Supplement to "Detecting and Localizing Differences in Functional Time Series Dynamics: A Case Study in Molecular Biophysics"[*]

Shahin Tavakoli[†] & Victor M. Panaretos[‡]

December 16, 2015

A Molecular Dynamics Simulation Protocol

The details of the atomistic simulation protocol have been described previously (Lankas et al. 2006, Gonzalez et al. 2013). In particular, initial DNA minicircle structures were created with `Jumna` (Lavery et al. 1995) of sequences studied using cryo-electron microscopy (Amzallag et al. 2006)—see also Table 1 in the main paper. The `AMBER 10` suite of programs (Case et al. 2008), with the `parmbsc0` force field (Pérez et al. 2007), were used to run the simulations in explicit water modeled with the SPC/E parameters, with periodic boundary conditions. The ion parameters developed by Joung & Cheatham III (2008) were used to add 150 mM (Mole/litre) of KCl. The structures were first equilibrated and then the production run was carried out with `PMEMD`. This treats long-range electrostatics interactions with the particle mesh Ewald summation method (Essmann et al. 1995) with a real space cutoff of 9 angstrom, cubic B-spline interpolation onto the charge grid with a spacing of ~ 1 angstrom. The Berendsen algorithm was used to maintain a constant temperature of 300 Kelvin and a constant pressure of 1 atm (standard atmosphere). A 50 nanosecond trajectory of each minicircle was generated using timesteps of 2 femtoseconds, with snapshots saved at 1 picosecond intervals. The DNA conformation at each snapshot was analysed with the program

[*]Research Supported by an ERC Starting Grant Award

[†]*Statistical Laboratory, University of Cambridge*, Email: `s.tavakoli@statslab.cam.ac.uk`. ST was partially supported by the EPSRC grant EP/K021672/2, and the research was partly carried out while ST was a Ph.D. student at the Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne.

[‡]*Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne*, Email: `victor.panaretos@epfl.ch`

Curves+ to obtain the base-pair coordinates (Lavery et al. 2009).

B Technical Assumptions

Our technical assumptions concerning the smoothness of the curves $\tau \mapsto \{X_t(\tau)\}$ and the decay of dependence between the elements of the sequence $\{X_t\}$ follow those in Panaretos & Tavakoli (2013*b*) and Tavakoli (2014), which extend to the functional setting the classical cumulant mixing conditions of Brillinger (2001). The following conditions, given in Panaretos & Tavakoli (2013*a*) and Tavakoli (2014), are used in the paper:

Conditions B.1. *X_t is a stationary times series in $L^2([0,1],\mathbb{R})$, satisfying:*

(1) $\mathbb{E}\|X_0\|^k < \infty$ *for all* $k \geq 1$

(2) $\sum_{t_1,\ldots,t_{k-1}=-\infty}^{\infty} \left\|\mathrm{cum}\left(X_{t_1},\ldots,X_{t_{k-1}},X_0\right)\right\| < \infty, \quad$ *for all* $k \geq 2$.

(3) $\sum_{t_1,\ldots,t_{k-1}=-\infty}^{\infty}(1+|t_j|)\left\|\mathrm{cum}\left(X_{t_1},\ldots,X_{t_{k-1}},X_0\right)\right\| < \infty$, *for* $k \in \{2,4\}$ *and* $j < k$.

(4) $\sum_{t\in\mathbb{Z}}(1+|t|)|||\mathscr{R}_t|||_1 < \infty$.

(5) $\sum_{t_1,t_2,t_3\in\mathbb{Z}} |||\mathscr{R}_{t_1,t_2,t_3}|||_1 < \infty$.

The definitions of the different objects involved are as follows:

$$r_t(\tau,\sigma) = \mathbb{E}\left[(X_t(\tau)-\mu(\tau))(X_0(\sigma)-\mu(\sigma))\right]$$

is the lag-t autocovariance kernel, $\mathrm{cum}\left(X_{t_1},\ldots,X_{t_k}\right)$ is a *k-th order cumulant kernel* (see Panaretos & Tavakoli 2013*b*). The cumulant kernel of order 4 gives rise to a corresponding *4-th order cumulant operator* $\mathscr{R}_{t_1,t_2,t_3} : L^2([0,1]^2,\mathbb{R}) \to L^2([0,1]^2,\mathbb{R})$, defined by

$$\mathscr{R}_{t_1,t_2,t_3}(u\otimes v) = \mathrm{cum}\left(X_{t_1}\otimes X_{t_2}\langle u, X_{t_3}\rangle\langle v, X_0\rangle\right), \quad u,v \in L^2\left([0,1],\mathbb{R}\right).$$

$|||\cdot|||_1$ denotes the trace norm (also known as nuclear norm). For a detailed discussion of the interpretation and role of these conditions, and a comparative discussion in relation with the finite-dimensional versions thereof (as given in Brillinger 2001), the reader is referred to Panaretos &

Tavakoli (2013*b*). Notice that moment-based conditions such as cumulant conditions are quite natural in our setting since we are considering the motion of a finite physical object, thus guaranteeing finiteness of all moments.

C Proof of Theorem 3.1

The proof is in two parts. First we will show the result for a modified version of the test, where we assume that the eigenfunctions and eigenvalues of the spectral density operators are known. Then we will show that the sample versions of the eigenfunctions and eigenvalues are consistent, and the conclusion will follow from Slutsky's Theorem for metric spaces.

We recall that under the hypothesis H_ω, we have $\mathscr{F}_\omega^1 = \mathscr{F}_\omega^2 =: \mathscr{F}_\omega$, and the latter operator has an eigendecomposition $\mathscr{F}_\omega = \sum_{j\geq 1} \mu_j^\omega \varphi_j^\omega \otimes \varphi_j^\omega$. Denote by $\mathscr{F}_\omega^{(T)} = (\mathscr{F}_\omega^{1,(T)} + \mathscr{F}_\omega^{2,(T)})/2$ the pooled sample spectral density operator, with eigendecomposition $\mathscr{F}_\omega^{(T)} = \sum_{j\geq 1} \tilde{\mu}_j^\omega \tilde{\varphi}_j^\omega \otimes \tilde{\varphi}_j^\omega$. Write $\varphi_{ij} = \varphi_i \otimes \varphi_j$ and $\tilde{\varphi}_{ij} = \tilde{\varphi}_i \otimes \tilde{\varphi}_j$. Fix $\omega \in [0,\pi]$, and omit any superscripts of the form $\cdot^\omega$ in order simplify notation (this will not affect the validity of the proof, since ω will be fixed throughout). Define

$$\check{\Delta}_K^{(T)}(\omega) = \sum_{i,j=1}^{K} \frac{\left|\left\langle D_\omega^{(T)} \varphi_j, \varphi_i \right\rangle\right|^2}{\left(1 + \mathbf{1}_{\{0,\pi\}}(\omega)\right) 4\pi\kappa^2 \mu_i \mu_j}, \tag{C.1}$$

where $D_\omega^{(T)} = \sqrt{TB_T}\left(\mathscr{F}_\omega^{1,(T)} - \mathscr{F}_\omega^{2,(T)}\right)$. By Panaretos & Tavakoli (2013*b*, Theorem 3.7), we know that $D_\omega^{(T)}$ converges in distribution to a random element $\check{\mathscr{F}}_\omega$, whose Karhunen–Loève expansion is given by $\check{\mathscr{F}}_\omega = \sum_{i,j=1}^{\infty} \eta_{ij}\varphi_{ij}$, see Lemmas 4.4 and 4.5 in Panaretos & Tavakoli (2013*a*) for details. In particular, $\{\eta_{ij} : 1 \leq i \leq j\}$ are all independent Gaussian random variables with mean zero. Hence, by the continuous mapping Theorem,

$$\check{\Delta}_K^{(T)}(\omega) \xrightarrow{d} \sum_{i,j=1}^{K} \frac{|\eta_{ij}|^2}{\left(1 + \mathbf{1}_{\{0,\pi\}}(\omega)\right) 4\pi\kappa^2 \mu_i \mu_j} =: \Delta_K(\omega) \tag{C.2}$$

We now need to distinguish two cases: if $\omega \in \{0,\pi\}$, then $\eta_{ji} = \eta_{ij}$ for $i < j$, all $\{\eta_{ij} : j \geq i \geq 1\}$ are *real* independent Gaussian random variables with mean zero, and $\operatorname{var}(\eta_{ii}) = 8\pi\kappa^2\mu_i^2$, $\operatorname{var}(\eta_{ij}) = 4\pi\kappa^2\mu_i\mu_j, i < j$. A direct calculation thus yields $\Delta_K(\omega) \sim \chi^2_{K(K+1)/2}$, since $\left(|\eta_{ij}|^2 + |\eta_{ji}|^2\right)/(8\pi\kappa^2\mu_i\mu_j) = |\eta_{ij}|^2/(4\pi\kappa^2\mu_i\mu_j) \sim \chi_1^2$, $i < j$ and $|\eta_{ii}|^2/(8\pi\kappa^2\mu_i^2) \sim \chi_1^2$. If $\omega \notin \{0,\pi\}$, the random variables η_{ij}

are real Gaussian variables for $i = j$ and circular complex Gaussian for $i < j$ (see e.g. Picinbono (1996), Schreier & Scharf (2010) for the definition of circular complex Gaussian random variables), with $\eta_{ji} = \overline{\eta_{ij}}$ for $i < j$ and $\text{var}(\eta_{ij}) = 4\pi\kappa^2\mu_i\mu_j$ for all $i \leq j$. Hence, $\Delta_K(\omega) \sim \chi^2_{K^2}$, since $|\eta_{ii}|^2/(4\pi\kappa^2\mu_i^2) \sim \chi^2_1$ and $\left(|\eta_{ij}|^2 + |\eta_{ji}|^2\right)/(4\pi\kappa^2\mu_i\mu_j) = 2|\eta_{ij}|^2/(4\pi\kappa^2\mu_i\mu_j) \sim \chi^2_2$, for $i < j$.

