The Annals of Statistics
2018, Vol. 46, No. 6A, 2960–2984
https://doi.org/10.1214/17-AOS1645

TESTING FOR PERIODICITY IN FUNCTIONAL TIME SERIES

BY SIEGFRIED HÖRMANN[1], PIOTR KOKOSZKA[2] AND GILLES NISOL[3]

Graz University of Technology, Colorado State University and Université libre de Bruxelles

We derive several tests for the presence of a periodic component in a time series of functions. We consider both the traditional setting in which the periodic functional signal is contaminated by functional white noise, and a more general setting of a weakly dependent contaminating process. Several forms of the periodic component are considered. Our tests are motivated by the likelihood principle and fall into two broad categories, which we term multivariate and fully functional. Generally, for the functional series that motivate this research, the fully functional tests exhibit a superior balance of size and power. Asymptotic null distributions of all tests are derived and their consistency is established. Their finite sample performance is examined and compared by numerical studies and application to pollution data.

1. Introduction. Periodicity is one of the most important characteristics of time series, and tests for periodicity go back to the very origins of the field, for example, Schuster (1898), Walker (1914), Fisher (1929), Jenkins and Priestley (1957), Hannan (1961), among many others. An excellent account of these early developments is given in Chapter 10 of Brockwell and Davis (1991).

We respond to the need to develop periodicity tests for time series of functions—short *functional time series (FTSs)*. Examples of FTSs include annual temperature or smoothed precipitation curves, for example, Gromenko, Kokoszka and Reimherr (2017), daily pollution level curves; Aue, Norinho and Hörmann (2015), various daily curves derived from high frequency asset price data; Horváth, Kokoszka and Rice (2014), yield curves; Hays, Shen and Huang (2012), daily vehicle traffic curves and Klepsch, Klüppelberg and Wei (2017). This work is motivated both by the need to address a general inferential problem and by specific data with which we have worked over the past decade. We first discuss the general motivation; then we illustrate it using the data.

Received June 2016; revised March 2017.

[1]Supported by the Communauté française de Belgique—Actions de Recherche Concertées, Projects Consolidation 2016–2021 and Interuniversity Attraction Poles Programme (IAP-network P7/06) of the Belgian Science Policy Office.

[2]Supported by NSF Grant DMS-1462067.

[3]Supported by the Fonds de la Recherche Scientifique—FNRS Grant MCF/FC 24535233.

MSC2010 subject classifications. Primary 62M15, 62G10; secondary 60G15, 62G20.

Key words and phrases. Functional data, time series data, periodicity, spectral analysis, testing, asymptotics.

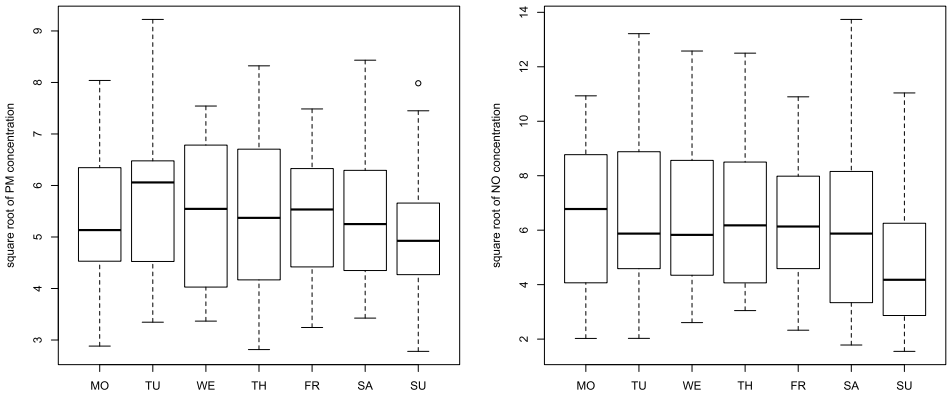

FIG. 1. *Boxplots of* PM10 (*left*) *and* NO (*right*) *for each to day of the week. The sample consists of* $N = 167$ *days.*

Most inferential procedures for FTSs require that the series be stationary; see Horváth and Kokoszka (2012). However, pollution levels, finance or traffic data may exhibit periodic (e.g., weekly) patterns, and then the stationarity assumption is violated. Horváth, Kokoszka and Rice (2014) propose several procedures to test stationarity of an FTS. Their approach is based on functionals of a CUSUM process, which makes it powerful when testing against changes in the mean or against integration of order 1. However, it is not designed for testing against a periodic signal.

Finding periodicity in a data set is also of direct relevance for understanding the problem at hand as will be illustrated in Section 6. Tests of periodicity for FTSs can be applied to the observed functions or to residual functions obtained after model fitting. If periodicity is found in the residuals, it may indicate an inadequate model.

The following motivating example, which is described in detail in Section 6, illustrates the need to develop tests that exploit the functional structure of the data. Figure 1 shows boxplots of daily averages of the pollutants PM10 (fine dust) and NO (nitrogen monoxide) measured in Graz, Austria, during the winter season 2015/16. The boxplots are grouped by weekdays and we want to infer if the corresponding group means differ significantly. Due to the traffic exposure of the measuring device in the city center and the weekday dependent traffic volumes reported in Stadlober and Pfeiler (2004), significant differences between the groups are expected. But although the boxplots indicate lower concentrations on Sundays, the variation within the groups is relatively large, and from a one-way ANOVA we do not obtain evidence against the null hypothesis of equal weekday means. The p-values are 0.75 (PM10) and 0.27 (NO), respectively. It needs to be stressed at this point, that formally ANOVA is not theoretically justified since we are analyzing time series data which are serially correlated. Nevertheless, we will see in

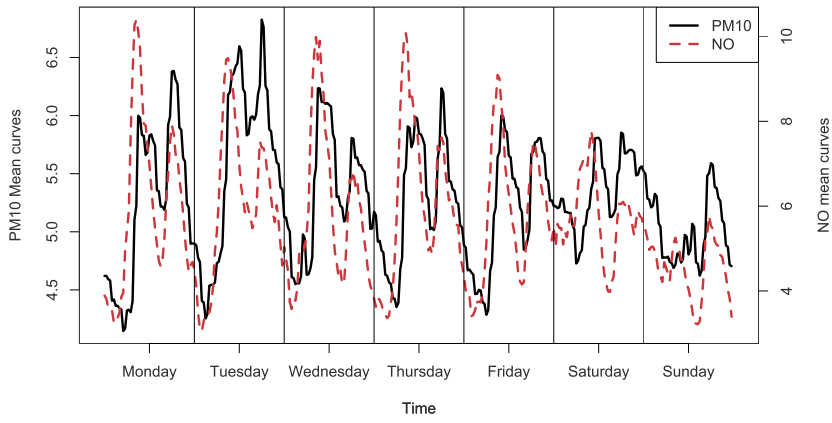

FIG. 2. *Weekday means of* PM10 (*solid black*) *and* NO (*red dashed*).

Section 6 that for the PM10 data set the conclusion remains the same even after adjusting the test for dependence. Now let us look at this problem from a functional data perspective. Figure 2 shows intraday mean curves (our raw pollution data are available up to half-hour resolution) of both pollutants during the same winter season. The plot suggests that Saturday and Sunday mean curves differ from those of working days. While they have smaller peaks, they have higher lows (presumably due to lower commuter traffic and higher nighttime activity on weekends). The methodology developed in subsequent sections, will allow us to judge whether the differences in the functional means are significant. In this particular example, the answer is affirmative. Hence, in contrast to daily averages, the intraday mean functions do show significant dependence on the day of the week.

One of the important contributions of this paper is the development of a *fully functional ANOVA test for dependent data*. Using a frequency domain approach, we obtain the asymptotic null distribution of the functional ANOVA statistic. This result is formulated in Corollary 4.1. The limiting distribution has an interesting form and can be written as a sum of independent hypoexponential variables whose parameters are eigenvalues of the spectral density operator of (Y_t). To the best of our knowledge, there exists no comparable asymptotic result in FDA literature.

Adapting ANOVA to stationary time series is one way to conduct periodicity analysis. It is suitable when the periodic component has no particular form. If, however, the alternative is more specific, then we can construct simpler and more powerful tests. In Section 2, we introduce three different models of increasing complexity, and in Section 3 we develop the appropriate test statistics. By considering specific local alternatives, the power advantage will be numerically illustrated in Section 8 and theoretical supported in the Supplementary Material [Hörmann, Kokoszka and Nisol (2018), Appendix E]. General consistency results are provided in Section 5.

We have emphasized so far fully functional procedures which are theoretically elegant and appealing. A common approach to inference for functional data is to project observations onto a low dimensional basis system, and then apply a suitable multivariate procedure to the vector of projections. This approach will be outlined in Section 3.1. Our multivariate results improve upon MacNeill (1974) in two ways: First, our tests are derived from a (Gaussian) likelihood-ratio approach. As we will see, this provides a power advantage over MacNeill's test. Second, we extend in Section 4 all of our tests to a general class of weakly dependent processes, which includes the class of linear processes studied in MacNeill (1974) and Hannan (1961).

Our methodology and theory for dependent FTSs are based on new developments in the Fourier methods for such series. The work of Panaretos and Tavakoli (2013a, 2013b) introduces the main concepts of this approach, such as the functional periodogram and spectral density operators. This framework has been recently extended and used in other contexts; see, for example, Hörmann, Kidziński and Hallin (2015) and Zhang (2016). Zamani, Haghbin and Shishebor (2016) use it in a setting that falls between our models (2.3) and (2.6) (i.i.d. Gaussian error functions), which also allows them to derive tests for hidden periodicities; the climate data they study may exhibit some a priori unspecified periods. For the data that motivate our work (pollution, traffic, temperature, economic and and finance data), the potential period is known (week, year, etc.), and they generally exhibit dependence under the null. This work therefore focuses on a fixed known period and weakly dependent functions.

