihood

Using the observations $\{X_1, \ldots, X_N\}$, we evaluate the MT spectral estimates $\widehat{S}^{(\mathrm{mt})}(f_j)$ on the set of $M = \lceil N/2 \rceil - 1$ non-zero, non-Nyquist (i.e., not equal to 1/2) (NZNN) **Fourier frequencies** defined by

$$\left\{ f_j = \frac{j}{N} : j = 1, \ldots, M \right\}.$$

Definition 1: Using MT spectral estimators, a quasi-likelihood function with expression

$$l_W(S(f)) = \sum_{j=1}^{M} \left\{ \log S(f_j) + \frac{\widehat{S}^{(\mathrm{mt})}(f_j)}{S(f_j)} \right\} \tag{3}$$

is called the **MT-Whittle likelihood** function.

MT-Whittle likelihood: asymptotics

Proposition 1 [Walden, 2000, Section 3.3]

Suppose that $\{X_t : t \in \mathbb{Z}\}$ is strictly stationary with all moments existing such that

$$\sum_{\tau_1,\ldots,\tau_{l-1}=-\infty}^{\infty} |\mathrm{cum}(X_{t+\tau_1}, \ldots, X_{t+\tau_{l-1}}, X_t)| < \infty,$$

for $l = 2, 3, \ldots$, where $\mathrm{cum}(X_{t_1}, \ldots, X_{t_l})$ denotes the joint cumulant function of order l (see, e.g., Brillinger [1981], sec. 2.3). Also for each N, let $\{h_{k,t} : k = 1, \ldots, K,\ t = 1, \ldots, N\}$ be a set of K orthonormal sine or DPSS data tapers. Then

$$\widehat{S}^{(\mathrm{mt})}(f) \to_d S(f)\frac{\chi^2_{2K}}{2K}, \ \text{ for } 0 < f < 1/2, \ \text{as } N \to \infty,$$

where χ^2_{2K} denotes a chisquared random variable (RV) with $2K$ degrees of freedom.

MT-Whittle likelihood: quasi-likelihood representation

In general MT-spectral estimators are correlated over frequencies [Thomson, 1982].

For a locally slowly varying spectrum, for $0 < f < f' < 1/2$, with f close to f',

$$\mathrm{Cov}\{\widehat{S}^{(\mathrm{mt})}(f), \widehat{S}^{(\mathrm{mt})}(f')\} \approx \frac{S^2(f)}{K^2} \sum_{k=1}^{K} \sum_{l=1}^{K} \left| \sum_{t=1}^{N} h_{t,k} h_{t,l} e^{i2\pi(f'-f)t} \right|^2 .$$

However, the next result shows that our MT-Whittle likelihood (3) can be reinterpreted as a gamma quasi-likelihood, which ignores these correlations between frequencies.

Proposition 2: The MT-Whittle likelihood (3) corresponds to a gamma quasi-likelihood assuming the asymptotic distribution of Proposition 1 at the NZNN Fourier frequencies, and assuming independence between the Fourier frequencies.

L_1 **Penalized MT-Whittle Method**

- Incorporating a Lasso-type penalty with the MT-Whittle likelihood (3), our optimization problem is then

$$\min_{\boldsymbol{\beta}} \; l_W(\boldsymbol{\Phi}\boldsymbol{\beta}) + \lambda \sum_{l=1}^{p} |\beta_l|, \tag{4}$$

where $\lambda \geq 0$ denotes the **tuning parameter** of the L_1 penalty.

- Specifically, by introducing the equality constraints $\boldsymbol{\zeta} = \boldsymbol{\Phi}\boldsymbol{\beta}$ and $\boldsymbol{\eta} = \boldsymbol{\beta}$, the original problem (4) is equivalent to

$$\min_{\boldsymbol{\zeta},\boldsymbol{\eta},\boldsymbol{\beta}} \quad l_W(\boldsymbol{\zeta}) + \lambda \sum_{l=1}^{p} |\eta_l|$$
$$\text{subject to } \boldsymbol{\Phi}\boldsymbol{\beta} = \boldsymbol{\zeta} \text{ and } \boldsymbol{\beta} = \boldsymbol{\eta}\,,$$

where $\boldsymbol{\zeta} \in \mathbb{R}^M$ and $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_p)^T \in \mathbb{R}^p$.