Turning to the second part of the proof, Lemma 4.2 of Bosq (2000), implies that $\mathbb{E}|\tilde{\mu}_i - \mu_i|^2 \leq \mathbb{E}\|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\|_2^2 \to 0, \quad T \to \infty$. Hence, for all $i = 1, 2, \ldots, K$, $\tilde{\mu}_i$ converges in L^2 to μ_i, and $\tilde{\mu}_i$ is therefore a consistent estimator of μ_i. We now turn to the eigenfunctions φ_i. We point out that these eigenfunctions are not uniquely defined, however the eigenprojectors $\Pi_i = \varphi_i \otimes \varphi_i$ are well defined. Using Panaretos & Tavakoli (2013*a*, Propositions 5.1 and 5.2), we get $|\langle D_\omega^{(T)}\tilde{\varphi}_j, \tilde{\varphi}_i\rangle|^2 = \langle D_\omega^{(T)} \mathbin{\widetilde{\otimes}} D_\omega^{(T)}, \tilde{\Pi}_i \otimes \tilde{\Pi}_j\rangle_{\mathcal{S}_2}$, where $A \mathbin{\widetilde{\otimes}} B$ denotes the Kronecker product of two operators A, B acting on a Hilbert space H, i.e. $(A \mathbin{\widetilde{\otimes}} B)C = ACB^\dagger$ for an operator C acting on H, $B^\dagger$ is the adjoint operator of B, and $\langle\cdot,\cdot\rangle_{\mathcal{S}_2}$ denotes the Hilbert-Schmidt scalar product, i.e. $\langle A, B\rangle_{\mathcal{S}_2} = \sum_{j\geq 1}\langle Ae_j, Be_j\rangle$ for operators A, B acting on $L^2([0,1],\mathbb{C})$, and $(e_j)_{j\geq 1}$ is any complete orthonormal sequence in $L^2([0,1],\mathbb{C})$. Since $\tilde{\Pi}_i$ is a consistent estimator of Π_i, $i = 1, \ldots, K$ (Panaretos & Tavakoli 2013*a*, Theorem 4.3), and the Kronecker and tensor products are continuous, the continuous mapping Theorem implies $\tilde{\Pi}_i \otimes \tilde{\Pi}_j \xrightarrow{p} \Pi_i \otimes \Pi_j$, and $D_\omega^{(T)} \mathbin{\widetilde{\otimes}} D_\omega^{(T)} \xrightarrow{d} \breve{\mathscr{F}}_\omega \mathbin{\widetilde{\otimes}} \breve{\mathscr{F}}_\omega$. Therefore, by Slutsky's Theorem,

$$\tilde{\Delta}_K^{(T)} \xrightarrow{d} \sum_{i,j=1}^{K} \frac{\left\langle \breve{\mathscr{F}}_\omega \mathbin{\widetilde{\otimes}} \breve{\mathscr{F}}_\omega, \Pi_i \otimes \Pi_j \right\rangle_{\mathcal{S}_2}}{\left(1 + \mathbf{1}_{\{0,\pi\}}(\omega)\right) 4\pi\kappa^2\mu_i\mu_j} = \Delta_K(\omega).$$

To complete the proof, it suffices to note that the independence of the $\Delta_{K_j}(\omega_j)$s follows directly from the independence of the $\breve{\mathscr{F}}_{\omega_j}$s.

D Numerical Simulations

D.1 Level and Power

In order to assess the finite sample performances of our testing procedure, we conducted some numerical simulations. The situation where the truncation level K is chosen using either AIC or AIC* is of particular interest, since our asymptotic framework requires having K fixed (deterministic) and $T \to \infty$, whereas K chosen with AIC/AIC* is random. For our simulations, we generate a stationary functional times series admitting the linear representation $X_t^{[\alpha]} = \sum_{s=0}^{2} \alpha_s A_s \varepsilon_{t-s}$,

where $\alpha_s \in \mathbb{R}$ are scaling parameters (described below), A_s are bounded operators, and the ε_t are i.i.d. random functions (the innovations), represented using a truncated Karhunen–Loève expansion: $\varepsilon_t(\tau) = \sum_{k=1}^{20} \xi_{k,t}\sqrt{\lambda_k}e_k(\tau)$, and $e_k(\tau) = \sqrt{2}\sin[(k-1/2)\pi\tau]$ is orthonormal system in $L^2([0,1],\mathbb{R})$, see e.g. Adler (1990). The $\xi_{k,t}$ are centered i.i.d. random variables with unit variance, whose distribution induces the distribution of the random functions ε_t. The λ_ks are numbers that describe the roughness of random curves ε_t. If $\lambda_k \to 0$ very fast, then the curves are smooth. Conversely, if λ_k decays slowly, the curves are rough. We will consider the three following scenarios for $\xi_{k,t}$ and λ_k:

`Wiener`: the $\xi_{k,t}$ are independent standard Gaussian random variables, and

$$\lambda_k = 1/[(k-1/2)^2\pi^2].$$

The random curves ε_t therefore correspond to an approximation of the Wiener process, where the approximation is due to the truncation of the Karhunen–Loève expansion of ε_t.

`White-noise`: the $\xi'_{k,t}$ are independent standard Gaussian random variables, and $\lambda_k = 1$ for all $k \geq 1$. This process corresponds to a rougher version of the `Wiener` scenario, and is a projection of a true Gaussian white noise process.

`Student5`: The $\xi_{k,t}$ are i.i.d. distributed random variables, following Student's t distribution with $\nu = 5$ degrees of freedom (rescaled to have unit variance), and $\lambda_k = 1$ for all $k \geq 1$. This process is similar to the `White-noise` process, except it is not Gaussian, and only its first 4 moments are finite. In particular, it does not satisfy Conditions B.1. It is presented here as an extreme scenario.

We have constructed the operators A_s so that their image be contained within a 20-dimensional subspace of $L^2([0,1],\mathbb{R})$, spanned by an orthonormal basis $\psi_1, \ldots, \psi_{20}$. Representing ε_t in the $(e_k)_{k=1}^{20}$ basis, and A_s in the $(\psi_m \otimes e_k)_{m,k=1}^{20}$ basis, we obtain a matrix representation of the process $X_t^{[\alpha]}$ as $\mathbf{X}_t^{[\alpha]} = \sum_{s=0}^{2} \alpha_s \mathbf{A}_s \boldsymbol{\varepsilon}_{t-s}$, where $\mathbf{X}_t^{[\alpha]}$ is a 20×1 matrix, each $\mathbf{A}_s$ is a 20×20 matrix, and $\boldsymbol{\varepsilon}_t$ is a 20×1 matrix.

The matrices $\mathbf{A}_s$ are constructed by drawing, for each of their coordinates, i.i.d. Gaussian variables with mean 1 and standard deviation 0.5. In practice, their construction is done by fixing

the random seed to a pre-chosen value and using the same generation scheme for each simulation run. The parameters $\alpha = (\alpha_0, \alpha_1, \alpha_2)$ are used to make the spectral density operators of $X_t^{[\alpha]}$ less constant. We chose

$$\alpha = \alpha(\texttt{ma.diff}) = (-1.4, 2.3, -2 + \texttt{ma.diff}) \tag{D.1}$$

where `ma.diff` is a parameter that is allowed to change. The trace of the spectral density operator for the `Wiener` scenario is shown is in Figure S1 for $\texttt{ma.diff} = 0, 0.1, \ldots, 0.5$. Visual appreciation of the roughness of each process can be obtained by plots of the percentage of explained variation at each frequency $\omega \in [0, \pi]$, that is, the proportion of the total variation of the infinitesimal increment process dZ_ω contained in its first k eigenspaces, i.e.

$$\frac{\sum_{j=1}^{k} \mu_j(\omega)}{\sum_{j=1}^{20} \mu_j(\omega)}, \quad k = 1, 2, \ldots, 20. \tag{D.2}$$

The percentages of explained variation per frequency are shown in Figure S3 for the `Wiener` scenario, and in Figure S4 for the `White-noise` and `Student5` scenarios. Notice that $k = 1$ already explains at least 80% of the variation at each frequency for the `Wiener` scenario, whereas we need to take $k = 8$ in the `White-noise` scenario to explain 80% of the variation at frequencies near 0.9.

For each $\texttt{ma.diff} \in \{0, 0.1, \ldots, 0.5\}$, and each $T \in \{2^6, 2^7, \ldots, 2^{10}\}$, we simulated for $b = 1, \ldots, B$ stretches of length T of the time series $X_t^{[\alpha(0)]}$ and $X_t^{[\alpha(\texttt{ma.diff})]}$. Denoting these observed times series by $X^{b,1}$ and $X^{b,2}$, we computed their spectral density operators using the bandwidth $B_T = T^{-1/5}$, (e.g. Grenander & Rosenblatt (1957, Par. 4.7), Brillinger (2001, Par. 7.4)) and took the weight function $W(x)$ to be the *Epanechnikov kernel* (e.g. Wand & Jones 1995), $W(x) = \frac{3}{4}(1-x^2)$ if $|x| < 1$, and zero otherwise. We then compute and store the p-values

$$p_{j,k}^b = \mathbb{P}\left(\tilde{\Delta}_k^{(T)}(\omega_j) < \chi_\nu^2(\omega_j)\right), \quad j = 1, \ldots, J; k = 1, \ldots, 10,$$

where $\tilde{\Delta}_k^{(T)}(\cdot)$ is defined in Theorem 3.1 of the paper, $\nu(\omega_j) = K(\omega_j)[K(\omega_j) - 1]/2$ if $\omega_j \in \{0, \pi\}$, and $\nu(\omega_j) = K(\omega_j)^2$ otherwise, and $\Gamma = \{\omega_1, \ldots, \omega_J\} \subset [0, \pi]$ is a grid of frequencies. Following the discussion of Section 3.3 of the paper, we choose $\omega_1 = 0$; $\omega_2 = B_T$; $\omega_{J-1} = \pi - B_T$; $\omega_J = \pi$; and $\omega_3, \ldots, \omega_{J-2}$ equispaced with spacing $\Delta\omega = \frac{2B_T}{10}$. The reason behind this choice is that the spectral density operators are computed by weighted averaging of the periodogram operator on