The remainder of the paper is organized as follows. In Sections 2 and 3, we consider models and tests under the null of iid Gaussian functions. Section 4 considers dependent, non-Gaussian functions. Consistency of the tests is established in Section 5. Applications to pollution data and a simulation study are presented, respectively, in Sections 6 and 7. In Section 8, we numerically assess asymptotic local power of the tests. The main contributions and findings are summarized in Section 9. All technical results and proofs that are not essential to understand and apply the new methodology are presented in the Supplementary Material [Hörmann, Kokoszka and Nisol (2018)].

2. Models for periodic functional time series. The classical model, Fisher (1929), for a (scalar) periodic signal contaminated by noise is

$$y_t = \mu + \alpha \cos(t\theta) + \beta \sin(t\theta) + z_t, \tag{2.1}$$

where the z_t are normal white noise, α, β and μ are unknown constants and $\theta \in [-\pi, \pi]$ is a known frequency which determines the period. Model (2.1) has been extended in several directions, for example, by replacing a pure harmonic wave by an arbitrary periodic component and/or by replacing the normal white noise by a more general stationary time series, as well as by considering multivariate series.

In this section, we list extensions to functional time series organizing them by increasing complexity. Our theory is valid in an arbitrary separable Hilbert space H, in which $\langle x, y\rangle$ denotes the inner product and $\|x\| = \sqrt{\langle x, x\rangle}$ the corresponding norm, $x, y \in H$. In most applications, it is the space L^2 of square integrable functions on a compact interval, in which case $\langle x, y\rangle = \int x(u)y(u)\,du$. For simplicity of presentation, we stick with this setting in our paper. A comprehensive exposition of Hilbert space theory for functional data is given in Hsing and Eubank (2015).

We begin by stating the following (preliminary) assumption on the functional noise process.

ASSUMPTION 2.1. The noise (Z_t) is an i.i.d. sequence in H, with each Z_t being a Gaussian element in H with zero mean and covariance operator Γ.

Recall that a random variable Z in H is Gaussian, in short $Z \sim \mathcal{N}_H(\mu, \Gamma)$, if and only if all projections $\langle Z, v\rangle$, $v \in H$, are normally distributed with mean $\langle \mu, v\rangle$ and variance $\langle \Gamma(v), v\rangle$. Working under Assumption 2.1 is convenient because we can motivate our tests proposed in Section 3 by a likelihood ratio approach and calculate exact distributions. Nevertheless, this framework is too restrictive for many applied problems. We devote Section 4 to procedures applicable in case of noise which is a general stationary functional time series. The testing problems remain the same, but the test statistics and/or critical values change.

To make the exposition more specific and focused on the main ideas, we introduce the following assumption.

ASSUMPTION 2.2. The sample size N is a multiple of the period, $N = dn$, where the period $d > 1$ is odd. We set $q = (d-1)/2$.

Section A discusses modifications needed in case of even d. Assuming that the sample size N is a multiple of d is not really restrictive and can easily be achieved by trimming up to $d-1$ data points.

The simplest extension of model (2.1) to a functional setting is

$$Y_t(u) = \mu(u) + \big[\alpha\cos(t\theta) + \beta\sin(t\theta)\big]w(u) + Z_t(u), \tag{M.1}$$

with $\mu, w \in H$ and $\alpha, \beta \in \mathbb{R}$. If $\rho := \sqrt{\alpha^2+\beta^2} = 0$, then $(Y_t : t \geq 1)$ is functional Gaussian white noise with a mean function μ. If $\rho > 0$, then a periodic pattern is added, which varies along the direction of a function w. To ensure identifiability, we assume that $\|w\|^2 := \int_0^1 w^2(u)\,du = 1$. The functions μ and w, as well as the parameters α and β are assumed to be unknown. As explained in the Introduction, the parameter θ, which determines the period d, is assumed to be a *known* positive fundamental frequency, that is,

$$\theta \in \Theta_N := \{\theta_j = 2\pi j/N,\ j = 1, \ldots, m := [(N-1)/2]\}.$$

The testing problem is

$$\mathcal{H}_0\colon \rho = 0 \quad \text{vs.} \quad \mathcal{H}_A\colon \rho > 0. \tag{2.2}$$

A first extension of (M.1) is to replace $\alpha\cos(\theta t) + \beta\sin(\theta t)$ by some arbitrary d–periodic sequence. A more general model thus is

$$Y_t(u) = \mu(u) + s_t w(u) + Z_t(u), \qquad s_t = s_{t+d},\ \mu, w \in H. \tag{M.2}$$

We wish to test

$$\mathcal{H}_0\colon s_1 = s_2 = \cdots = s_d = 0 \quad \text{against} \quad \mathcal{H}_A\colon \max_{1\le t\le d} |s_t| > 0. \tag{2.3}$$

Here, we impose the identifiability constraints $\|w\| = 1$ and $\sum_{k=1}^{d} s_t = 0$. The latter ensures that the vector $(s_1, \ldots, s_d)'$ is contained in the subspace spanned by the orthogonal vectors

$$\begin{pmatrix} \cos(\theta_n) \\ \cos(2\theta_n) \\ \vdots \\ \cos(d\theta_n) \end{pmatrix}, \begin{pmatrix} \sin(\theta_n) \\ \sin(2\theta_n) \\ \vdots \\ \sin(d\theta_n) \end{pmatrix}, \ldots, \begin{pmatrix} \cos(\theta_{nq}) \\ \cos(2\theta_{nq}) \\ \vdots \\ \cos(d\theta_{nq}) \end{pmatrix}, \begin{pmatrix} \sin(\theta_{nq}) \\ \sin(2\theta_{nq}) \\ \vdots \\ \sin(d\theta_{nq}) \end{pmatrix},$$

cf. Assumption 2.2. With the convention

$$\vartheta_k := \theta_{nk} = 2\pi k/d, \tag{2.4}$$

model (M.2) can be written as

$$Y_t(u) = \mu(u) + \left(\sum_{k=1}^{q} \big(\alpha_k \cos(t\vartheta_k) + \beta_k \sin(t\vartheta_k)\big)\right) w(u) + Z_t(u), \tag{2.5}$$

with some coefficients α_k and β_k.

Model (M.2) assumes that at any point of time, the periodic functional component is proportional to a single function w. A model which imposes periodicity in a very general sense is

$$Y_t(u) = \mu(u) + w_t(u) + Z_t(u), \qquad \mu, w_t \in H, \tag{M.3}$$

with $w_t = w_{t+d}$ and $\sum_{t=1}^{d} w_t = 0$. In this context, we test

$$\mathcal{H}_0\colon w_1 = w_2 = \cdots = w_d = 0 \quad \text{against} \quad \mathcal{H}_A\colon \max_{1\le t\le d} \|w_t\| > 0. \tag{2.6}$$

The three models are nested, but coincide under $\mathcal{H}_0$. Test procedures presented in Section 3 are motivated by specific models as they point toward specific alternatives. However, they can be applied to any data, and, as we demonstrate in Sections 6 and 7, tests motivated by simple models often perform very well for more complex alternatives.

3. Test procedures in presence of Gaussian noise. Throughout this section, we work under Assumptions 2.1 and 2.2. In Section 4 and Section A, respectively, we show how to remove these assumptions. Details of mathematical derivations are given in Section B.

Let us start by introducing the necessary notation and notational conventions. Given a vector time series $(\boldsymbol{Y}_t : 1 \le t \le N)$ the discrete Fourier transform (DFT) is $\boldsymbol{D}(\theta) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \boldsymbol{Y}_k e^{-\mathrm{i}k\theta}$, $\theta \in [-\pi, \pi]$. We will use the decomposition into real and complex parts: $\boldsymbol{D}(\theta) = \boldsymbol{R}(\theta) + \mathrm{i}\boldsymbol{C}(\theta)$. At some places, we may add a subscript to indicate the dependence on the sample size and/or a superscript to refer to the underlying data [e.g., $\boldsymbol{R}_N^{\boldsymbol{Y}}(\theta)$]. We proceed analogously for a functional time series $(Y_t : 1 \le t \le N)$. Then the DFT is denoted by $D(\theta) = R(\theta) + \mathrm{i}C(\theta)$.

Set $\boldsymbol{A}(\theta_{i_1}, \ldots, \theta_{i_k}) = [\boldsymbol{R}(\theta_{i_1}), \ldots, \boldsymbol{R}(\theta_{i_k}), \boldsymbol{C}(\theta_{i_1}), \ldots, \boldsymbol{C}(\theta_{i_k})]'$ and analogously define $A(\theta_{i_1}, \ldots, \theta_{i_k})$ to be a $2k$-vector of functions with components $R(\theta_{i_j})$ and $C(\theta_{i_j})$. If $A = (A_1, \ldots, A_k)'$ is any k-vector of functions, then AA' is the $k \times k$ matrix of scalar products $\langle A_i, A_j \rangle$. We use $\|M\|$ for the usual (Euclidean) norm and $\|M\|_{\mathrm{tr}}$ for the trace norm of some generic matrix M. Finally, $\mathbf{W}_p(n)$ denotes the real $p \times p$ Wishart matrix with n degrees of freedom and $q_\alpha(X)$ is the α-quantile of some variable X.

3.1. *Projection based approaches.* Typically, functional data are represented in a smoothed form by finite dimensional systems, such as B-splines, Fourier basis, wavelets, etc. Additional dimension reduction can be achieved by functional principal components or similar data-driven systems. It is thus natural to search for a periodic pattern within a lower dimensional approximation of the data.