An alternating direction method of multipliers (ADMM) algorithm

Advantages:

- Favors distributed computing;
- Per-iteration cost is often much lower than that of the interior point algorithm in literature for optimizing penalized non-Gaussian likelihood;
- An attractive choice when solutions of medium accuracy are sufficient, such as parameter estimation problems.

Overall computational complexity for the ADMM algorithm to reach an ϵ-optimal solution is:

- $O(\epsilon^{-1}M^2 + M^3)$ for a general basis;
- $O(\epsilon^{-1}M)$ for orthogonal basis functions, such as wavelets.

(Remember M is the number of NZNN Fourier frequencies.)

An ADMM algorithm

Step 0. Initialize $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\zeta}^{(0)}$, $\boldsymbol{\eta}^{(0)}$, $\boldsymbol{u}_1^{(0)}$ and $\boldsymbol{u}_2^{(0)}$;

Step 1. Alternately update the **primal variables** $(\boldsymbol{\zeta}, \boldsymbol{\eta}, \boldsymbol{\beta})$ and the associated **dual variables** $(\boldsymbol{u}_1, \boldsymbol{u}_2)$. The $(n+1)$-th set of updates are:

$$
\begin{aligned}
\boldsymbol{\beta}^{(n+1)} &= (\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \boldsymbol{I}_p)^{-1}\left\{\boldsymbol{\Phi}^T(\boldsymbol{\zeta}^{(n)} - \boldsymbol{u}_1^{(n)}) + \boldsymbol{\eta}^{(n)} - \boldsymbol{u}_2^{(n)}\right\};\\
\zeta_j^{(n+1)} &= \arg\min_{\zeta_j}\{\zeta_j + \widehat{S}^{(\mathrm{mt})}(f_j)\exp(-\zeta_j) + \frac{\rho}{2}\{\boldsymbol{\phi}^T(f_j)\boldsymbol{\beta}^{(n+1)} - \zeta_j + u_{1j}^{(n)}\}^2\},\ j = 1, \ldots, M;\\
\eta_l^{(n+1)} &= \mathrm{ST}\left(\beta_l^{(n+1)} + u_{2l}^{(n)}, \frac{\lambda_l}{\rho}\right),\ l = 1, \ldots, p;\\
\boldsymbol{u}_1^{(n+1)} &= \boldsymbol{u}_1^{(n)} + \boldsymbol{\Phi}\boldsymbol{\beta}^{(n+1)} - \boldsymbol{\zeta}^{(n+1)};\\
\boldsymbol{u}_2^{(n+1)} &= \boldsymbol{u}_2^{(n)} + \boldsymbol{\beta}^{(n+1)} - \boldsymbol{\eta}^{(n+1)},
\end{aligned}
$$

Step 2. Iterate Step 1. until when both the primal and dual residuals are smaller than prespecified precisions following the criterion in Boyd et al. [2011].

$\rho > 0$ is a **penalty parameter**; $\mathrm{ST}(x, a) = \mathrm{Sign}(x)\max(|x| - a, 0)$ is **soft-thresholding function** with threshold $a \geq 0$.

Tuning parameter selection

1. Scale-calibrated universal threshold, adapted from Donoho and Johnstone [1994],

$$\lambda^{univ} = \sqrt{1/K}\sqrt{2\log p}.$$

(1/K follows from asymptotic quasi-likelihood theory for the MT-Whittle estimator of $\boldsymbol{\beta}$.)

2. Generalized information criterion [Fan and Tang, 2013],

$$\lambda^{GIC} = \arg\min_{\lambda} \left\{2K\ l_W(\boldsymbol{\Phi}\boldsymbol{\beta}_{\lambda}) + c_M|p_{\lambda}|\right\}, \tag{5}$$

where $\boldsymbol{\beta}_{\lambda}$ is the optimizer of L_1 penalized MT-Whittle likelihood with tuning parameter λ, $|p_{\lambda}|$ denotes the number of non-zero elements in $\boldsymbol{\beta}_{\lambda}$, and c_M is the penalty parameter.