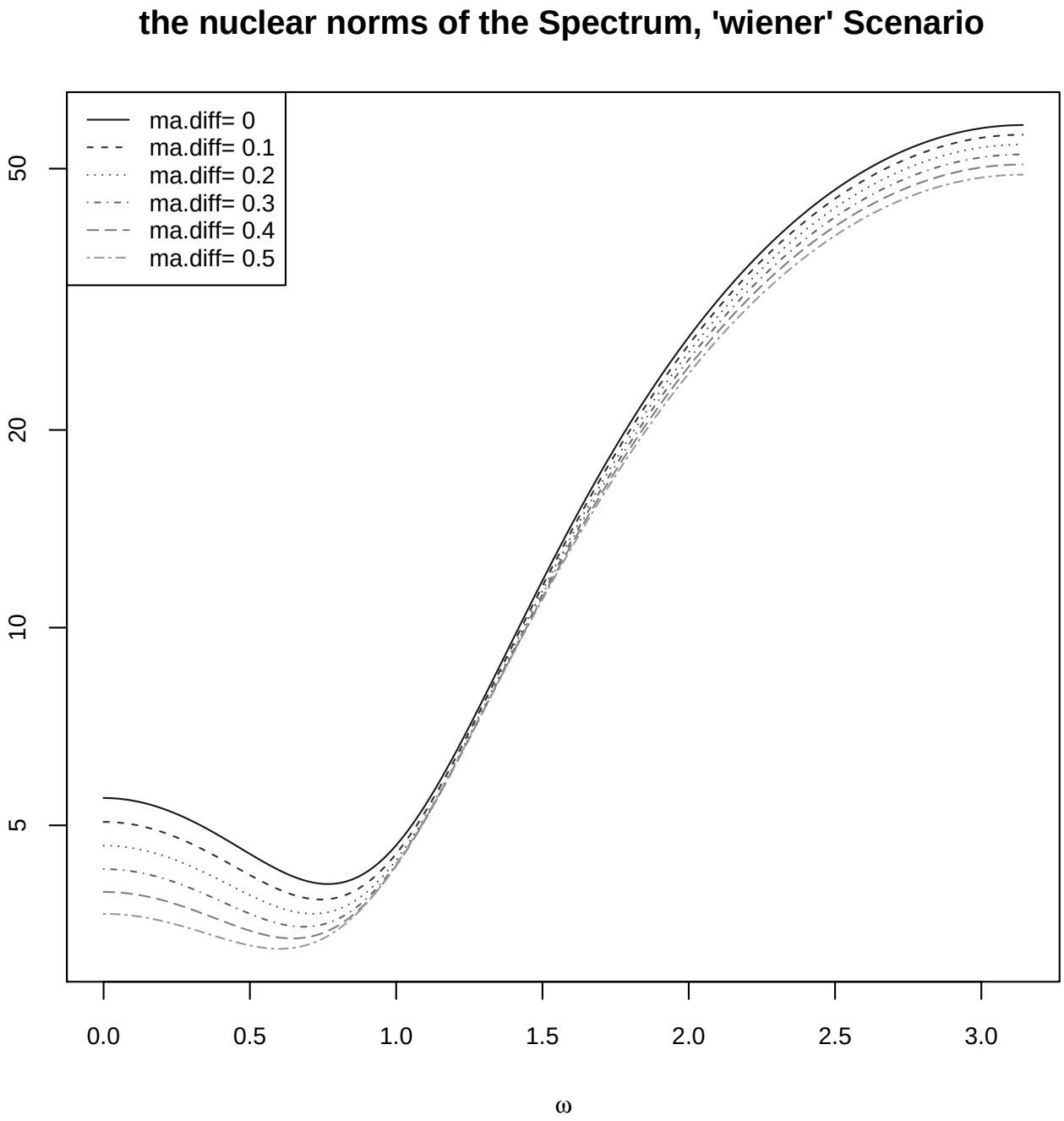


Figure S1: The trace of the spectral density operators of $X_t^{[\alpha(\texttt{ma.diff})]}$ for different values of `ma.diff`.

frequencies belonging to an interval of length $2B_T$. Therefore the spectral density operators will be very variable (approximately independent) for two frequencies separated by $2B_T$, but not very variable if the frequencies are at distance less than $2B_T$. Our choice of $\Delta\omega$ corresponds to a coverage of each interval of length $2B_T$ with at least $9 = 10 - 1$ points of the grid Γ, except of course near the extremities $\{0, \pi\}$.

For each depth level k, and each b, we adjust the p-values $(p^b_{j,k})_{j=1,\dots,J}$ using the Benjamini-Hochberg procedure. We then estimate the probability of rejecting the global null hypothesis $H_G := \bigcap_{\omega\in\Gamma} H_\omega$ at level α, within the depth k, by

$$B^{-1}\sum_{b=1}^{B} \mathbf{1}\left(\min_{j=1,\dots,J} p^b_{j,k} \leq \alpha\right). \tag{D.3}$$

The results of this procedure—with $\alpha = 5\%$, and $k \in \{1, \dots, 5\}$—are shown in Figures S5, S6 and S7 for the `Wiener`, `White-noise` and `Student5` scenarios, respectively. We notice that the level is respected in each scenario, and for each sample size $T = 128, \dots, 1024$. For the `Wiener` scenario, choosing $k = 1$ does not yield a powerful test, even with a sample size of $T = 1024$. However, taking $k = 2$ seems to already be reasonably powerful, even at sample size $T = 256$, at which it is almost as competitive as $k = 3, 4, 5$. For larger sample sizes, choosing k larger yields big differences in power: for instance in the case $T = 1024$, at `ma.diff` $= 0.2$, the power increases by roughly 0.2 if we increase k by one. The `White-noise` and `Student5` scenarios are more surprising: quite often, the case $k = 1$ yields the most (or near to the most) powerful test amongst $k = 1, 2, \dots, 5$. At sample size $T = 1024$, and `ma.diff` $= 0.3$, the difference in power between $k = 1$ and $k = 5$ is about 40%, in favour of $k = 1$.

We also show in Table S1 the true level of our testing procedure, and in Figure S8 the estimated probabilities of rejection for the global null, when the truncation level k is chosen either by AIC or by AIC*. That is, if we denote by $K(\omega)$, respectively $K^*(\omega)$, the truncation level K that minimizes the AIC, respectively the AIC* criteria, we plot the values of

$$B^{-1}\sum_{b=1}^{B} \mathbf{1}\left(\min_{j=1,\dots,J} p^b_{j,K(\omega_j)} \leq \alpha\right) \tag{D.4}$$

and

$$B^{-1}\sum_{b=1}^{B}\mathbf{1}\left(\min_{j=1,\ldots,J} p^{b}_{j,K^*(\omega_j)} \leq \alpha\right) \tag{D.5}$$

for each `ma.diff` $\in \{0, 0.1, \ldots, 0.5\}$ and $T = 64, 128, \ldots, 1024$, for $\alpha = 5\%$. We see that our testing procedure respects the level of the test in all the settings considered, except for $T = 64$ when $K(\omega)$ is chosen by AIC (this is due $K(\omega)$ taking too large values in this setting). Moreover, our procedure is generally conservative, due to the multiple testing approach taken (Benjamini & Hochberg 1995). Concerning the power of our test, we see that in the `Wiener` scenario, criterion AIC is much better than AIC*. However, in the `White-noise` or `Student5` scenarios, the opposite holds. Moreover, for the low sample sizes $T = 64$, AIC seems to fail. Except for that, in every scenario and every other sample size, both criteria seem to respect the level $\alpha = 5\%$. The power using either AIC criterion is not necessarily larger than that obtained using a fixed pre-chosen value for the truncation level k (that is, without correcting for multiplicity in the choice of k). However, it is not clear how to choose k a priori. Therefore, for most power in rejecting the global hypothesis H_G, we recommend using the AIC criterion in settings that are quite smooth (when the eigenvalues of the spectral density operators decay quickly), and the AIC* criterion in rougher settings (when the eigenvalues of the spectral density operators decay slowly).

D.2 Power as gridsize varies

Our testing method assumes the choice of a discretization of the interval $[0, \pi]$ as a given, but in practice this choice is up to the analyst and may affect the power of the test, as described in Section 3.3 of the main paper. In order to better understand the effect of the discretization at a finite sample level—particularly in the absence of prior knowledge on differences between the spectral density operators—we investigated by means of numerical simulations the level/power of our procedure under various discretizations of the interval $[0, \pi]$. We focused on discretizations $\Gamma_d \subset [B_T, \pi - B_T]$, where $d \in [1, 100]$ denotes the density of the grid: $d = 1$ corresponds to a grid with frequencies separated by at least $2B_T$ (a sparse grid), and $d = 100$ corresponds to the densest grid, the grid of Fourier frequencies within $[B_T, \pi - B_T]$:

$$\Gamma_* = \Gamma_{100} = \{\pi s/T : s = 0, 1, \ldots, T\} \cap [B_T, \pi - B_T].$$

Figure S2 shows the discretization grid Γ_d for various values of d. We show in Figure S9 how the FDR-adjusted p-values change for different values of d, given two functional time series of length $T = 1024$ that have different spectral density operators. Notice that there seems to be a convergence of the adjusted p-values as $d \to 100$. However the convergence is not perfect. This is due in part to the randomness of $K(\omega)$, but also to the continuity (in ω) of the test statistic. Looking at the level/power for different values of d, given in Figure S10, it seems that, as $d \to 100$, the quantity of information contained in the raw p-values increases less and less, and the multiplicity correction leads to a slight loss of power.

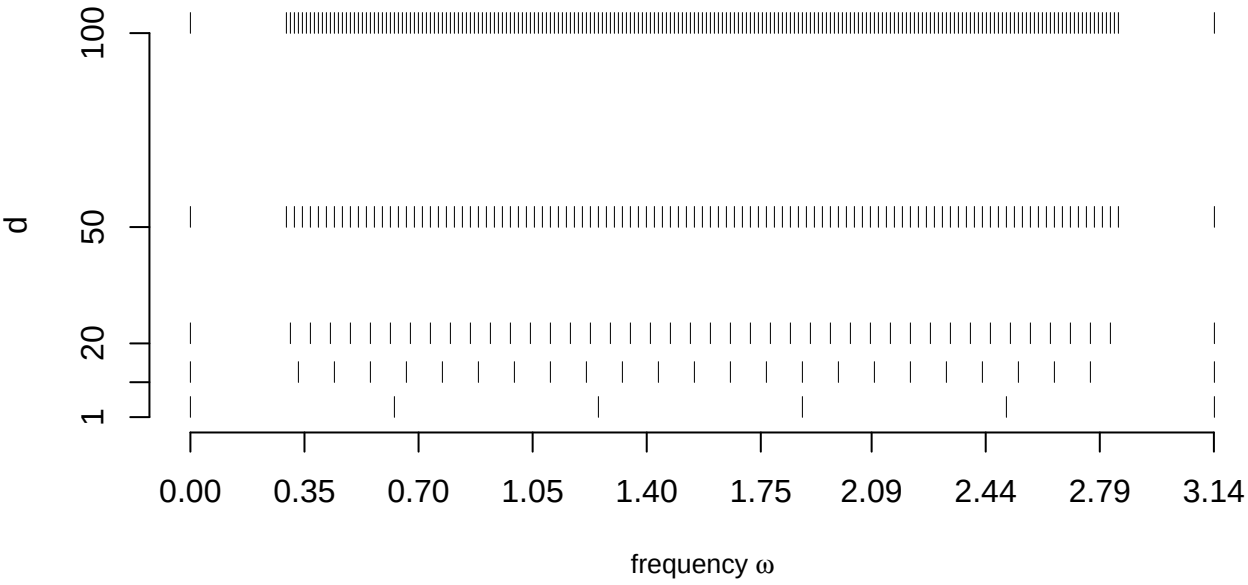


Figure S2: The discretizations Γ_d for different values of d, for $T = 512$ and $B_T = T^{-1/5} = 0.287$.