In this section, we assume that $v_1, v_2, \ldots, v_p$ is a suitably chosen set of linearly independent functions. Section F shows that if the v_k are replaced by consistent estimates, then under mild condition (e.g., conditions met by empirical functional principal components), the asymptotic distribution of the test statistics is not affected. Setting $\boldsymbol{Y}_t := (\langle Y_t, v_1 \rangle, \ldots, \langle Y_t, v_p \rangle)'$, we obtain vector observations. Under $\mathcal{H}_0$, the time series $(\boldsymbol{Y}_t)$ is i.i.d. Gaussian with covariance matrix $\boldsymbol{\Sigma} = (\langle \Gamma(v_i), v_j \rangle : 1 \le i, j \le p)$. Under $\mathcal{H}_A$, we can write the projected version of model (M.3) as

$$\boldsymbol{Y}_t = \boldsymbol{\mu} + \boldsymbol{w}_t + \boldsymbol{Z}_t, \tag{3.1}$$

with $\boldsymbol{\mu} = (\langle \mu, v_1 \rangle \cdots \langle \mu, v_p \rangle)'$, $\boldsymbol{w}_t = (\langle w_t, v_1 \rangle \cdots \langle w_t, v_p \rangle)'$ and the innovations $\boldsymbol{Z}_t = (\langle Z_t, v_1 \rangle \cdots \langle Z_t, v_p \rangle)'$. This in turn can be specialized to projected versions of models (M.1) and (M.2). The periodic component can be detected if it is not orthogonal to $\operatorname{span}\{v_1, v_2, \ldots, v_p\}$. In the following theorem, we state the likelihood ratio tests. Recall the definition of the frequencies ϑ_k in (2.4) and the notation $q = (d-1)/2$.

THEOREM 3.1. *Assuming known* $\mathbf{\Sigma}$, *the likelihood-ratio tests for the multivariate analogues of testing problems* (2.2), (2.3) *and* (2.6) [*related to the projected models* (M.1), (M.2) *and* (M.3), *respectively*] *are given as follows*: *Reject the null-hypothesis at level* α *if*

$$T^{\mathrm{MEV}_1} := \big\|\boldsymbol{A}(\vartheta_1)\mathbf{\Sigma}^{-1}\boldsymbol{A}'(\vartheta_1)\big\| > q_{1-\alpha}\big[\|\mathbf{W}_p(2)\|/2\big];$$

$$T^{\mathrm{MEV}_2} := \big\|\boldsymbol{A}(\vartheta_1,\dots,\vartheta_q)\mathbf{\Sigma}^{-1}\boldsymbol{A}'(\vartheta_1,\dots,\vartheta_q)\big\| > q_{1-\alpha}\big[\|\mathbf{W}_p(d-1)\|/2\big];$$

$$T^{\mathrm{MTR}_2} := \big\|\boldsymbol{A}(\vartheta_1,\dots,\vartheta_q)\mathbf{\Sigma}^{-1}\boldsymbol{A}'(\vartheta_1,\dots,\vartheta_q)\big\|_{\mathrm{tr}} > q_{1-\alpha}\big[\mathrm{Erlang}(pq,1)\big].$$

Some remarks are in order:

1. The superscript MEV in our tests stands for Multivariate EigenValue. *Multivariate*, as opposed to functional, and *eigenvalue*, refers to the fact that the Euclidean matrix norm of a symmetric matrix is equal to its largest eigenvalue. MTR abbreviates Multivariate TRace.

2. By Lemma B.1, the columns of $\mathbf{\Sigma}^{-1/2}\boldsymbol{A}'(\vartheta_1,\dots,\vartheta_q)$ are i.i.d. $\mathcal{N}_p(0,\frac{1}{2}I_p)$. This explains the Wishart distribution. For explicit computation of the quantiles $q_{1-\alpha}[\|\mathbf{W}_p(k)\|]$, we refer to Chiani (2014).

3. An alternative to the test based on T^{MEV_1} is

$$T^{\mathrm{MTR}_1} = \big\|\boldsymbol{A}(\vartheta_1)\mathbf{\Sigma}^{-1}\boldsymbol{A}'(\vartheta_1)\big\|_{\mathrm{tr}} > q_{1-\alpha}\big[\mathrm{Erlang}(p,1)\big].$$

The latter can be seen to be equivalent to the test proposed by MacNeill (1974) for a multivariate version of model (M.1). The likelihood ratio and MacNeill's test statistic are related to different matrix norms of $\boldsymbol{A}(\vartheta_1)\mathbf{\Sigma}^{-1}\boldsymbol{A}'(\vartheta_1)$. By the Neyman–Pearson lemma, a likelihood ratio test, even in an approximate form, can be expected to have good and sometimes even optimal power properties. Likewise, replacing the matrix norm in T^{MEV_2} by the trace norm leads to T^{MTR_2}. As Figure 3

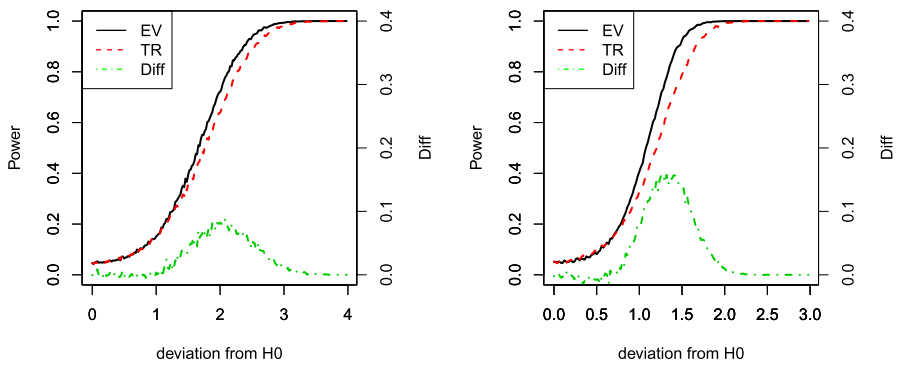

FIG. 3. *Local asymptotic power curves of tests* T^{MEV_2} (*EV*) *and* T^{MTR_2} (*TR*), *and their difference* (*Diff*, *right scale*). *Left panel*: $p = 5$ *and* $d = 7$, *right panel*: $p = 5$ *and* $d = 31$. *Details of the implementation are given in Section* 8.

illustrates, the difference in power between the two tests can be quite noticeable, especially when d is large.

4. In practice, $\mathbf{\Sigma}$ must be replaced by a consistent estimator. General construction of estimators, which remain consistent under $\mathcal{H}_A$, is discussed in Section D. A natural choice is

$$\hat{\mathbf{\Sigma}}_A := \frac{1}{N}\sum_{k=1}^{d}\sum_{i=1}^{n}(\boldsymbol{Y}_{ki} - \overline{\boldsymbol{Y}}_k)(\boldsymbol{Y}_{ki} - \overline{\boldsymbol{Y}}_k)'$$

with $\boldsymbol{Y}_{ki} = \boldsymbol{Y}_{k+(i-1)d}$ and $\overline{\boldsymbol{Y}}_k = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Y}_{ki}$. With this plug-in estimator, one can show that the resulting test statistics are asymptotically equivalent to the LR approach with unknown $\mathbf{\Sigma}$. It is possible to directly formulate time domain likelihood ratio tests based on unknown $\mathbf{\Sigma}$ (Wilks' Lambda), but it is not evident how to extend them to the fully functional setting. Using a spectral domain formulation with known $\mathbf{\Sigma}$ points toward an extension to the fully functional tests introduced in the next section and will allow for adapting the method for stationary noise; see Section 4.

3.2. *Fully functional tests.* The projection based approaches of the previous section may be sensitive to the choice of the basis and to the number of basis functions. It is therefore desirable to develop some fully functional procedures to bypass this problem. Before we introduce fully functional test statistics, let us observe that $T^{\mathrm{MEV_i}}$ and $T^{\mathrm{MTR_i}}$ $(i = 1, 2)$ are computed from the rescaled sample $\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_1, \ldots, \mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_N$, which results in asymptotically pivotal tests. The rescaling guarantees that the component processes with larger variation are not concealing potential periodic patterns in components with little variance. While this is clearly a very desirable property in multivariate analysis, one may favor a different perspective for functional data. If $\boldsymbol{Y}_t$ are principal component scores, then $\mathbf{\Sigma} = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$, where λ_i are the eigenvalues of $\operatorname{Cov}(Z_1)$. Suppose that $Y_t(u) = \sqrt{\lambda_\ell}\cos(2\pi t/d)v_\ell(u) + Z_t(u)$, $\ell \geq 1$. Then, due to $\lambda_\ell \to 0$, the bigger ℓ, the smaller and more negligible the periodic signal is. However, it is easily seen that for any of our multivariate tests, the probability of rejecting $\mathcal{H}_0$ is the same for all values $1 \leq \ell \leq p$.