When the number of predictors, p, increases exponentially as the sample size M increases Fan and Tang [2013] suggest $c_M = \log\log M\ \log p$.

Theory – assumptions

Assumption A1: (Sparsity condition) Assume that

$$\log S(f_j) = \langle \phi(f_j)\,, \boldsymbol{\beta}^0 \rangle, \quad j = 1, \ldots, M,$$

for some sparse vector $\boldsymbol{\beta}^0 \in \mathbb{R}^p$ and basis functions $\phi(f_j)$ satisfying $\|\phi(f)\|_\infty \leq B$ for $f \in [-1/2, 1/2]$ and some constant B.

Assumption A2: (Compatibility condition) Let $S = \{l : \boldsymbol{\beta}_l^0 \neq 0\}$ and $s_0 = |S|$. Assume that for any $v \in \mathbb{R}^p$ with $\|v_{S^c}\|_1 \leq 3\|v_S\|_1$ that

$$\frac{1}{M} \sum_{j=1}^{M} R_j \left\{ \exp(v^\top \phi(f_j)) - v^\top \phi(f_j) - 1 \right\} \; \geq \; \min \left\{ \frac{c_0}{s_0} \|v_S\|_1^2 \,, c_1 M^{\gamma - \frac{1}{2}} \|v_S\|_1 \right\},$$

with probability tending to 1 as $M \to \infty$ for some constants $c_0 > 0$, $c_1 > 0$, $0 < \gamma \leq \frac{1}{2}$ and where we define $R_j \equiv \widehat{S}^{(\mathrm{mt})}(f_j)/S(f_j)$ for $j = 1, \ldots, M$.

Theorem

Under Assumptions (A1) and (A2) and on the event

$$\left\| \sum_{j=1}^{M} (R_j - 1)\, \phi(f_j) \right\|_\infty < \frac{c_1}{3} M^{\gamma + \frac{1}{2}}, \tag{6}$$

we have that $\left\| \widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}^0 \right\|_1 \leq \frac{3\lambda s_0}{2c_0 M}$ for any λ satisfying $2 \left\| \sum_{j=1}^{M} (R_j - 1)\, \phi(f_j) \right\|_\infty \leq \lambda < \frac{2}{3} c_1 M^{\gamma + \frac{1}{2}}$. In particular, when

$$\lambda = 2 \left\| \sum_{j=1}^{M} (R_j - 1)\, \phi(f_j) \right\|_\infty ,$$

we have, on the event (6), that

$$\left\| \widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}^0 \right\|_1 \leq \frac{3 s_0}{c_0 M} \left\| \sum_{j=1}^{M} (R_j - 1)\, \phi(f_j) \right\|_\infty ,$$

and

$$\sup_{f \in [-\frac{1}{2}, \frac{1}{2}]} |\log \widehat{S}(f) - \log S(f)| \leq \frac{3 B s_0}{c_0 M} \left\| \sum_{j=1}^{M} (R_j - 1)\, \phi(f_j) \right\|_\infty ,$$

where $\log \widehat{S}(f)$ is the L_1 penalized MT-Whittle estimator of the log SDF.

Implications

- By Proposition 1, we have that $R_j - 1$ behaves like $\chi^2_{2K}/2K - 1$ RVs asymptotically.

  We conjecture this leads to the following rate of convergence for $\widehat{\boldsymbol{\beta}}$ and the log SDF:

$$\begin{aligned} \left\| \widehat{\boldsymbol{\beta}}(\lambda) - \boldsymbol{\beta}^0 \right\|_1 &= O_p\left(s_0 \frac{\log(M)}{\sqrt{M}} \right), \\ \sup_{f \in [-\frac{1}{2}, \frac{1}{2}]} |\log \widehat{S}(f) - \log S(f)| &= O_p\left(s_0 \frac{\log(M)}{\sqrt{M}} \right). \end{aligned}$$

- If we further assume that s_0 is quite small, that is, the true log SDF has a sparse basis representation, a parametric rate $M^{-1/2}$ can be achieved (up to a log factor).