E Approximate P-values for Spatially Localizing Differences within Frequencies

Theorem 3.7 of Panaretos & Tavakoli (2013*b*) implies that, under $\bigcap_{i,j=1}^{80} H_\omega(i,j)$, $\sqrt{TB_T}(\mathbf{f}_\omega^{1,(T)} - \mathbf{f}_\omega^{2,(T)})$ will be asymptotically distributed as a random matrix $\breve{\mathbf{f}}_\omega$, which follows a complex Gaussian

	T	64	128	256	512	1024
AIC	`Wiener`	0.71	0.00	0.03	0.03	0.02
	`White-noise`	0.66	0.00	0.02	0.02	0.02
	`Student5`	0.66	0.00	0.03	0.03	0.04
AIC*	`Wiener`	0.03	0.00	0.02	0.02	0.03
	`White-noise`	0.01	0.00	0.03	0.03	0.03
	`Student5`	0.01	0.00	0.03	0.04	0.04

Table S1: Empirical level of the simulation study. The standard deviation, based on a Gaussian approximation, is at most $sd = 0.015$.

distribution (which is not necessarily circular; see Picinbono (1996), Schreier & Scharf (2010) for a definition) with mean zero, and second-order structure given by

$$\mathbb{E}\left[\breve{\mathbf{f}}_{\omega_1}(i,j)\breve{\mathbf{f}}_{\omega_2}(k,l)\right] = 4\pi\kappa^2 \cdot \left[\eta(\omega_1+\omega_2)\mathbf{f}_{\omega_1}(i,k)\mathbf{f}_{-\omega_1}(j,l) + \eta(\omega_1-\omega_2)\mathbf{f}_{\omega_1}(i,l)\mathbf{f}_{-\omega_1}(j,k)\right], \quad \text{(E.1)}$$

where $\kappa^2 = \int_{\mathbb{R}} W(x)^2 dx$ and $\eta(\omega) = 1$ if $\omega \in \{0, \pm\pi, \pm 2\pi\}$, and zero otherwise. We shall use this asymptotic distribution to obtain, for each (i,j), an approximate p-value for the null hypothesis $H_\omega(i,j)$, by renormalizing properly the difference $\mathbf{D}_\omega^{(T)} = \sqrt{TB_T}(\mathbf{f}_\omega^{1,(T)} - \mathbf{f}_\omega^{2,(T)})$. In the following, we shall use the notation $\tilde{\mathbf{f}}_\omega = (\mathbf{f}_\omega^{1,(T)} + \mathbf{f}_\omega^{2,(T)})/2$. The test statistic we use is

$$L^{(T)}(\omega,i,j) = \begin{cases} \left|\mathbf{D}_\omega^{(T)}(i,j)\right|^2 / \left\{4\pi\kappa^2\left[|\tilde{\mathbf{f}}_\omega(i,j)|^2 + \tilde{\mathbf{f}}_\omega(i,i)\tilde{\mathbf{f}}_\omega(j,j)\right]\right\} & \text{if } \omega \in \{0,\pi\} \\ \left|\mathbf{D}_\omega^{(T)}(i,i)\right|^2 / \left\{4\pi\kappa^2|\tilde{\mathbf{f}}_\omega(i,i)|^2\right\} & \text{if } \omega \in (0,\pi),\ i=j, \\ \left\{|\mathbf{D}_\omega^{(T)}(i,j)|^2/P - \Re\left[(\mathbf{D}_\omega^{(T)}(i,j))^2 R\right]/P\right\}/(2\pi\kappa^2) & \text{if } \omega \in (0,\pi),\ i\neq j, \end{cases}$$

where $\Re(\cdot)$ denotes the real part of a complex number,

$$P = P(\omega,i,j) = \tilde{\mathbf{f}}_\omega(i,i)\tilde{\mathbf{f}}_\omega(j,j) - |\tilde{\mathbf{f}}_\omega(i,j)|^4/[\tilde{\mathbf{f}}_\omega(i,i)\tilde{\mathbf{f}}_\omega(j,j)]$$

and $R = R(\omega,i,j) = [\,\overline{\tilde{\mathbf{f}}_\omega(i,j)}\,]^2/[\tilde{\mathbf{f}}_\omega(i,i)\tilde{\mathbf{f}}_\omega(j,j)]$.

The following proposition gives the asymptotic distribution of $L^{(T)}$, and its proof follows easily from results in Picinbono (1996), Panaretos & Tavakoli (2013*b*) and Tavakoli (2014):

Proposition E.1. *Assume Conditions B.1* ***(1)-(5)*** *hold, and* $B_T \to 0$ *&* $TB_T \to \infty$ *as* $T \to \infty$. *Under* $H_\omega(i,j)$, *if* $\mathbf{f}_\omega(i,i)\mathbf{f}_\omega(j,j) \neq 0$, *the asymptotic distribution of the test statistic* $L^{(T)}(\omega,i,j)$ *is*

χ^2_1 *if* $\omega \in \{0, \pi\}$ *or* $i = j$*, and* χ^2_2 *otherwise (i.e. if* $i \neq j$ *and* $\omega \in (0, \pi)$*).*

The reason the form of the test statistic $L^{(T)}(\omega, i, j)$ is quite complicated in the case $i \neq j \,\&\, \omega \in (0, \pi)$ is that in this case, $\mathbf{D}^{(T)}_\omega(i, j)$ follows a complex distribution that is not circular, and its studentization cannot be accomplished via the "usual" formula for Gaussian random vectors (see Picinbono (1996), Schreier & Scharf (2010)). Using Proposition E.1, we can compute the approximate p-values $p(\omega, i, j), \omega \in [0, \pi], i \leq j$. Notice that within a frequency, the p-values $\{p(\omega, i, j) : i \leq j\}$ are correlated, with a complicated correlation structure. For instance, for $\omega \in \{0, \pi\}$ or $i = j \;\&\; k = l$, the asymptotic covariance is given by

$$\lim_{T\to\infty} \operatorname{cov}\left(L^{(T)}(\omega, i, j), L^{(T)}(\omega, k, l)\right) = 2 \operatorname{corr}\left(\check{\mathbf{f}}_\omega(i, j), \check{\mathbf{f}}_\omega(k, l)\right)^2, \tag{E.2}$$

which is in general non-zero, see (E.1).

References

Adler, R. J. (1990), *An introduction to continuity, extrema, and related topics for general Gaussian processes.*, IMS Lecture Notes - Monograph Series. 12. Hayward, CA: Institute of Mathematical Statistics. vii, 160 p. .

Amzallag, A., Vaillant, C., Jacob, M., Unser, M., Bednar, J., Kahn, J. D., Dubochet, J., Stasiak, A. & Maddocks, J. H. (2006), '3D reconstruction and comparison of shapes of DNA minicircles observed by cryo-electron microscopy', *Nucleic acids research* **34**(18), e125–e125.

Benjamini, Y. & Hochberg, Y. (1995), 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 289–300.

Bosq, D. (2000), *Linear Processes in Function Spaces*, Springer.

Brillinger, D. R. (2001), *Time Series: Data Analysis and Theory*, Classics in Applied Mathematics, (classics edition) edn, SIAM.

Case, D., Darden, T., Cheatham, T., Simmerling, C., Wang, J., Duke, R., Luo, R., Crowley, M., Walker, R. C., Zhang, W. et al. (2008), Amber 10, Technical report, University of California.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995), 'A smooth particle mesh Ewald method', *The Journal of chemical physics* **103**(19), 8577–8593.

Gonzalez, O., Petkevičiūtė, D. & Maddocks, J. (2013), 'A sequence-dependent rigid-base model of DNA', *The Journal of chemical physics* **138**(5), 055102.

Grenander, U. & Rosenblatt, M. (1957), *Statistical analysis of stationary time series*, Wiley, New York.

Joung, I. S. & Cheatham III, T. E. (2008), 'Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations', *The journal of physical chemistry B* **112**(30), 9020–9041.

Lankas, F., Lavery, R. & Maddocks, J. H. (2006), 'Kinking occurs during molecular dynamics simulations of small DNA minicircles.', *Structure* **14**(10), 1527–34.

Lavery, R., Moakher, M., Maddocks, J., Petkeviciute, D. & Zakrzewska, K. (2009), 'Conformational analysis of nucleic acids revisited: Curves+', *Nucleic acids research* **37**(17), 5917–5929.

Lavery, R., Zakrzewska, K. & Sklenar, H. (1995), 'JUMNA (junction minimisation of nucleic acids)', *Computer physics communications* **91**(1), 135–158.

Panaretos, V. M. & Tavakoli, S. (2013*a*), 'Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series', *Stochastic Process. Appl.* **123**(7), 2779–2807.
URL: *http://dx.doi.org/10.1016/j.spa.2013.03.015*

Panaretos, V. M. & Tavakoli, S. (2013*b*), 'Fourier analysis of stationary time series in function space', *Ann. Statist.* **41**(2), 568–603.

Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A. & Orozco, M. (2007), 'Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers', *Biophysical journal* **92**(11), 3817–3829.

Picinbono, B. (1996), 'Second-order complex random vectors and normal distributions', *IEEE Transactions on Signal Processing* **44**(10), 2637–2640.

Schreier, P. J. & Scharf, L. L. (2010), *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*, Cambridge University Press, Cambridge.

Tavakoli, S. (2014), Fourier Analysis of Functional Time Series, with Applications to DNA Dynamics, PhD thesis, EPFL.

Wand, M. P. & Jones, M. C. (1995), *Kernel smoothing*, Vol. 60 of *Monographs on statistics and applied probability*, Chapman & Hall, London.

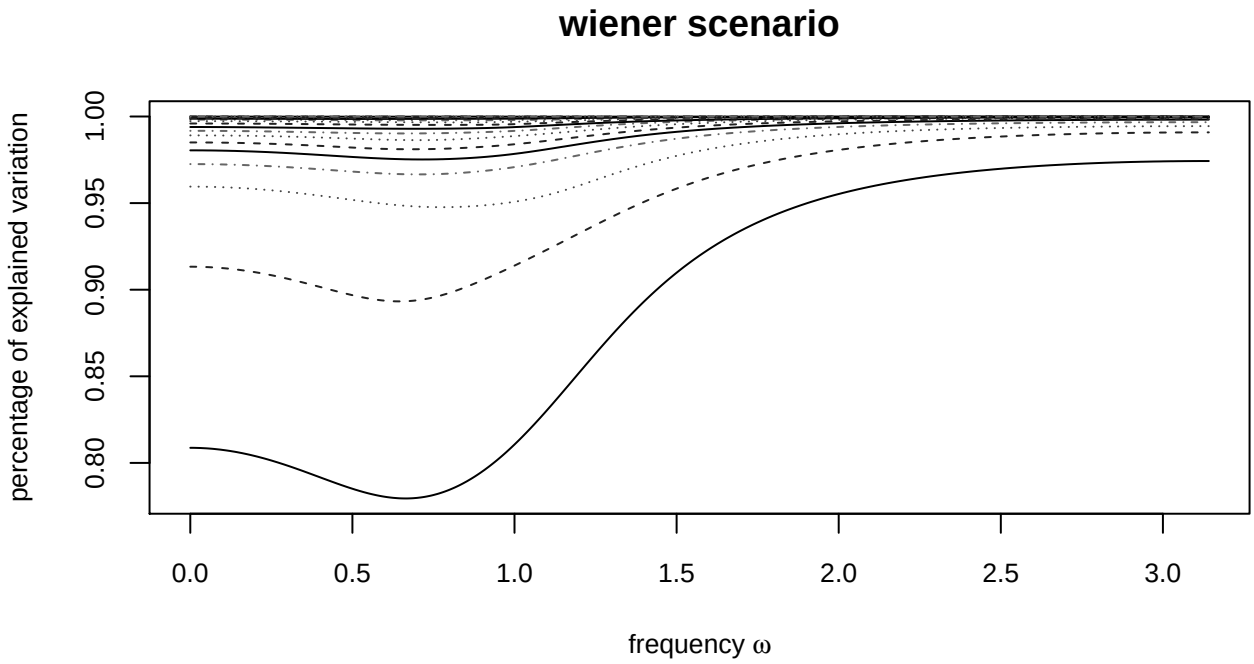


Figure S3: The percentage of variation per frequency (computed using (D.2)), for the `Wiener` scenario. The bottom curve corresponds to the percentage of variation explained by $k = 1$, the one just above corresponds to $k = 2$, as so on. Notice the scale of the y-axis.