A way to account for the functional nature of the data is to base the test statistics directly on the unscaled and fully functional observations, that is, to define analogues of the test statistics in Theorem 3.1 with the matrices $A(\vartheta_1)A'(\vartheta_1)$ (in $\mathbb{R}^{2\times 2}$) and $A(\vartheta_1, \ldots, \vartheta_q)A'(\vartheta_1, \ldots, \vartheta_q)$ (in $\mathbb{R}^{(d-1)\times(d-1)}$). Since, to the best of our knowledge, there is no result available on the distribution of $\|A(\vartheta_1, \ldots, \vartheta_q)A'(\vartheta_1, \ldots, \vartheta_q)\|$, we shall only consider the trace norm for which we can get explicit formulas. Hence, for model (M.1) we propose a test which rejects $\mathcal{H}_0$ at level α if

$$T^{\mathrm{FTR_1}} := \big\|A(\vartheta_1)A'(\vartheta_1)\big\|_{\mathrm{tr}} > q_{1-\alpha}\big[\mathrm{HExp}(\lambda_1, \lambda_2, \ldots)\big].$$

Here, $\mathrm{HExp}(\lambda_1, \lambda_2, \ldots)$ denotes a random variable which is distributed as $\sum_{i\geq 1} \lambda_i E_i$, where the E_i are i.i.d. Exp(1) variables. If $\lambda_i = 0$ for $i > k$, then this is a so-called hypoexponential distribution, whose distribution function is explicitly known; see, for example, Ross (2010), Section 5.2.4. For models (M.2) and (M.3) we propose the test which rejects $\mathcal{H}_0$ at level α if

$$T^{\mathrm{FTR}_2} := \big\|A(\vartheta_1, \ldots, \vartheta_q)A'(\vartheta_1, \ldots, \vartheta_q)\big\|_{\mathrm{tr}} > q_{1-\alpha}\left[\sum_{k=1}^{q} \Xi_k\right], \tag{3.2}$$

where $\Xi_k \overset{\text{i.i.d.}}{\sim} \mathrm{HExp}(\lambda_1, \lambda_2, \ldots)$. Lemma B.2 provides the justification of (3.2).

In practice, we will approximate $\mathrm{HExp}(\lambda_1, \lambda_2, \ldots)$ by $\mathrm{HExp}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_k)$ with eigenvalues $\hat{\lambda}_i$ of $\widehat{\Gamma}$ and some fixed k to obtain critical values. (See Section D.) Since the sample covariance has only a finite number of nonzero eigenvalues, we can either use all of them or chose the smallest $k \geq 1$ such that $\mathrm{tr}(\widehat{\Gamma}) - (\hat{\lambda}_1 + \cdots + \hat{\lambda}_k) \leq \varepsilon$ for some small ε. Other details, including the rate of the approximation of $\mathrm{HExp}(\lambda_1, \lambda_2, \ldots)$ by $\mathrm{HExp}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_k)$, are presented in Section D.

3.3. *Relation to MANOVA and functional ANOVA.* A possible strategy for our testing problem is to embed it into the ANOVA framework as it was sketched in the Introduction. If the period is d, we can think of the data as coming from d groups, and the objective is to test if all groups have the same mean. ANOVA can be applied to models (M.1) and (M.2), but it is particularly suitable for model (M.3) where we impose no structural assumptions on the periodic component. As in the previous sections, we can either adopt a multivariate setting, where we consider projections onto specific directions, or a fully functional approach.

The likelihood ratio statistic in the multivariate setting is the classical MANOVA test based on Wilk's lambda [see Mardia, Kent and Bibby (1979)], which is given as the ratio of the determinants of the empirical covariance under $\mathcal{H}_0$ in the numerator and of the empirical covariance under $\mathcal{H}_A$ in the denominator. Such an object is not easy to extend to the fully functional setting. If, however, we work with a fixed $\mathbf{\Sigma}$ (later it can be replaced by an estimator), then the LR statistic takes the form

$$T^{\mathrm{MAV}} = \frac{1}{d}\sum_{k=1}^{d} n(\overline{\boldsymbol{Y}}_k - \overline{\boldsymbol{Y}})'\mathbf{\Sigma}^{-1}(\overline{\boldsymbol{Y}}_k - \overline{\boldsymbol{Y}}), \tag{3.3}$$

where $\overline{\boldsymbol{Y}}_k = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{Y}_{(t-1)d+k}$, $1 \leq k \leq d$, and $\overline{\boldsymbol{Y}}$ is the grand mean. Translating this, with the same line of argumentation as in Section 3.2, into the fully functional setting we obtain

$$T^{\mathrm{FAV}} = \frac{1}{d}\sum_{k=1}^{d} n\|\overline{Y}_k - \overline{Y}\|^2, \tag{3.4}$$

where $\overline{Y}_k$ and $\overline{Y}$ are defined analogously. This formally coincides with the functional ANOVA test statistics considered in Cuevas, Febrero and Fraiman (2004) assuming a balanced design.

The following important result shows that the test statistics (3.3) and (3.4) are equivalent to T^{MTR_2} and T^{FTR_2}, respectively.

PROPOSITION 3.1. *It holds that* $T^{\mathrm{MAV}} = \frac{2}{d} T^{\mathrm{MTR}_2}$ *and* $T^{\mathrm{FAV}} = \frac{2}{d} T^{\mathrm{FTR}_2}$.

Proposition 3.1 is proven in Section B. We stress that the equalities in this result are of an algebraic nature, so they hold for any process $(Y_t : t \in \mathbb{Z})$. The limiting distribution of T^{FTR_2} with general stationary noise will follow from the theory developed in Section 4. Hence, we obtain the asymptotic null distribution of the functional ANOVA statistics T^{FAV} for stationary FTSs. This is formulated as Corollary 4.1. The result is of independent interest, as it relaxes the independence assumption in the functional ANOVA methodology.

4. Dependent non-Gaussian noise. In this section, we derive extensions of the testing procedures proposed in Section 3 to the setting of a general stationary noise sequence (Z_t); we drop the assumptions of Gaussianity and independence. We require that (Z_t) be a mean zero stationary sequence in H, which satisfies the following dependence assumption.

ASSUMPTION 4.1 (L^r–m-approximability). The sequence $(Z_t : t \in \mathbb{Z})$ can be represented as $Z_t = f(\delta_t, \delta_{t-1}, \delta_{t-2}, \ldots)$, where the δ_i's are i.i.d. elements taking values in some measurable space S and f is a measurable function $f : S^\infty \to H$. Moreover, if $\delta_1', \delta_2', \ldots$ are independent copies of $\delta_1, \delta_2, \ldots$ defined on the same measurable space S, then for

$$Z_t^{(m)} := f\big(\delta_t, \delta_{t-1}, \delta_{t-2}, \ldots, \delta_{t-m+1}, \delta_{t-m}', \delta_{t-m-1}', \ldots\big),$$

we have

$$\sum_{m=1}^{\infty} \big(E\|Z_m - Z_m^{(m)}\|^r\big)^{1/r} < \infty. \tag{4.1}$$

In the context of functional time series, the above assumption was introduced by Hörmann and Kokoszka (2010), and used in many subsequent papers including Hörmann, Horváth and Reeder (2013), Horváth, Kokoszka and Rice (2014), Zhang (2016), among many others. Similar conditions were used earlier by Wu (2005) and Shao and Wu (2007), to name representative publications. Of special note is the work of Lin and Liu (2009) who derive a test for the presence of a hidden periodicity (unknown d) for a scalar time series whose stationary component admits a Bernoulli shift representation used in Assumption 4.1. In the following, we will use this assumption with $r = 2$. The asymptotic theory could most likely be

developed under different weak dependence assumptions. The advantage of using Assumption 4.1 is that it has been verified for many functional time series models, and a number of asymptotic results exist, which we can use as components of the proofs.

Denote by $C_h = E(Z_h \otimes Z_0)$ the lag h autocovariance operator. If H is the space of square integrable functions, C_h is a kernel operator, $C_h : L^2 \to L^2$, which maps a function f to the function $C_h(f)(u) = \int E[Z_h(u)Z_0(s)]f(s)\,ds$. If Assumption 4.1 holds with $r = 2$, then

$$\sum_{h\in\mathbb{Z}} \|C_h\|_{\mathcal{S}} < \infty, \tag{4.2}$$

where $\|\cdot\|_{\mathcal{S}}$ denotes the Hilbert–Schmidt norm. As shown in Hörmann, Kidziński and Hallin (2015), this ensures the existence of the *spectral density operator*:

$$\mathcal{F}_\theta := \sum_{h\in\mathbb{Z}} C_h e^{-\mathrm{i}h\theta}.$$

This operator was defined in Panaretos and Tavakoli (2013a) (with an additional scaling factor $\frac{1}{2\pi}$). It plays a crucial role in frequency domain analysis of functional time series. We will see in Theorem 4.1 below that the spectral density operator is the asymptotic covariance operator of the discrete Fourier transform $D_N^Z(\theta)$, and hence it will enter the construction of our test statistics in a similar way as $\Gamma = \mathrm{Var}(Z_1)$ does in the case of independent noise. We recall hereby the definition of a complex-valued functional Gaussian random variable with mean μ, variance operator $F(v) = E[(X-\mu)\langle v, X-\mu\rangle]$ and relation operator $C(v) = E[(X-\mu)\langle v, \overline{X-\mu}\rangle]$. Then $Z = Z_0 + \mathrm{i}Z_1$ with $Z_0, Z_1 \in H$ is complex Gaussian $\mathcal{N}_H(\mu, F, C)$ if

$$\begin{pmatrix} Z_0 \\ Z_1 \end{pmatrix} \sim \mathcal{N}_{H\times H}\left(\begin{pmatrix}\mu_0\\ \mu_1\end{pmatrix}, \frac{1}{2}\begin{pmatrix} \mathrm{Re}(F+C) & -\mathrm{Im}(F-C) \\ \mathrm{Im}(F+C) & \mathrm{Re}(F-C)\end{pmatrix}\right),$$

where $\mu = \mu_0 + i\mu_1 = E(Z_0 + iZ_1) = \mu$. When the relation operator is null, we will write $Z \sim \mathcal{CN}_H(0, F)$. Theorem 4.1 follows from Theorem 5 in Cerovecki and Hörmann (2017).