- This is in contrast to the slower nonparametric rate for typical one-dimensional nonparametric regression or density estimation problems (see, e.g., Tsybakov [2009]).

- Our theory suggests that by exploring sparsity, if it is indeed present in the signal, a significant improvement in estimation efficiency can be achieved using our proposed method.

Simulation study

We use the following processes:

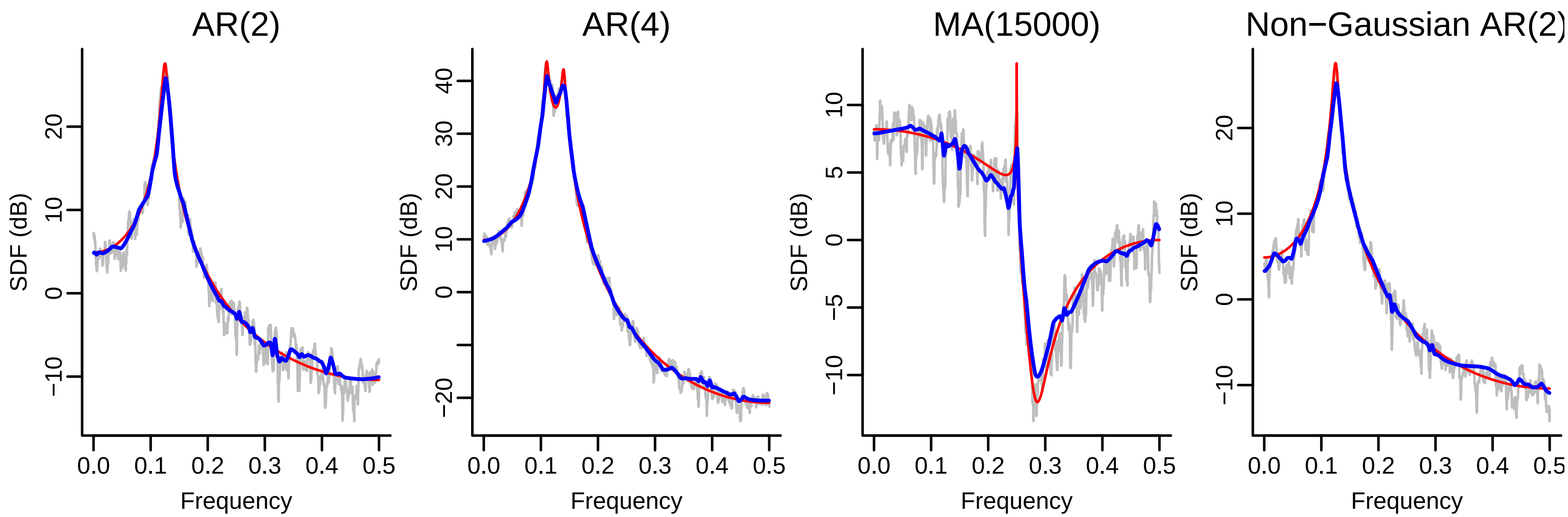
1. AR(2) process: $X_t = \varphi_{1,1} X_{t-1} + \varphi_{1,2} X_{t-2} + \varepsilon_t$ with $\varphi_{1,1} = 0.97\sqrt{2}$, $\varphi_{1,2} = -0.97^2$;

2. AR(4) process: $X_t = \varphi_{2,1} X_{t-1} + \varphi_{2,2} X_{t-2} + \varphi_{2,3} X_{t-3} + \varphi_{2,4} X_{t-4} + \varepsilon_t$ with $\varphi_{2,1} = 2.7607$, $\varphi_{2,2} = -3.8106$, $\varphi_{2,3} = 2.6535$, $\varphi_{2,4} = -0.9238$;

3. High-order MA process: $X_t = \sum_{l=0}^{15000} \theta_l \varepsilon_{t-l}$ with $\theta_0 = 1$, $\theta_1 = \pi/4$, and $\theta_l = \sin(\pi(l-1)/2)/(l-1)$ for $l = 2, 3, \ldots, 15000$.