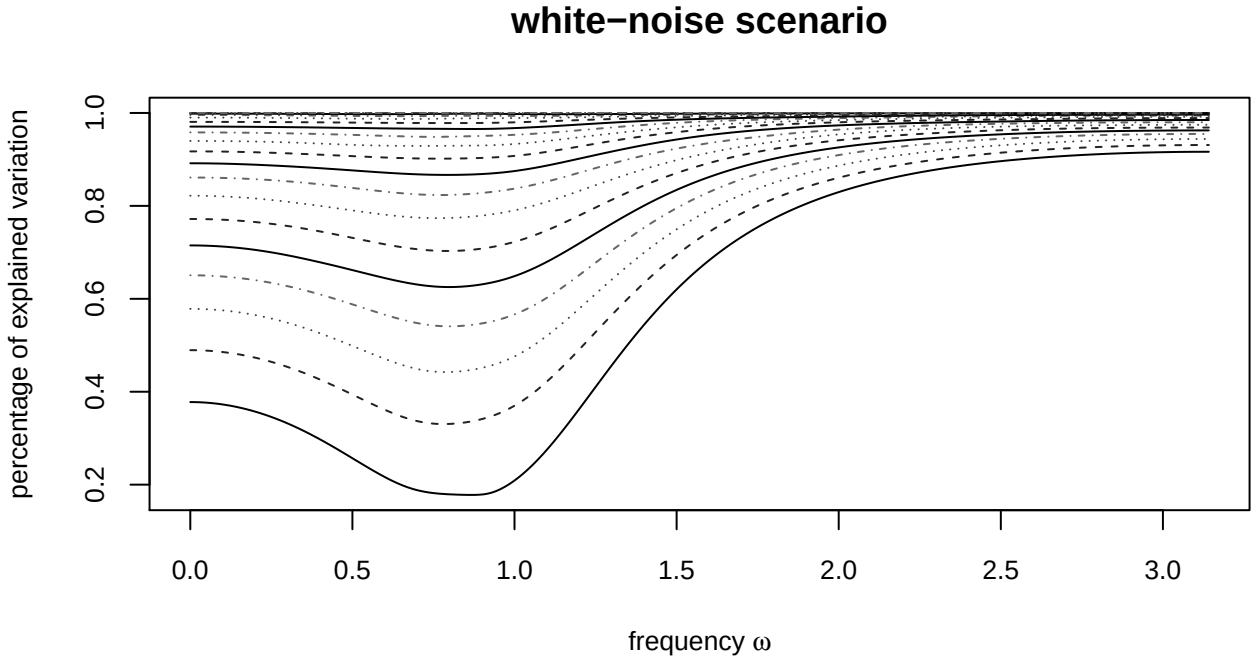


Figure S4: The percentage of variation per frequency, for the `White-noise` and `Student5` scenarios. These are computed using (D.2). The bottom curve corresponds to the percentage of variation explained by $k = 1$, the one just above corresponds to $k = 2$, as so on. Notice the scale of the y-axis.

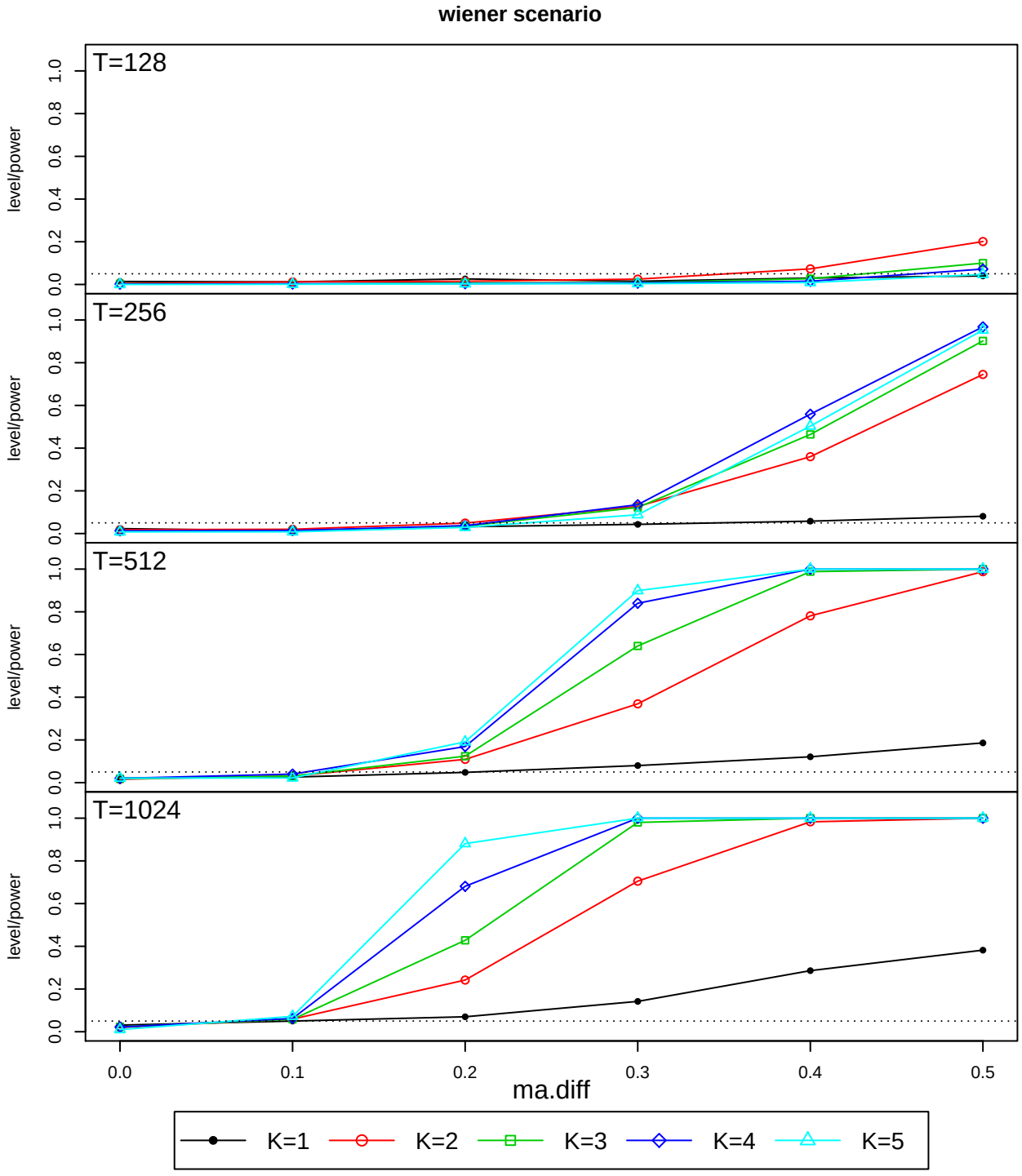


Figure S5: The estimated probability of rejection of the global null hypothesis at level 5% in the `Wiener` scenario for each depth $k \in \{1, \ldots, 5\}$, see (D.3). The x-axis represents the coefficient `ma.diff`, and each plot corresponds to a different sample size, ranging from $T = 128$ (top) to $T = 1024$ (bottom). The horizontal dotted line corresponds to the 5% level, and the standard deviation, based on a normal approximation, is at most 0.016.

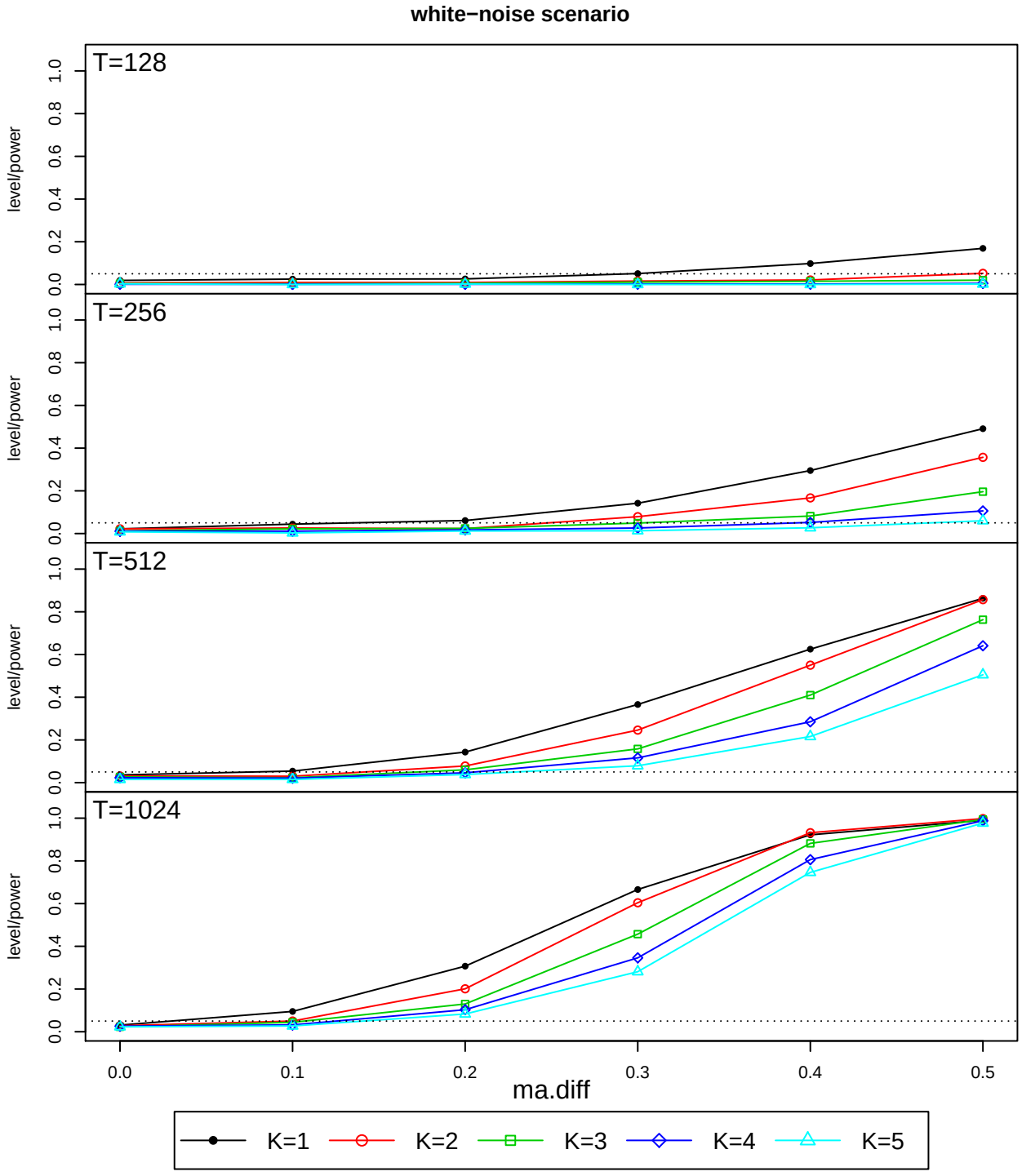


Figure S6: The estimated probability of rejection of the global null hypothesis at level 5% in the `White-noise` scenario for each depth $k \in \{1, \ldots, 5\}$, see (D.3). The x-axis represents the coefficient `ma.diff`, and each plot corresponds to a different sample size, ranging from $T = 128$ (top) to $T = 1024$ (bottom). The horizontal dotted line corresponds to the 5% level, and the standard deviation, based on a normal approximation, is at most 0.016.

