JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **39**: 731–747 (2018)
Published online 2 May 2018 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12302

ORIGINAL ARTICLE

TESTING SEPARABILITY OF FUNCTIONAL TIME SERIES

PANAYIOTIS CONSTANTINOU,[a] PIOTR KOKOSZKA[b] AND MATTHEW REIMHERR[a]

[a] *Department of Statistics Pennsylvania State University University Park, PA, USA*
[b] *Department of Statistics Colorado State University Fort Collins, CO, USA*

We derive and study a significance test for determining whether a panel of functional time series is separable. In the context of this paper, separability means that the covariance structure factors into the product of two functions, one depending only on time and the other depending only on the coordinates of the panel. Separability is a property that can dramatically improve computational efficiency by substantially reducing model complexity. It is especially useful for functional data, as it implies that the functional principal components are the same for each member of the panel. However, such an assumption must be verified before proceeding with further inference. Our approach is based on functional norm differences and provides a test with well-controlled size and high power. We establish our procedure quite generally, allowing one to test separability of autocovariances as well. In addition to an asymptotic justification, our methodology is validated by a simulation study. It is applied to functional panels of particulate pollution and stock market data.

Received 02 May 2017; Accepted 23 March 2018

Keywords: Functional data; spatiotemporal separability; time series
MOS subject classification 62M10; 62M30; 62G10; 62G20.

1. INTRODUCTION

Suppose $\left\{X(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^2, t \in \mathbb{R}\right\}$ is a real-valued spatiotemporal random field, with the coordinate $\mathbf{s}$ referring to space and t to time. The field $X(\cdot, \cdot)$ is said to be *separable* if

$$\mathrm{Cov}(X(\mathbf{s}_1, t_1), X(\mathbf{s}_2, t_2)) = u(\mathbf{s}_1, \mathbf{s}_2)v(t_1, t_2),$$

where u and v are, respectively, spatial and temporal covariance functions. Separability is discussed in many textbooks, e.g. Cressie and Wikle (2015, Chap. 6). It has been extensively used in spatiotemporal statistics because it leads to theoretically tractable models and computationally feasible procedures; some recent references are Hoff (2011), Paul and Peng (2011), Sun *et al.* (2012). Before separability is assumed for the reasons noted above, it must be tested. Tests of separability are reviewed in Mitchell *et al.* (2005, 2006) and Fuentes (2006).

Time series of weather- or pollution-related measurements obtained at spatial locations typically exhibit strong periodic patterns. An approach to accommodate this periodicity is to divide the time series of such type into segments, each segment corresponding to a natural period. For example, a periodic time series of maximum daily temperatures at some location can be viewed as a stationary time series of functions, with one function per year. If the measurements are available at many locations $\mathbf{s}_k$, this gives rise to a data structure of the form

$$X_n(\mathbf{s}_k; t_i),\ k = 1, \ldots, S,\ i = 1, \ldots, I(= 365),\ n = 1, \ldots, N,$$

* Correspondence to: Matthew Reimherr, Department of Statistics, Pennsylvania State University, 411 Thomas Building, University Park, PA 16802, USA. E-mail: mreimherr@psu.edu

where n indexes the year, and t_i the day within a year. Time series of functions are discussed in several books, e.g. Bosq (2000), Horváth and Kokoszka (2012), Kokoszka and Reimherr (2017), but research on spatial fields or panels of time series of functions is relatively new, e.g. Kokoszka *et al.* (2016), Gromenko *et al.* (2016, 2017), French *et al.* (2016), Tupper *et al.* (2017), Liu *et al.* (2017), and Shang and Hyndman (2017). Testing separability of spatiotemporal functional data of the above form is investigated in Constantinou *et al.* (2017), Aston *et al.* (2017), and Bagchi and Dette (2017) under the assumption that the fields $X_n(\cdot,\cdot)$, $1 \le n \le N$, are independent. No tests are currently available for testing separability in the presence of temporal dependence across n. In a broader setting, the data that motivate this research have the form of functional panels:

$$\boldsymbol{X}_n(t) = [X_{n1}(t), X_{n2}(t), ..., X_{nS}(t)]^{\mathrm{T}}, \quad 1 \le n \le N. \tag{1}$$

Each $X_{ns}(\cdot)$ is a curve, and all curves are defined on the same time interval. The index n typically stands for day, week, month, or year. For instance, $X_{ns}(t)$ can be the exchange rate (against the Euro or the US Dollar) of currency s at minute t of the nth trading day, or $X_{ns}(t)$ can be the stock price of company s at minute t of the nth trading day. Another extensively studied example is daily or monthly yield curves for a panel of countries, e.g. Ang and Bekaert (2002), Bowsher and Meeks (2008), Hays *et al.* (2012), Kowal *et al.* (2017), among others. As for scalar data, the assumption of separability has numerous benefits including a simpler covariance structure, increased estimation accuracy, and faster computational times. In addition, in the contexts of functional time series, separability implies that the optimal functions used for temporal dimension reduction are the same for each member (coordinate) of the panel; information can then be pooled across the coordinates to get better estimates of these functions. We elaborate on this point in the following. However, if separability is incorrectly assumed, it leads to serious biases and misleading conclusions. A significance test, which accounts for the temporal dependence present in all the examples listed above, is therefore called for. The derivation of such a test, as well as the examination of its properties, is the purpose of this work. Our procedure is also applicable to testing separability of the autocovariance at any lag. We will demonstrate that it works well in situations where the tests of Constantinou *et al.* (2017) and Aston *et al.* (2017) fail.