We demonstrate how our estimation method performs when the innovations $\{\varepsilon_t\}$ are N(0, 1) RVs, but also present the same AR(2) process case 1 with innovations generated by a shifted Exponential distribution with mean 0 and variance 1.

A comparison of spectral estimates for different four processes

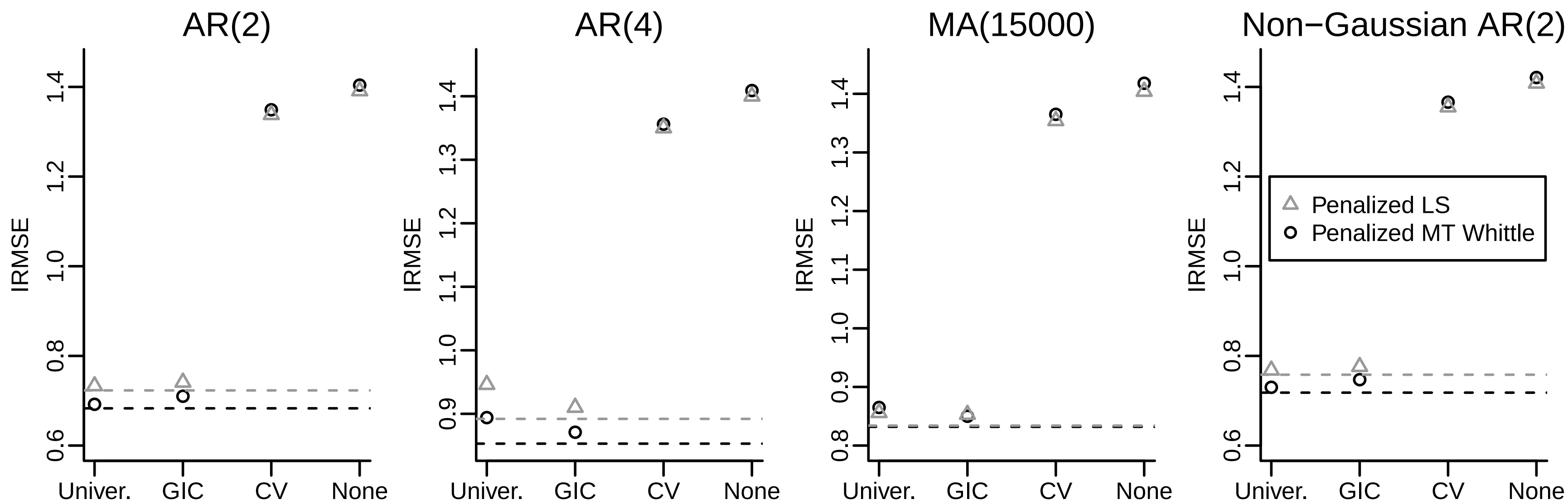


Red: true SDF.

Blue: L_1 penalized MT-Whittle estimates with LA(8) wavelet basis and universal threshold.

Gray: corresponding raw MT estimates with $K = 10$ tapers.