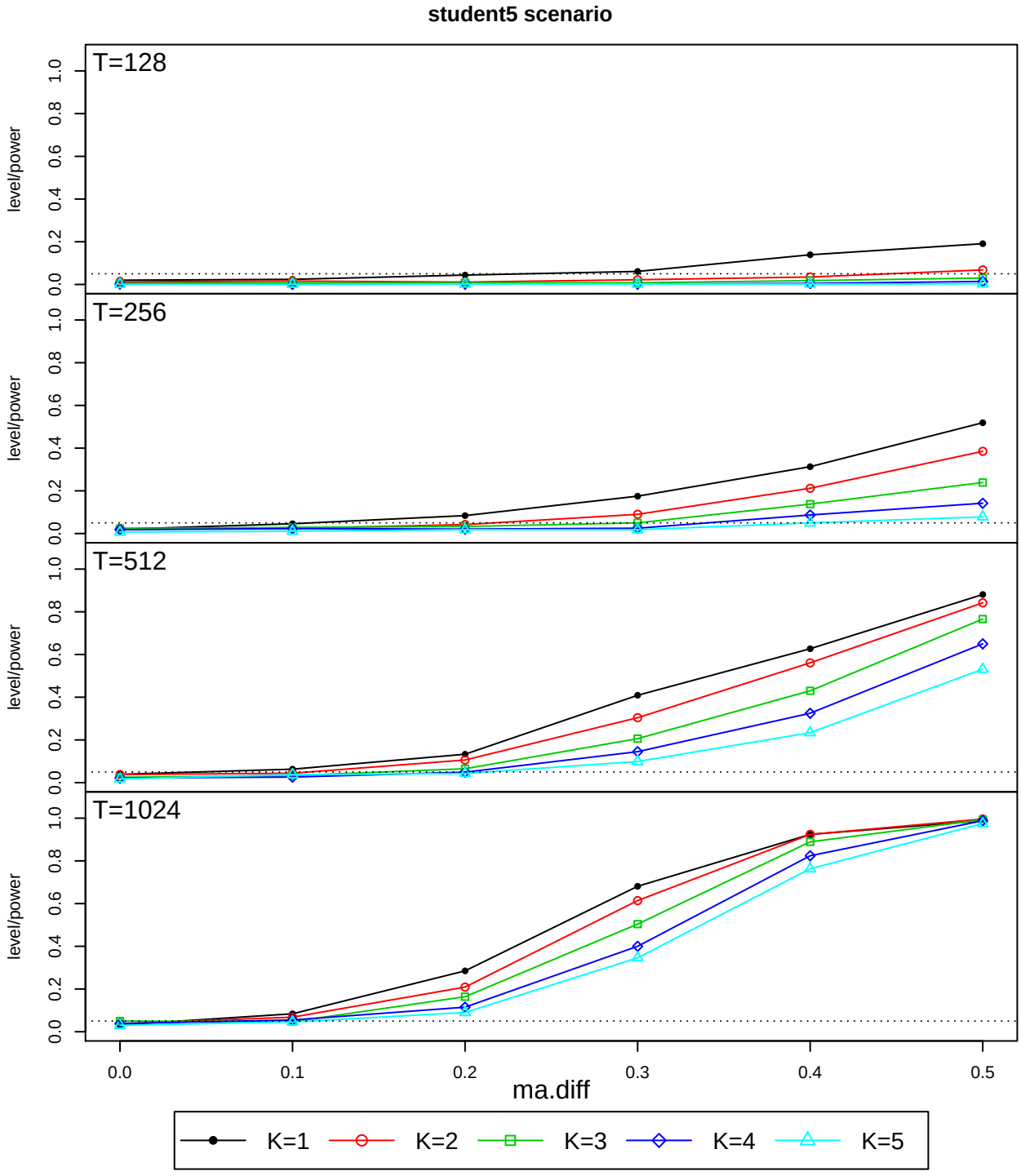


Figure S7: The estimated probability of rejection of the global null hypothesis at level 5% in the `Student5` scenario for each depth $k \in \{1, \ldots, 5\}$, see (D.3). The x-axis represents the coefficient `ma.diff`, and each plot corresponds to a different sample size, ranging from $T = 128$ (top) to $T = 1024$ (bottom). The horizontal dotted line corresponds to the 5% level, and the standard deviation, based on a normal approximation, is at most 0.016.

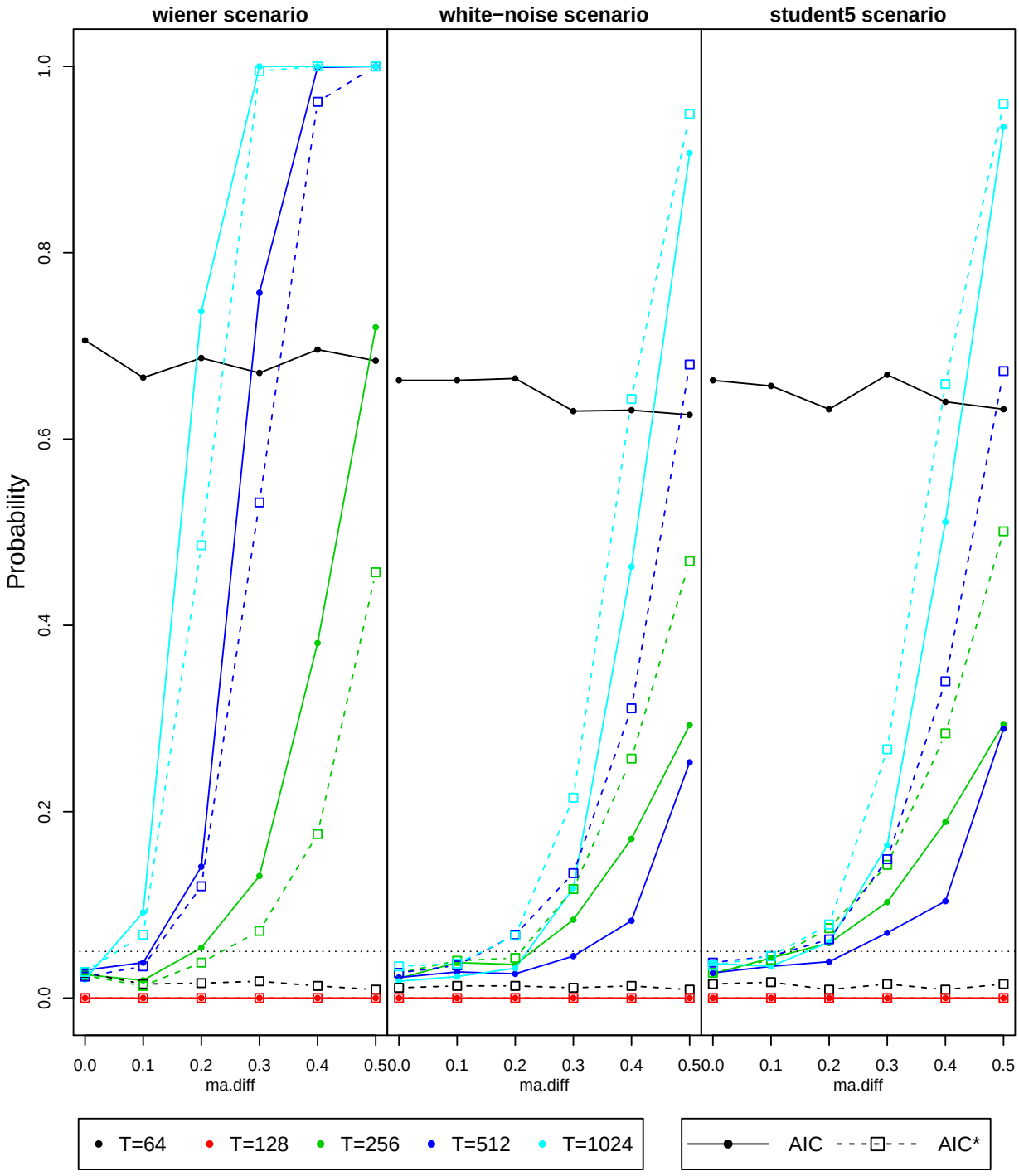


Figure S8: The estimated probability of rejecting the global null hypothesis—see (D.4) and (D.5)—when the truncation levels K are chosen either by the AIC or the AIC* criterion. The p-values have been adjusted for multiplicity by FDR, and the standard deviation, based on a normal approximation, is at most $sd = 0.016$.