THEOREM 4.1. *If (Z_t) is an $L^2 - m$–approximable time series with values in a separable Hilbert space H, then for any $\theta \in [-\pi, \pi]$*

$$D_N^Z(\theta) \xrightarrow{d} \mathcal{CN}_H(0, \mathcal{F}_\theta).$$

Furthermore:

(i) $\mathrm{Var}(D_N^Z(\theta))$ *converges in weak operator topology to* $\mathcal{F}_\theta$.

(ii) *The components of* $(D_N^Z(\theta), D_N^Z(\theta'))$ *are asymptotically independent whenever* $\theta + \theta' \neq 0$ *and* $\theta - \theta' \neq 0$.

Using Theorem 4.1, which is applicable to both functional and multivariate data, we are now ready to explain how to construct tests when Assumption 2.1 is dropped and replaced by $L^2 - m$-approximability. These tests, justified in Section C, have asymptotic (rather than exact) size α.

Independent noise: The tests of Section 3 remain unchanged for general i.i.d. noise with second order moments.

Projection based approach: If we project the functional data onto a basis $(v_1, \ldots, v_p)$, then the resulting multivariate time series $\boldsymbol{Y}_t$ inherits L^2–m-approximability. Let $\boldsymbol{\mathcal{F}}_\theta$ denote the spectral density matrix of this process. Assuming that the $\boldsymbol{\mathcal{F}}_{\vartheta_j}$ are of full rank, we need to replace the matrix

$$\boldsymbol{A}(\vartheta_1, \ldots, \vartheta_\ell)\boldsymbol{\Sigma}^{-1}\boldsymbol{A}'(\vartheta_1, \ldots, \vartheta_\ell), \qquad \ell = 1 \text{ or } \ell = q$$

in the definition of the multivariate tests by

$$\boldsymbol{H}(\vartheta_1, \ldots, \vartheta_\ell)\boldsymbol{H}'(\vartheta_1, \ldots, \vartheta_\ell),$$

where the columns of $\boldsymbol{H}'(\vartheta_1, \ldots, \vartheta_\ell)$ are given by

$$\big[\mathrm{Re}\big(\boldsymbol{\mathcal{F}}_{\vartheta_1}^{-1/2}\boldsymbol{D}(\vartheta_1), \ldots, \boldsymbol{\mathcal{F}}_{\vartheta_\ell}^{-1/2}\boldsymbol{D}(\vartheta_\ell)\big), \mathrm{Im}\big(\boldsymbol{\mathcal{F}}_{\vartheta_1}^{-1/2}\boldsymbol{D}(\vartheta_1), \ldots, \boldsymbol{\mathcal{F}}_{\vartheta_\ell}^{-1/2}\boldsymbol{D}(\vartheta_\ell)\big)\big].$$

The critical values remain the same as in Section 3.

Fully functional approach: In contrast to the multivariate setting the fully functional test statistics remain unchanged, but the critical values need to be adapted according to the following result.

PROPOSITION 4.1. *If (Y_t) is L^2–m–approximable then for any frequencies $0 < \omega_1 < \omega_2 < \cdots < \omega_\ell < \pi$,*

$$\big\|A(\omega_1, \ldots, \omega_\ell)A'(\omega_1, \ldots, \omega_\ell)\big\|_{\mathrm{tr}} \xrightarrow{d} \sum_{k=1}^{\ell} \Xi_k,$$

where $\Xi_k \overset{ind.}{\sim} \mathrm{HExp}(\lambda_1(\omega_k), \lambda_2(\omega_k), \ldots)$, and $\lambda_\ell(\omega_k)$ are the eigenvalues of $\mathcal{F}_{\omega_k}$.

In practice, we do not know the spectral densities which are necessary for our tests. In Section D, we show how to construct their estimators.

We conclude this section with a corollary to Proposition 4.1. This result is new and interesting in itself. It broadly extends the applicability of functional ANOVA by revealing its asymptotic distribution when the underlying data are weakly dependent.

COROLLARY 4.1. *Under the assumptions of Proposition 4.1, the functional ANOVA test statistic satisfies*

$$\frac{1}{d}\sum_{k=1}^{d} n\|\overline{Y}_k - \overline{Y}\|^2 \xrightarrow{d} \frac{2}{d}\sum_{k=1}^{q} \Xi_k,$$

where $\Xi_k \overset{ind.}{\sim} \mathrm{HExp}(\lambda_1(\vartheta_k), \lambda_2(\vartheta_k), \ldots)$, and $\lambda_\ell(\vartheta_k)$ are the eigenvalues of $\mathcal{F}_{\vartheta_k}$.

5. Consistency of the tests. In this section, we state consistency results for the tests developed in the previous sections. The proofs are presented in Section C. We focus on the general model (M.3) with the noise (Z_t) satisfying Assumption 4.1 with $r = 2$, but we also consider the simpler tests and alternatives introduced in Section 2. We assume throughout that Assumption 2.2 holds.

To state the consistency results, we decompose the DFT of the functional observations as follows:

$$D_N^Y(\theta) = D_N^w(\theta) + D_N^Z(\theta) = \sqrt{n} D_d^w(\theta) + D_N^Z(\theta), \tag{5.1}$$

where $D_N^w(\theta)$, $D_d^w(\theta)$ and $D_N^Z(\theta)$ are the DFT's of $(Y_1, \dots, Y_N)$, $(w_1, \dots, w_d)$ and $(Z_1, \dots, Z_N)$, respectively.

PROPOSITION 5.1. *Assume model* (M.3) *and that the noise* (Z_t) *is* L^2*–m-approximable. Then if* $\sum_{j=1}^q \|D_d^w(\vartheta_j)\|^2 > 0$, *we have that* $T^{\mathrm{FTR}_2} \to \infty$ *with probability* 1. *Moreover, if* $\|D_d^w(\vartheta_1)\|^2 > 0$, *we have that* $T^{\mathrm{FTR}_1} \to \infty$ *with probability* 1 $(N \to \infty)$.

Observe that

$$\sum_{j=1}^q \|D_d^w(\vartheta_j)\|^2 = \frac{1}{2}\sum_{t=1}^d \|w_t\|^2 =: \frac{d}{2}\mathrm{MSS}_{\mathrm{sig}}.$$

Explicit forms for $\|D_d^w(\vartheta_1)\|^2$ and $\sum_{j=1}^q \|D_d^w(\vartheta_j)\|^2$ when specialized to the alternatives considered in models (M.1), (M.2) and (M.3) are summarized in Table 1. We infer that if (Z_t) satisfies Assumption 4.1 with $r = 2$, then the functional tests based on T^{FTR_2} (or equivalently on T^{FAV}) are consistent under the alternatives in models (M.1), (M.2) and (M.3). The test based on T^{FTR_1} is consistent for model (M.1). It remains consistent for model (M.2) provided $\alpha_1^2 + \beta_1^2 > 0$, and it is consistent for model (M.3) if $\|D_d^w(\vartheta_1)\|^2 > 0$.

Consistency for the multivariate tests can be stated similarly. Consider the representation (3.1) of the projections.

TABLE 1
Explicit forms for $\|D_d^w(\vartheta_1)\|^2$ *and* $\sum_{j=1}^q \|D_d^w(\vartheta_j)\|^2$ *when specialized to the alternatives in models* (M.1), (M.2) *and* (M.3)

Model	$\|D_d^w(\vartheta_1)\|^2$	$\sum_{j=1}^q \|D_d^w(\vartheta_j)\|^2$
(M.1)	$\frac{d}{4}\rho^2$	$\frac{d}{4}\rho^2$
(M.2)	$\frac{d}{4}(\alpha_1^2 + \beta_1^2)$	$\frac{d}{4}\sum_{k=1}^d(\alpha_k^2 + \beta_k^2)$
(M.3)	$\|D_d^w(\vartheta_1)\|^2$	$\frac{d}{2}\mathrm{MSS}_{\mathrm{sig}}$

PROPOSITION 5.2. *Consider model* (M.3) *such that the noise is* L^2*–m-approximable. Let* $\boldsymbol{D}_d^{\boldsymbol{w}}(\theta) = \frac{1}{\sqrt{d}} \sum_{t=1}^{d} \boldsymbol{w}_t e^{-\mathrm{i}t\theta}$. *If* $\sum_{j=1}^{q} \|\boldsymbol{D}_d^{\boldsymbol{w}}(\vartheta_j)\|^2 > 0$, *we have that* $T^{\mathrm{MEV}_2} \to \infty$ *and* $T^{\mathrm{MTR}_2} \to \infty$ *with probability* 1. *If* $\|\boldsymbol{D}_d^{\boldsymbol{w}}(\vartheta_1)\|^2 > 0$, *we have that* $T^{\mathrm{MEV}_1} \to \infty$ *and* $T^{\mathrm{MTR}_1} \to \infty$ *with probability* 1 $(N \to \infty)$.

As before, we can specialize the result to models (M.1) and (M.2). Similar conditions as for the functional case are needed.

In Section E, we will present some results on local consistency, that is, we consider the case where the periodic signal shrinks to zero with growing sample size. This study gives some insight to the question in which situations a particular test can be recommended. In this context, we also refer to a numerical study in Section 8.