Decibel-scale integrated root mean squared error (IRMSE) comparisons



Comparison of IRMSEs for different L_1 penalizations methods for estimating SDF

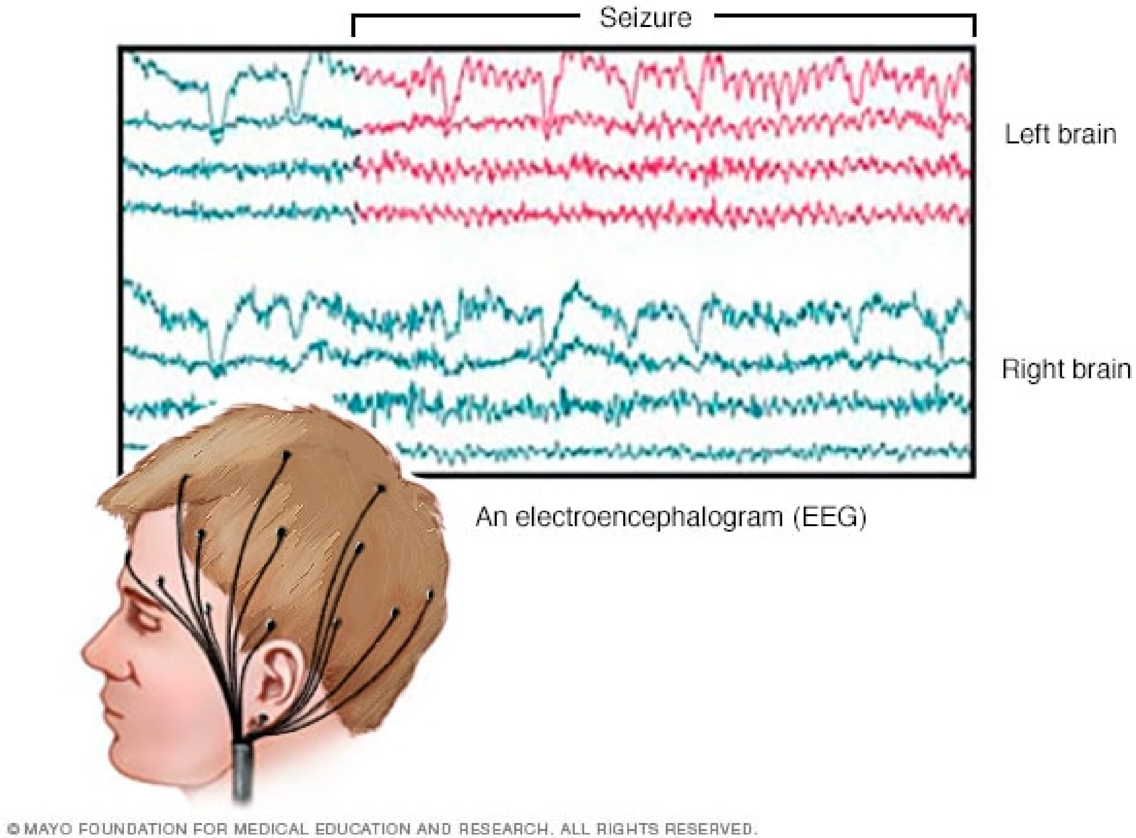
(Symbol heights are larger than 95% bootstrap-based confidence intervals.)

IRMSEs are averaged over 1,000 realizations.

Horizontal dashed lines: the best possible IRMSEs for each method.

Biomedical application: Electroencephalography (EEG)

Electroencephalogram (EEG) signals are often used to monitor brain activity and diagnose disease such as epileptic seizures.