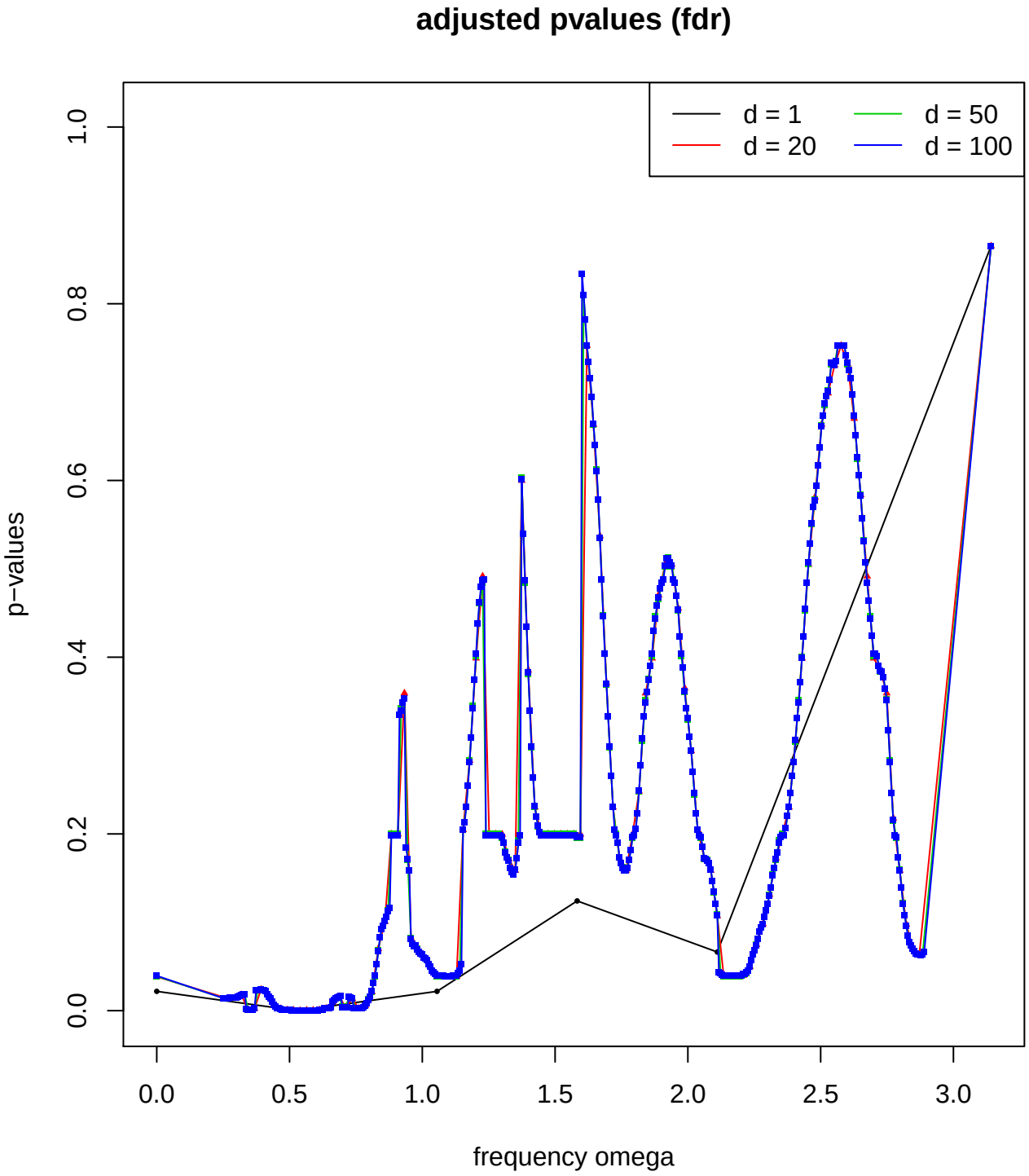


Figure S9: FDR-adjusted p-values based on the test $\tilde{\Delta}_K^{(T)}(\omega), \omega \in \Gamma_d$, with K chosen by AIC*, for various values of d, based on a two realizations of a time series of length $T = 1024$, with different spectral density operators.

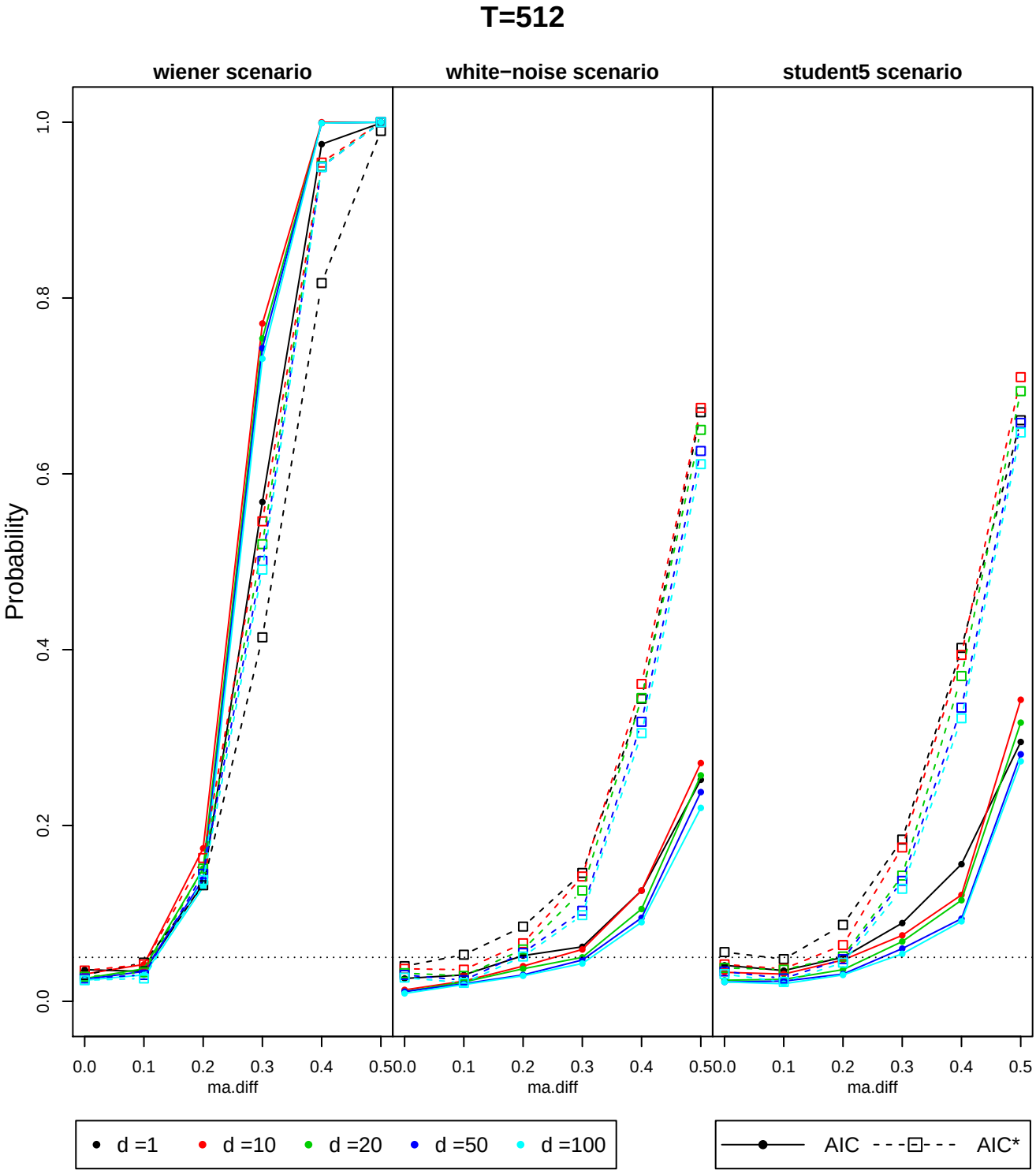


Figure S10: The estimated probability of rejecting the global null hypothesis when the truncation levels K are chosen either by the AIC or the AIC* criterion, for various discretizations Γ_d of the interval of frequencies $[0, \pi]$. The standard deviation, based on a normal approximation, is at most $sd = 0.016$.

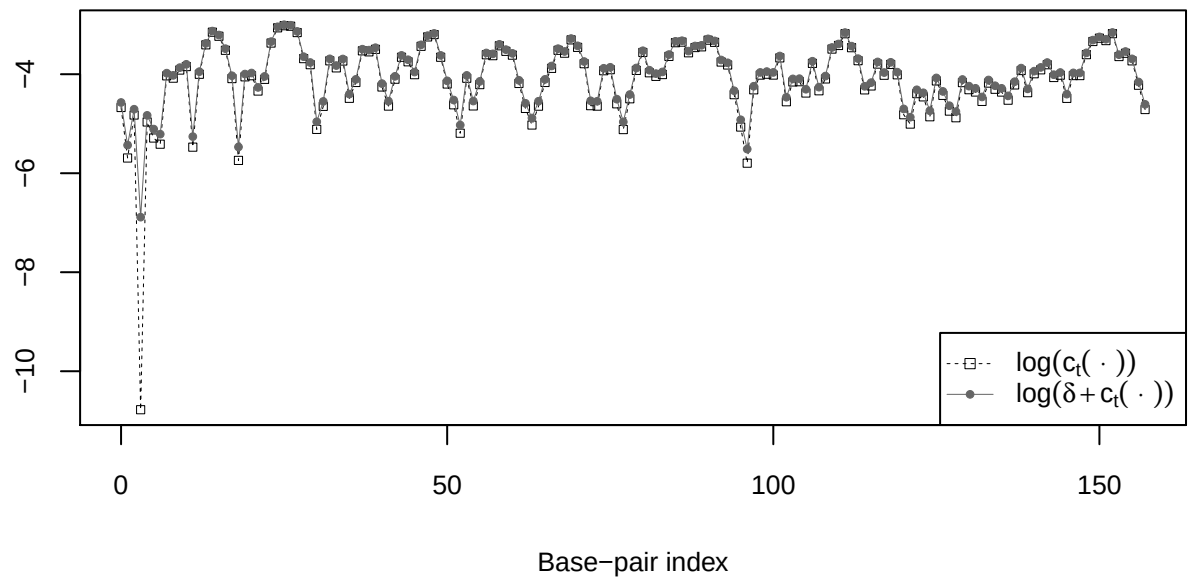


Figure S11: The effect of the constant $\delta = 10^{-3}$ in the linearization of the curvature (see (2.2) in the main paper). The dashed curve is $\log(c_t(\cdot))$, and the gray solid curve is $d_t(\cdot) = \log(\delta + c_t(\cdot))$. Notice that the d_t smooths the downward peaks that are very deep, while changing only a little other points.