6. Application to pollution data. We analyze measurements of `PM10` (particulate matter) and `NO` (nitrogen monoxide) in Graz, Austria, collected during one cold season, between October 1, 2015, and March 15, 2016. The measurement unit for both pollutants is $\mu g/m^3$. Due to the geographic location of Graz in a basin and unfavorable meteorological conditions (temperature inversion), the EU air quality standards are often not met during the winter months. The measurement station we consider is in the city center (Graz-Mitte). Observations are available in 30 minute resolution. The data were preprocessed in order to account for a few missing values. To improve the stability of our L^2 based methodology, we follow Stadlober, Hörmann and Pfeiler (2008) and base our investigations on the square-root transformed data. The resulting discrete sample has been transformed into functional data objects with the `fda` package in `R` using nine B-spline basis functions of order four.

Our preliminary analysis, referred to in the Introduction, was based on standard ANOVA for daily averages, not taking into account the dependence of the data. Viewing them as projections onto $v(u) \equiv 1$, we can apply our tests T^{MEV_1} and T^{MEV_2} (or equivalently T^{MTR_1} and T^{MTR_2} since $p = 1$) adjusted for dependence as explained in Section 4. The spectral density of the daily averages is obtained as explained in Section D. The corresponding p-values are given in Tables 2 and 3 in the rows tagged $v(u) \equiv 1$. While there is still no indication of a weekly period in the `PM10` data, the test T^{MEV_2} is rejecting $\mathcal{H}_0$ for daily `NO`-averages at a 5% significance level.

Next, we conduct a periodicity analysis using our projection based and fully functional (FF) tests. For the projections, we use the first p PCA basis curves ($p \in \{1, 2, 3, 5\}$). We adjust the procedures for dependence, as explained in Section 4. The spectral density and the covariance operator (the latter is needed to compute principal components) are estimated as described in Section D. The results are presented in Tables 2 and 3. It can be seen that the fully functional ANOVA procedures do give strong evidence of a weekly pattern for `PM10` and

TABLE 2
The p-values for PM10 *data. In parentheses, the percentage of variance explained by the first p principal components on which the curves are projected*

	T^{MEV_1}	T^{MTR_1}	T^{MEV_2}	T^{MTR_2}	T^{FTR_1}	T^{FTR_2}
FF					0.081	0.026
$v(u) \equiv 1$	0.466	0.466	0.261	0.261		
$p = 1$ (71%)	0.397	0.397	0.216	0.216		
$p = 2$ (82%)	0.027	0.031	0.004	0.003		
$p = 3$ (88%)	0.031	0.030	0.010	0.005		
$p = 5$ (96%)	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$		

NO. When $p \geq 3$, all multivariate tests also give strong evidence for a weekly period. It is likely that the periodic component is mainly concentrated in the second (PM10) or third (NO) principal component. In contrast to the ANOVA not taking into account dependence of the NO data (see the Introduction), we now observe some significant weekly period also for the daily averages. In general, the p-values suggest that we are facing one of the alternatives described by model (M.2) or (M.3).

We extend this illustrating example by fitting a functional regression model to the PM10 curves:

$$\mathrm{PM}_t(u) = \int b_0(u, v)\mathrm{PM}_{t-1}(v)\,dv + \int b_1(u, v)\mathrm{NO}_t(v)\,dv + \varepsilon_t(u).$$

The kernel functions are estimated using B-spline expansions; see, for example, Ramsay, Hooker and Graves (2009). We analyze the residual curves $\hat{\varepsilon}_t(u)$. The p-values are summarized in Table 4. At 5% nominal level, none of our tests yields significant evidence that there remains a weekly periodicity in the residual curves.

TABLE 3
The p-values for NO *data. In parentheses, the percentage of variance explained by the first p principal components on which the curves are projected*

	T^{MEV_1}	T^{MTR_1}	T^{MEV_2}	T^{MTR_2}	T^{FTR_1}	T^{FTR_2}
FF					0.007	$<10^{-3}$
$v(u) \equiv 1$	0.161	0.161	0.016	0.016		
$p = 1$ (68%)	0.115	0.115	0.009	0.009		
$p = 2$ (81%)	0.331	0.331	0.029	0.017		
$p = 3$ (87%)	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$		
$p = 5$ (97%)	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$		

TABLE 4
The p-values for residuals of regression PM10 (*independent variables*) *onto first lag values and* NO *time series*

	T^{MEV_1}	T^{MTR_1}	T^{MEV_2}	T^{MTR_2}	T^{FTR_1}	T^{FTR_2}
FF					0.215	0.127
$p=1$	0.160	0.160	0.076	0.076		
$p=2$	0.277	0.316	0.200	0.272		
$p=3$	0.379	0.389	0.348	0.512		
$p=5$	0.673	0.674	0.678	0.740		

7. Assessment based on simulated data. Our goal is to assess empirical rejection rates of our tests in realistic finite sample settings. For this purpose, we consider the functional time series of PM10, preprocessed as explained in Section 6. We remove the weekday mean curves $\widehat{w}_k$, $1 \le k \le 7$, (from every Monday curve, we remove Monday's mean $\widehat{w}_1$, etc.). We then generate series of functional data by bootstrapping (with replacement) the times series of these residuals. The resulting i.i.d. data are denoted $\varepsilon_t, t = 1, \ldots, M$. Next, we generate dependent errors by setting

$$Z_t = \varepsilon_t + a_1\varepsilon_{t-i} + \cdots + a_5\varepsilon_{t-5}, \qquad t = 6, \ldots, M,$$

where $a_k = \varrho^k$ are scalar coefficients. We chose $M = 215$ and 425 so that the length of the time series, N, is 210 and 420. We consider $\varrho = 0.3$ and $\varrho = 0.6$. For the projection based tests, we project on the first p functional principal components, where $p \in \{1, 2, 3, 5\}$. Then we run our tests with the procedures adjusted for dependence as explained in Section 4. The empirical size, Table 5, is close to nominal level for the fully functional tests and the projection based tests when $p = 1, 2, 3$. The projection based tests with $p = 5$ are not well calibrated if $N = 215$, but improve considerably if we increase the sample size to $N = 420$ (last two rows in Table 5). The general tendency is that empirical rejection rates are slightly too big for $\varrho = 0.3$ and slightly too small for $\varrho = 0.6$. This observation remains true if we increase the sample size and suggests that estimation of the spectral density may need further tuning. We have experimented with other simulation setups, not reported here. For larger values of p, the fully functional tests seem more reliable than the multivariate tests in terms of their empirical size. This is most likely explained by the fact that the fully functional methods are not very sensitive to the effect of estimation errors for small eigenvalues. The distributions of $\mathrm{HExp}(\lambda_1, \lambda_2, \ldots)$ and $\mathrm{HExp}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots)$ are typically close because they mainly depend on a few large eigenvalues for which the relative estimation error is small. For the multivariate tests, eigenvalues enter as reciprocals. If λ_k is close to $\hat{\lambda}_k$, it does not necessarily mean that $1/\lambda_k$ and $1/\hat{\lambda}_k$ are close, if the eigenvalues are small.

TABLE 5

Empirical size (in %) at the nominal level α of 5% and 10% for dependent time series with sample size $N = 210$, $\varrho = 0.3$ (top rows) and $\varrho = 0.6$ (bottom rows). Results are based on 5,000 Monte Carlo simulation runs. The last two rows show the rates for $p = 5$ and $N = 420$

	$\alpha = 5\%$						$\alpha = 10\%$					
	$T^{\mathtt{MEV}_1}$	$T^{\mathtt{MTR}_1}$	$T^{\mathtt{MEV}_2}$	$T^{\mathtt{MTR}_2}$	$T^{\mathtt{FTR}_1}$	$T^{\mathtt{FTR}_2}$	$T^{\mathtt{MEV}_1}$	$T^{\mathtt{MTR}_1}$	$T^{\mathtt{MEV}_2}$	$T^{\mathtt{MTR}_2}$	$T^{\mathtt{FTR}_1}$	$T^{\mathtt{FTR}_2}$
FF					6.3	5.5					11.9	10.2
					4.2	3.6					8.6	7.8
$p = 1$	6.2	6.2	5.1	5.1			11.9	11.9	10.1	10.1		
	4.3	4.3	4.3	4.3			8.9	8.9	8.8	8.8		
$p = 2$	6.0	6.2	5.4	5.2			12.2	12.2	10.0	9.9		
	5.1	4.9	4.6	4.8			9.5	9.5	9.3	9.2		
$p = 3$	5.8	5.9	5.3	5.3			11.8	11.7	9.7	9.4		
	4.6	5.0	4.6	4.8			9.9	9.7	9.5	9.6		
$p = 5$	9.6	10.1	8.2	8.6			16.9	17.4	14.7	14.9		
	7.9	8.7	9.1	9.3			14.4	14.6	15.0	15.9		
$p = 5$	7.6	7.8	6.0	6.3			13.5	14.0	11.8	11.5		
	5.9	6.2	6.1	5.8			10.7	10.6	11.4	11.3		

To see how well the tests can detect a realistic alternative, we use the same data generating process as above and periodically add the weekday means $\widehat{w}_1, \dots \widehat{w}_7$ to the stationary noise (Z_t). We thus get the series $V_t = \widehat{w}_{(t)} + Z_t$ where $(t) = t \bmod 7$ with the convention that $\widehat{w}_0 = \widehat{w}_7$. This construction entails that we are in the setting of Model (M.3), and hence, in view of Theorem 3.1, we expect the multi-frequency and trace based tests to be most powerful. This is confirmed in Table 6 where we show empirical rejection rates. In this example, $\mathrm{MSS}_{\mathrm{sig}} = \frac{1}{7}\sum_{k=1}^{7} \|\widehat{w}_k - \overline{\widehat{w}}\|^2 \approx 0.1$ and $E\|Z_k\|^2 \approx 3.1$. Given the relatively small signal-to-noise ratio, we can see that overall the tests perform very well in finite samples.