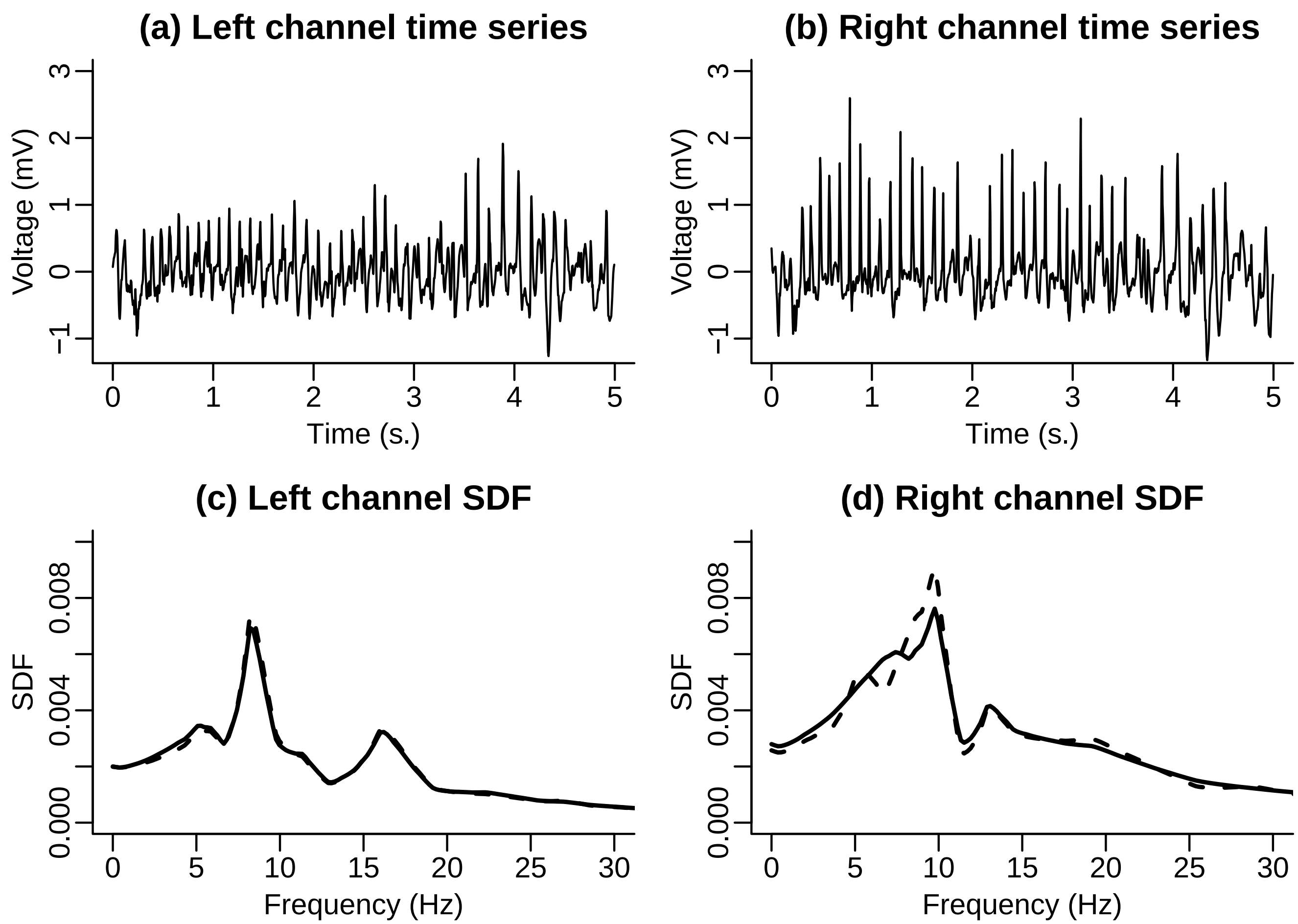


(Source: https://www.mayoclinic.org/tests-procedures/eeg/about/pac-20393875)

Application: EEG signals

- We analyze two channels of EEG data collected from the left and right front cortex of one male rat. Quiroga et al. [2002] argue that, genetically, analyzing these series is relevant to the study of human epilepsy.
- The data is presented in van Luijtelaar [1997], and was originally downloaded from `http://www.vis.caltech.edu/~rodri` (no longer available).
- Each channel contains 1,000 voltages recorded in units of microvolts (mV) collected at a sampling rate of 200 Hz.

Application: EEG signals, continued



Plots of EEG signals and estimated SDFs. In (c) and (d), **solid line:** estimates with the universal threshold; **dashed line:** estimates with GIC-based threshold.

Conclusions

- Enforcing sparsity, our L_1 penalized MT-Whittle estimator performs better or as good as previous methods for estimating the SDF.
- Extends to the broader classes of basis functions and their mixtures, beyond those traditionally used with wavelet thresholding.
- Simulations demonstrate a clear advantage of using the GIC and universal threshold over cross-validation for tuning parameter selection.
- Computationally, universal threshold is data-invariant (it only relies on the number of tapers K) whereas the calculation of GIC is data-dependent.
- ADMM algorithm can be accelerated when parallel computing and orthogonal bases used.

Extensions for univariate spectral analysis

- Sandwich-based interval estimation based on **sample splitting** and **de-sparsifying the lasso** to improve statistical inference [e.g. Meinshausen et al., 2009, Dezeure et al., 2015, Faraway, 2016]
- Varying the initial number of basis, p, may further improve the estimation. Related approaches include the truncated lasso by Shen et al. [2012].
- A dictionary of different types of basis functions, and how to automatically choose the best basis type [e.g. Wasserman, 2006].
- Time series prediction using estimated SDF [e.g., Brockwell and Davis, 1991].