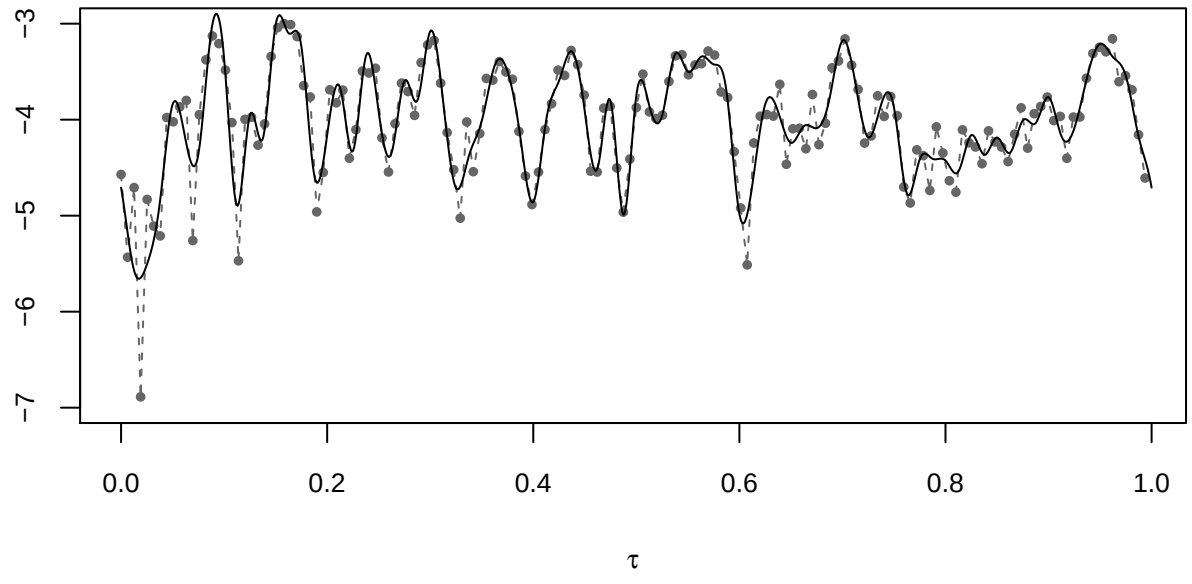


Figure S12: Illustration of the smoothing process. The dashed curve with the solid grey dots represents the scatterplot (2.3) (in the main paper), and the solid black curve represents its smoothed version obtained by projection on a basis of 80 periodic cubic B-splines.

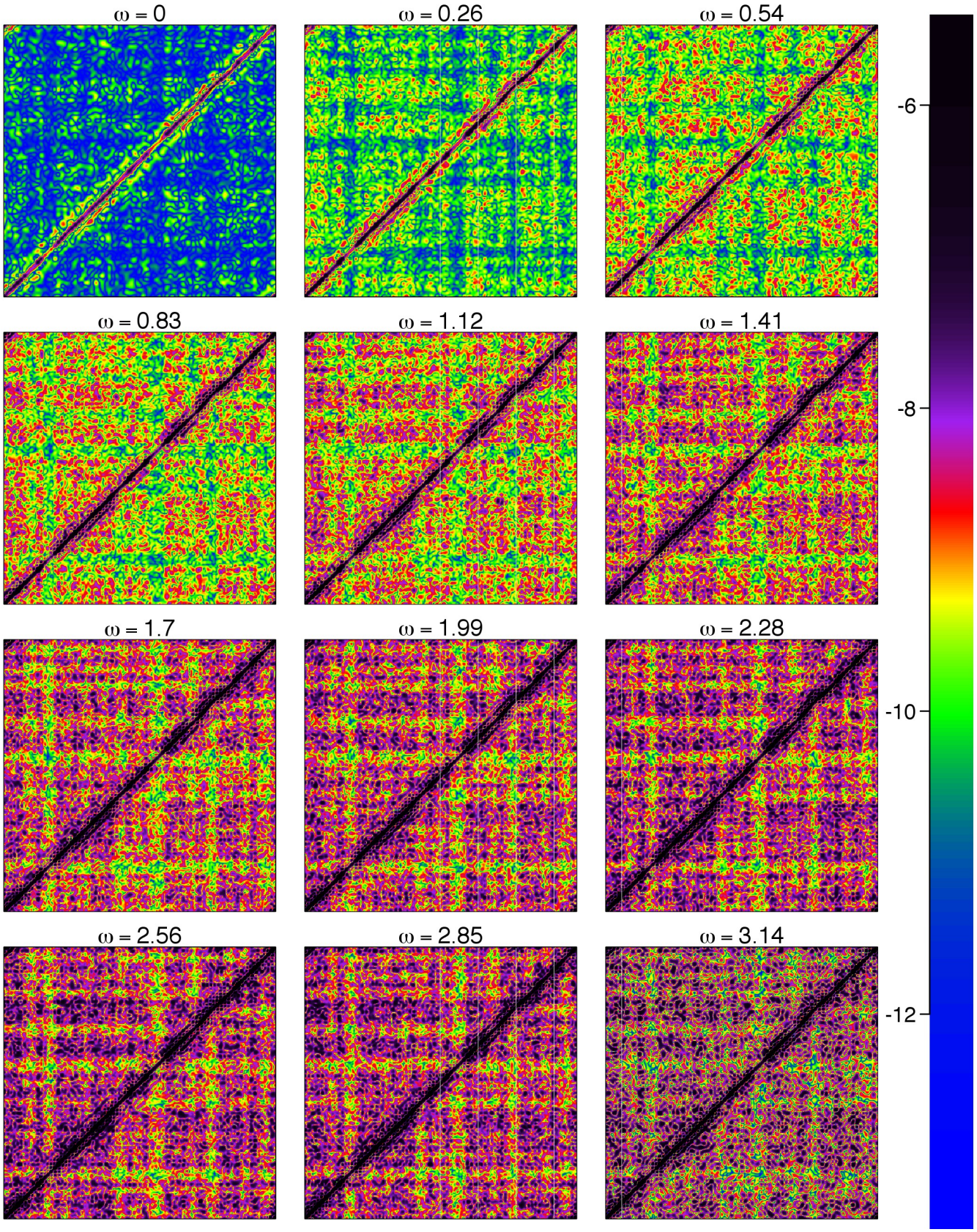
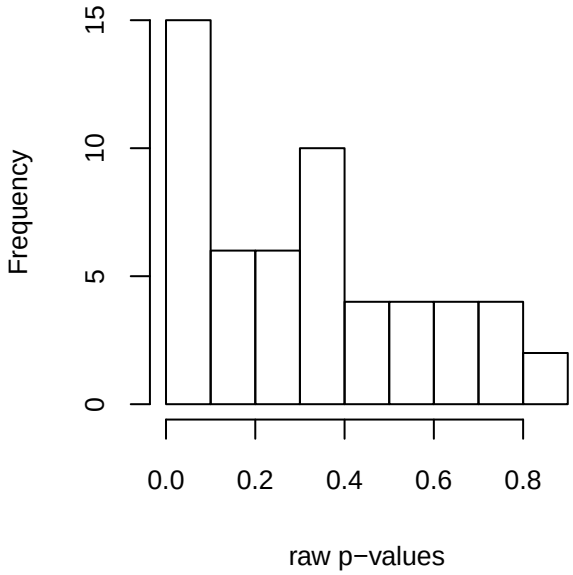
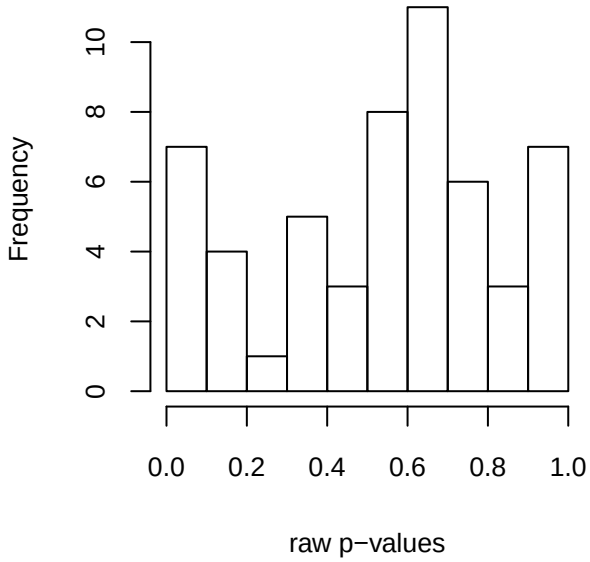


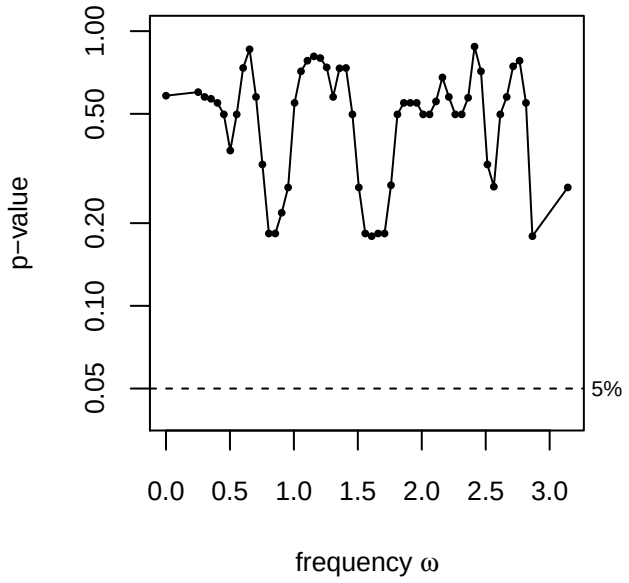
Figure S13: Plot of the sample spectral density kernels for 12 equispaced frequencies on $[0, \pi]$. For each ω, the modulus of the sample spectral density kernel of the minicircles are plotted: the upper-left part of each square represents the modulus of the sample spectral density kernel of CAP, and the lower-right part represents the corresponding quantity for TATA. Lack of symmetry between the upper-left and lower-right is a sign of differences in the sample spectral density operator of CAP and TATA.



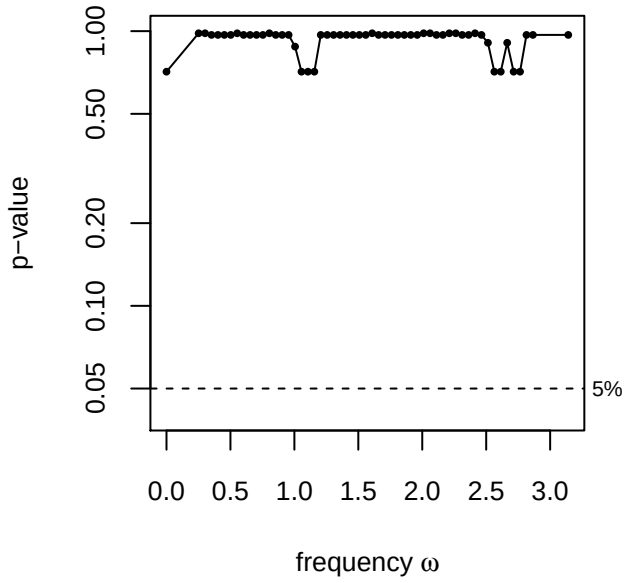
(a) Histogram of raw p-values for CAP.



(b) Histogram of raw p-values for TATA.



(c) Adjusted p-values (BH procedure) for CAP.



(d) Adjusted p-values (BH procedure) for TATA.

Figure S14: Histogram of raw p-values for testing the equality of the spectral density operator of CAP (a) and TATA (b) between the beginning of the time stretch ($t = 1, \ldots, 1000$) and the end of the time stretch ($t = 9001, \ldots, 10000$). The p-values adjusted by the Benjamini-Hochberg Procedure for CAP (c) and TATA (d). The truncation level $K(\omega)$ was chosen automatically at each frequency ω with AIC*. The dashed line represents the 5% level threshold.