The rejection rates reported in this section are based on a specific example and a specific estimator of the covariance structure, the same one as used in Section 6. To gain insights into the *asymptotic* rejection rates, we perform in Section 8 a numerical study which does not use a specific estimator, but assumes a known covariance structure. This approach allows us to isolate the effect of estimation from the intrinsic properties of the tests.

8. Local asymptotic power. A power study must necessarily involve a larger number of data generating processes (DGPs) which satisfy the various alternatives considered in this paper. We consider here 18 DGPs, indexed by the period $d = 7, 31$ and $i, j = 1, 2, 3$, which have the general form

$$Y_t(u) = s_t^{(i,d)}\left(\sum_{k=1}^{9} \psi_k^{(j)} v_k(u)\right) + \sum_{k=1}^{9} z_{t,k} v_k(u), \qquad i, j = 1, 2, 3. \tag{8.1}$$

The $v_1, v_2, \dots, v_9$ are orthonormal basis functions. We note right away that the results do not depend on what specific form the v_k take, as long as they are orthonormal. The $s_t^{(i,d)}$ is a real d-periodic signal with $\sum_{t=1}^{d} s_t^{(i,d)} = 0$ and $\psi_k^{(j)}$ are real coefficients. The exact specifications are given below. The variables $\boldsymbol{z_t} = (z_{t,1}, z_{t,2}, \dots, z_{t,9})'$ are i.i.d. Gaussian vectors with zero mean and covariance $\mathrm{diag}(1, 2^{-1}, 2^{-2}, 2^{-3}, \dots, 2^{-8})$. Then (Y_t) follows the functional model (M.2) with $w(u) = w^{(j)}(u) = \sum_{k=1}^{9} \psi_k^{(j)} v_k(u)$. Our assumptions imply that the v_k are the functional principal components of Y_t. We consider periods of length $d = 7$ and $d = 31$. For the periodic signal, we consider the following variants $s_t^{(i,d)}$ for $1 \le t \le d$:

$$\begin{aligned}
s_t^{(1,d)} &= \cos(2\pi t/d);\\
s_t^{(2,d)} &= I\{1 \le t \le 2(d-1)/3\} - 2I\{(2d-1)/3 + 1 < t \le d\};\\
s_t^{(3,d)} &= v_t - \bar{v} \qquad \text{where } v_t \overset{\text{i.i.d.}}{\sim} N(0,1).
\end{aligned}$$

TABLE 6

Empirical power (in %) when testing at nominal level α of 5%. We choose $\varrho = 0.6$. Results are based on 5000 Monte Carlo simulation runs

	$N = 210$						$N = 420$					
	$T^{\mathtt{MEV}_1}$	$T^{\mathtt{MTR}_1}$	$T^{\mathtt{MEV}_2}$	$T^{\mathtt{MTR}_2}$	$T^{\mathtt{FTR}_1}$	$T^{\mathtt{FTR}_2}$	$T^{\mathtt{MEV}_1}$	$T^{\mathtt{MTR}_1}$	$T^{\mathtt{MEV}_2}$	$T^{\mathtt{MTR}_2}$	$T^{\mathtt{FTR}_1}$	$T^{\mathtt{FTR}_2}$
FF					91.0	100					100	100
$p = 1$	25.5	25.5	77.7	77.7			51.1	51.1	98.6	98.6		
$p = 2$	84.2	84.9	100	100			99.3	99.3	100	100		
$p = 3$	96.1	97.0	100	100			99.9	100	100	100		
$p = 5$	100	100	100	100			100	100	100	100		

These functions are periodically extended to $t \geq 1$. We consider the following parameters $\boldsymbol{\psi}^{(j)} = (\psi_1^{(j)}, \ldots, \psi_9^{(j)})'$:

$$\boldsymbol{\psi}^{(1)} \propto (1, 0, 0, 0, 0, \ldots, 0)';$$
$$\boldsymbol{\psi}^{(2)} \propto \left(1, 2^{-1/2}, 2^{-1}, 2^{-3/2}, \ldots, 2^{-4}\right)';$$
$$\boldsymbol{\psi}^{(3)} \propto (0, 0, 0, 1, 0, \ldots, 0)'.$$

The vectors $\boldsymbol{\psi}^{(j)}$ determine $w(u)$ and are scaled such that the mean square sum of the signal $\mathrm{MSS}_{\mathrm{sig}} = 1$. Under parametrization $\boldsymbol{\psi}^{(1)}$ ($\boldsymbol{\psi}^{(3)}$), we have $w(u)$ varying in direction of the first (fourth) principal component, while under $\boldsymbol{\psi}^{(2)}$ we take into account all principal components. The DGP is determined by the pair $(\boldsymbol{\psi}, s) = (\boldsymbol{\psi}^{(j)}, s^{(i,d)})$.

We study the *local asymptotic power* functions defined by

$$LP(x|\boldsymbol{\psi}, s) = \lim_{N\to\infty} P\left(T_N > q_{0.95} \,\middle|\, \text{DGP is } \left(\boldsymbol{\psi}, \frac{x}{\sqrt{N/d}} s\right)\right),$$

where T_N stands for one of the test statistics we derived, and $q_{0.95}$ is its (asymptotic) 95th quantile under the null. The scaling is such that in any setup the total variation of the signal $\sum_{t=1}^{N} \|w_t\|^2 = x^2$. We use a superscript to indicate which statistic is used, for example, LP^{MEV_2}, LP^{FTR_1}, etc. It can be easily seen that if the covariance operator Γ is known, then due to our Gaussian setting, $P(T_N > q_{0.95} \mid \text{DGP is } (\boldsymbol{\psi}, \frac{x}{\sqrt{N/d}} s))$ does not depend on N for any of our tests. Since we let $N \to \infty$, we can use a Slutzky argument and compute $LP(x|\boldsymbol{\psi}, s)$ directly with the true Γ. It is not obvious how to obtain closed analytic forms for $LP(x|\boldsymbol{\psi}, s)$, and hence we compute them numerically by Monte-Carlo simulation based on 5000 replications.

1. *Comparing T^{MEV_2} and T^{MTR_2}: eigenvalue vs. trace based test statistic.*

We project data onto the space spanned by $v_1, \ldots, v_5$, which guarantees that at least 95% of variance are explained. In Figure 3, the asymptotic local power curves $LP^{\mathrm{MEV}_2}(x|\boldsymbol{\psi}^{(2)}, s^{(2,d)})$ and $LP^{\mathrm{MTR}_2}(x|\boldsymbol{\psi}^{(2)}, s^{(2,d)})$ with $d = 7$ and $d = 31$ are presented. We have done the same exercise with $\boldsymbol{\psi}^{(1)}$ and $\boldsymbol{\psi}^{(3)}$ and obtained very similar results.

2. *Comparing T^{MEV_1} and T^{MEV_2}: test for sinusoidal vs. test for general periodic pattern.*

We project again onto $v_1, \ldots, v_5$. In Figure 4, the asymptotic local power curves $LP^{\mathrm{MEV}_1}(x|\boldsymbol{\psi}^{(2)}, s)$ and $LP^{\mathrm{MEV}_2}(x|\boldsymbol{\psi}^{(2)}, s)$ are shown with $s = s^{(2,7)}$ (left panel), $s = s^{(2,31)}$ (middle panel) and $s = s^{(3,7)}$ (right panel). We see that the LR-test for the simpler model (M.1) can significantly outperform the LR-test for model (M.2) even if $s_t^{(2,d)}$ is not sinusoidal. However, the conclusion is very different if s is more erratic. When $s = s_t^{(3,7)}$, then T^{MEV_2} becomes a lot more powerful than T^{MEV_1}. Simulations not reported here show that the above described effects

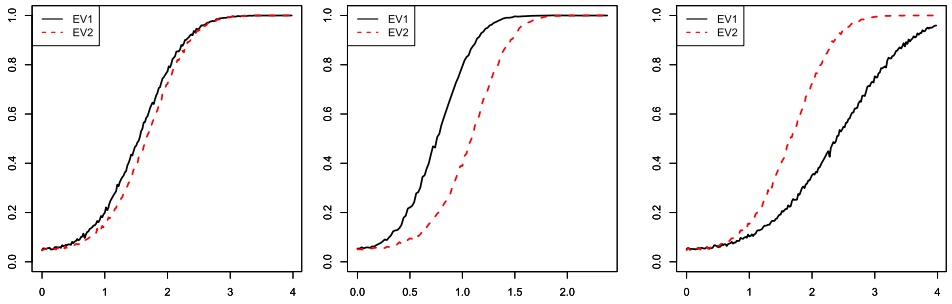

FIG. 4. *Local power curves* $LP^{\mathrm{MEV}_1}(x|\boldsymbol{\psi}^{(2)}, s)$ *(EV*1*) and* $LP^{\mathrm{MEV}_2}(x|\boldsymbol{\psi}^{(2)}, s)$ *(EV*2*) with* $s = s^{(2,7)}$ *(left panel),* $s = s^{(2,31)}$ *(middle panel) and* $s = s^{(3,7)}$ *(right panel), with the realization of* $(s_t^{(3,7)} : 1 \le t \le 7) = (-0.24, 0.42, -1.69, 0.37, 0.07, 1.12, -0.05)$.

become stronger the larger we choose the period d. This finding is supported by Proposition E.1, which provides a theoretical result on local consistency.