Multivariate and beyond

- Cholesky-based approaches for estimating the spectral density matrix (SDM).
- Comparison to non-Cholesky-based approaches for estimating the SDM [e.g. Holan et al., 2017, Chau and von Sachs, 2020].
- Asymptotic theory of the L_1 penalized multivariate MT-Whittle estimators for SDM.
- An extension of the sandwich method to interval estimation of the SDM elements, and the construction of a joint confidence region.
- Spectral analysis of nonstationary time series
- Spatio-temporal processes.

References

S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3:1–122, 2011.

D. Brillinger. *Time Series: Data Analysis and Theory*. Holt, New York, NY, 1981.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods (Second Edition)*. Springer Science & Business Media, New York, NY, 1991.

J. Chau and R. von Sachs. Intrinsic wavelet regression for curves of Hermitian positive definite matrices. *Journal of the American Statistical Association*, 0:1–14, 2020.

R. Cogburn and H. T. Davis. Periodic splines and spectral estimation. *The Annals of Statistics*, 2:1108–1126, 1974.

R. Dezeure, P. Bühlmann, L. Meier, and N. Meinshausen. High-dimensional inference: Confidence intervals, p-values and R-software hdi. *Statistical Science*, 30:533–558, 2015.

D. L. Donoho and J. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

Y. Fan and C. Y. Tang. Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75: 531–552, 2013.

J. J. Faraway. Does data splitting improve prediction? *Statistics and computing*, 26:49–60, 2016.

H.-Y. Gao. *Wavelet Estimation of Spectral Densities in Time Series Analysis*. PhD thesis, Department of Statistics, University of California, Berkeley, 1993.

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

S. H. Holan, T. S. McElroy, and G. Wu. The cepstral model for multivariate time series: the vector exponential model. *Statistica Sinica*, 27:23–42, 2017.

N. Meinshausen, L. Meier, and P. Bühlmann. P-values for high-dimensional regression. *Journal of the American Statistical Association*, 104:1671–1681, 2009.

P. Moulin. Wavelet thresholding techniques for power spectrum estimation. *IEEE Transactions on Signal Processing*, 42:3126–3136, 1994.

Y. Pawitan and F. O'Sullivan. Nonparametric spectral density estimation using penalized Whittle likelihood. *Journal of the American Statistical Association*, 89:600–610, 1994.

D. B. Percival and A. T. Walden. *Spectral Analysis for Physical Applications*. Cambridge University Press, Cambridge, England, 1993.

M. B. Priestley. *Spectral Analysis and Time Series*. Academic Press, New York, NY, 1981.

R. Q. Quiroga, T. Kreuz, and P. Grassberger. Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Physical Review E*, 66:1–9, 2002.

X. Shen, W. Pan, and Y. Zhu. Likelihood-based selection and sharp parameter estimation. *Journal of the American Statistical Association*, 107:223–232, 2012.

S. Tang, P. F. Craigmile, and Y. Zhu. Spectral estimation using multitaper Whittle methods with a lasso penalty. *IEEE Transactions on Signal Processing*, 67:4992–5003, 2019.

D. J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70:1055–1096, 1982.

A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, NY, 2009.

G. van Luijtelaar. *The WAG/Rij Rat Model of Absence Epilepsy: Ten Years of Research: a Compilation of Papers*. Nijmegen University Press, Nijmegen, Netherlands, 1997.

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

G. Wahba and S. Wold. Periodic splines for spectral density estimation: The use of cross validation for determining the degree of smoothing. *Communications in Statistics*, 4: 125–141, 1975.

A. T. Walden. A unified view of multitaper multivariate spectral estimation. *Biometrika*, 87:767–788, 2000.

A. T. Walden, D. B. Percival, and E. J. McCoy. Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Transactions on Signal Processing*, 46:3153–3165, 1998.

L. A. Wasserman. *All of Nonparametric Statistics.* Springer, New York, NY, 2006.

P. Whittle. Estimation and information in stationary time series. *Arkiv för Matematik*, 2:423–434, 1953.