3. *Comparing* T^{MEV_1} *and* T^{FTR_1}*: projection based method vs. fully functional method.*

Now the objective is to compare the projection based methods with the fully functional ones. By fixing $s = s^{(1,7)}$, we focus on the simple model (M.1). The local power curves $LP^{\mathrm{FTR}_1}(x|\boldsymbol{\psi}^{(i)}, s^{(1,7)})$ and $LP^{\mathrm{MEV}_1}(x|\boldsymbol{\psi}^{(i)}, s^{(1,7)})$ for values $p = 1, 2, 3$ and $p = 5$ and $i = 1, 2, 3$ are shown in Figure 5. We see that the fully functional test performs well in all settings. Not surprisingly, the better the basis onto which we project describes $w(u)$, the better the projection based method becomes. For all DGPs $(\boldsymbol{\psi}^{(i)}, s^{(1,7)})$, $i = 1, 2, 3$, there is one projection based test that outperforms the functional one. The disadvantage of the projection method is, however, its sensitivity with respect to the chosen basis. For example, while for DGP $(\boldsymbol{\psi}^{(1)}, s^{(1,7)})$ the test with $p = 1$ is performing best, it is the least powerful for DGPs $(\boldsymbol{\psi}^{(2)}, s^{(1,7)})$ and $(\boldsymbol{\psi}^{(3)}, s^{(1,7)})$.

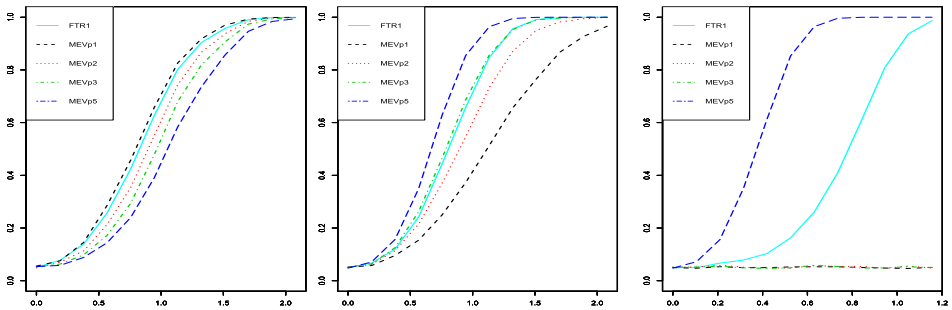

FIG. 5. *Local power curves* $LP^{\mathrm{FTR}_1}(x|\boldsymbol{\psi}^{(i)}, s^{(1,7)})$ *and* $LP^{\mathrm{MEV}_1}(x|\boldsymbol{\psi}^{(i)}, s^{(1,7)})$ *for values* $p = 1, 2, 3$ *and* $p = 5$ *and* $i = 1$ *(left panel),* $i = 2$ *(middle panel) and* $i = 3$ *(right panel).*

9. Summary. We have proposed several tests for the existence of a known period in functional time series. They fall into two broad categories, which we refer to as multivariate and fully functional approaches. The multivariate tests use projections on fixed or data-driven basis systems and the Gaussian likelihood ratio. The fully functional tests use L^2 norms. Our multivariate tests, in general, have the expected power advantage for multivariate time series for which other tests exist. Allowing general weak dependence of errors is also new even for multivariate data. For functional data, all tests are new. In what follows, we summarize the main conclusions of our work:

- Generally, the functional approach is a more adequate and safer option. The multivariate approach can be more powerful, but it is sensitive to the choice of the subspace on which the data are projected.
- If the signal is close to sinusoidal, then the simple single frequency test is more powerful, otherwise the opposite is true. The effect becomes stronger with length of the period. This empirical finding is theoretically confirmed in Section E of the Supplementary Material.
- For the multivariate tests, we have seen that the eigenvalue statistics can have a considerable power advantage over the traditionally used trace based statistics. Theoretically, we have shown that T^{MEV_1} and T^{MEV_2} can be justified by a LR procedure when the periodic signal is proportional to a single function $w(u)$. There exists an easy algorithm to compute critical values.

If no prior knowledge on the periodic component is available, we recommend using the ANOVA based approach or to base the decision on more than one test. Simultaneous acceptance or simultaneous rejection by several tests will lend confidence in the conclusion.

SUPPLEMENTARY MATERIAL

Supplemental material of "Testing for periodicity in functional time series" (DOI: 10.1214/17-AOS1645SUPP; .pdf). In Section A, we show how to apply our tests when Assumption 2.2 is not fulfilled. In Sections B and C, we prove the main theoretical results of this paper. In Section D, we discuss the estimation of the covariance and the spectral density matrix (operator) and we show their consistency also under the alternative. We also explain a data-driven manner to select the bandwidth for spectral smoothing. Finally, we study the rate of the approximation of the distribution $\mathrm{HExp}(\lambda_1, \lambda_2, \ldots)$ by $\mathrm{HExp}(\hat{\lambda}_1, \ldots, \hat{\lambda}_k)$. In Section E, we do a theoretically study for the local consistency under several alternatives. In Section F, we show that in typical situations (like when using PCA) the projection-based tests remain consistent if the basis is chosen in a data-driven way.

REFERENCES

AUE, A., NORINHO, D. D. and HÖRMANN, S. (2015). On the prediction of stationary functional time series. *J. Amer. Statist. Assoc.* **110** 378–392. MR3338510

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. Springer, New York. MR1093459

CEROVECKI, C. and HÖRMANN, S. (2017). On the CLT for discrete Fourier transforms of functional time series. *J. Multivariate Anal.* **154** 282–295. MR3588570

CHIANI, M. (2014). Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy–Widom distribution. *J. Multivariate Anal.* **129** 69–81. MR3215980

CUEVAS, A., FEBRERO, M. and FRAIMAN, R. (2004). An ANOVA test for functional data. *Comput. Statist. Data Anal.* **47** 111–122. MR2087932

FISHER, R. A. (1929). Tests of significance in harmonic analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **125** 54–59.

GROMENKO, O., KOKOSZKA, P. and REIMHERR, M. (2017). Detection of change in the spatiotemporal mean function. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 29–50.

HANNAN, E. J. (1961). Testing for a jump in the spectral function. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **23** 394–404.

HAYS, S., SHEN, H. and HUANG, J. Z. (2012). Functional dynamic factor models with application to yield curve forecasting. *Ann. Appl. Stat.* **6** 870–894. MR3012513

HÖRMANN, S., HORVÁTH, L. and REEDER, R. (2013). A functional version of the ARCH model. *Econometric Theory* **29** 267–288.

HÖRMANN, S., KIDZIŃSKI, L. and HALLIN, M. (2015). Dynamic functional principal components. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 319–348.

HÖRMANN, S. and KOKOSZKA, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884.

HÖRMANN, S., KOKOSZKA, P. and NISOL, G. (2018). Supplement to "Testing for periodicity in functional time series." DOI:10.1214/17-AOS1645SUPP.

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer.

HORVÁTH, L., KOKOSZKA, P. and RICE, G. (2014). Testing stationarity of functional time series. *J. Econometrics* **179** 66–82.

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis*, *with an Introduction to Linear Operators*. Wiley, Chichester. MR3379106

JENKINS, G. M. and PRIESTLEY, M. B. (1957). The spectral analysis of time-series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **19** 1–12 (discussion 47–63). MR0092329

KLEPSCH, J., KLÜPPELBERG, C. and WEI, T. (2017). Prediction of functional ARMA processes with an application to traffic data. *Econom. Stat.* **1** 128–149. MR3669993

LIN, Z. and LIU, W. (2009). On maxima of periodograms of stationary processes. *Ann. Statist.* **37** 2676–2695. MR2541443

MACNEILL, I. B. (1974). Tests for periodic components in multiple time series. *Biometrika* **61** 57–70. MR0378317

MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. Academic Press, London.

PANARETOS, V. M. and TAVAKOLI, S. (2013a). Fourier analysis of stationary time series in function space. *Ann. Statist.* **41** 568–603. MR3099114

PANARETOS, V. M. and TAVAKOLI, S. (2013b). Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Process. Appl.* **123** 2779–2807.

RAMSAY, J., HOOKER, G. and GRAVES, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer.

ROSS, S. M. (2010). *Introduction to Probability Models*. Elsevier, Amsterdam. MR3307945

SCHUSTER, A. (1898). On the investigation of hidden periodicities with application to the supposed 26 day period od meteorological phenomena. *Terr. Mag.* **3** 13–41.

SHAO, X. and WU, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. MR2351105

STADLOBER, E., HÖRMANN, S. and PFEILER, B. (2008). Quality and performance of a PM10 daily forecasting model. *Athmospheric Environment* **42** 1098–1109.

STADLOBER, E. and PFEILER, B. (2004). Explorative Analyse der Feinstaub-Konzentrationen von Oktober 2003 bis März 2004. Technical report, TU Graz.

WALKER, G. (1914). On the criteria for the reality of relationships or periodicities. *Calcutta Ind. Met. Memo* **21**.

WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215

ZAMANI, A., HAGHBIN, H. and SHISHEBOR, Z. (2016). Some tests for detecting cyclic behavior in functional time series with application in climate change. Technical report, Shiraz Univ.

ZHANG, X. (2016). White noise testing and model diagnostic checking for functional time series. *J. Econometrics* **194** 76–95. MR3523520

S. HÖRMANN
INSTITUTE OF STATISTICS
GRAZ UNIVERSITY OF TECHNOLOGY
KOPERNIKUSGASSE 24/III
A-8010 GRAZ
AUSTRIA
E-MAIL: shoermann@tugraz.at

P. KOKOSZKA
DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523-1877
USA
E-MAIL: piotr.kokoszka@colostate.edu

G. NISOL
ECARES
UNIVERSITÉ LIBRE DE BRUXELLES
50 AVENUE FRANKLIN ROOSEVELT
B-1050 BRUSSELS
BELGIUM
E-MAIL: gnisol@ulb.ac.be