The remainder of the paper is organized as follows. In Section 2, we formulate the assumptions, the definitions, and the problem. In Section 3, we derive the test and provide the required asymptotic theory. Section 4 focuses on details of the implementation. In Section 5, we present results of a simulation study, and, finally, in Section 6 we apply our procedure to functional panels of nitrogen dioxide levels on the east coast of the United States and to U.S. stock market data.

2. ASSUMPTIONS AND PROBLEM FORMULATION

We assume that $\boldsymbol{X}_n$ in (1) form a strictly stationary functional time series of dimension S. To simplify notation, we assume that all functions are defined on the unit interval $[0, 1]$ (integrals without limits indicate integration over $[0, 1]$). We assume that they are square-integrable in the sense that $E\|X_{ns}\|^2 = E\int X_{ns}^2(t)\,\mathrm{d}t < \infty$. Stationarity implies that the *lagged* covariance function can be expressed as

$$\mathrm{Cov}(X_{ns}(t), X_{n+h,s'}(t')) = c^{(h)}(s, t, s', t').$$

We aim to test the null hypothesis for a fixed value of h. The most important setting is when $h = 0$, i.e. testing separability of the covariance function, but other lags can be considered as well.

$$H_0 : \ c^{(h)}(s, t, s', t') = c_1^{(h)}(s, s')c_2^{(h)}(t, t'), \quad s, s' \in \{1, 2, \dots, S\};\ t, t' \in [0, 1]. \tag{2}$$

To derive the asymptotic distribution of our test statistic, we impose a weak dependence condition on $\boldsymbol{X}_n$. We use the concept of L^p–m-approximability introduced in Hörmann and Kokoszka (2010), see also Chapter 16 of Horváth

DOI: 10.1111/jtsa.12302

and Kokoszka (2012). Suppose $\mathbb{H}$ is a separable Hilbert space. Let $p \geq 1$ and let $L^p_{\mathbb{H}}$ be the space of $\mathbb{H}$-valued random elements X such that

$$\nu_p(X) = \left(E\|X\|^p\right)^{1/p} < \infty.$$

Definition 1. The sequence $\{Z_n\}$, $Z_n \in L^p_{\mathbb{H}}$ is *L^p–m-approximable* if the following conditions hold:

1. There exists a sequence $\{u_n\}$ of i.i.d. elements in an abstract measurable space $\mathcal{U}$ such that

$$Z_n = f(u_n, u_{n-1}, \dots),$$

for a measurable function $f : \mathcal{U}^\infty \to \mathbb{H}$;

2. For each integer $M > 0$, consider an approximating sequence $Z_{n,M}$ defined by

$$Z_{n,M} = f(u_n, u_{n-1}, \dots, u_{n-M}, u^\star_{n-M-1}, u^\star_{n-M-2}, \dots),$$

where the sequences $\{u^\star_n\} = \{u^\star_n(n,m)\}$ are copies of $\{u_n\}$ independent across m and n and independent of the original sequence $\{u_n\}$. We assume that $Z_{n,M}$ well approximates Z_n in the sense that

$$\sum_{M=1}^{\infty} \nu_p(Z_n - Z_{n,M}) < \infty. \tag{3}$$

Condition 1 of Definition 1 implies that the sequence is strictly stationarity and ergodic. The essence of Condition 2 is that the dependence of f on the innovations far in the past decays so fast that these innovations can be replaced by their independent copies. Such a replacement is asymptotically negligible in the sense quantified by (3). Similar conditions, which replace the more restrictive assumption of a linear moving average with summability conditions on its coefficients, have been used for at least a decade, see e.g. Shao and Wu (2007) and references therein. We work with Definition 1, as it is satisfied by most time series models, including functional time series, and provides a number of desirable asymptotic properties including the central limit theorem, see Chapter 16 of Horváth and Kokoszka (2012) and Kokoszka and Reimherr (2013a), among many other references. The conditions in Definition 1 cannot be verified; they are analogous to mixing or summability of cumulants conditions that have been imposed in theoretical time series analysis research. We therefore make the following assumption:

Assumption 1. The $\boldsymbol{X}_n$ form an L^4–m-approximable sequence in $\mathbb{H} = (L^2([0,1]))^S$.

We use tensor notation analogous to Aston *et al.* (2017). Let $\mathbb{H}_1$ and $\mathbb{H}_2$ denote two real separable Hilbert spaces with bases $\{u_i\}$ and $\{v_j\}$ respectively. We define $\mathbb{H} = \mathbb{H}_1 \otimes \mathbb{H}_2$ to be the tensor product Hilbert space. The tensors $\{u_i \otimes v_j\}$ form a basis for $\mathbb{H}$. In other words, the tensor product Hilbert space can be obtained by completing of the set span$\{u_i \otimes v_j \; : \; i = 1,\dots \; j = 1,\dots\}$, under the following inner product:

$$\langle u_i \otimes v_j, u_k \otimes v_\ell \rangle = \langle u_i, u_k\rangle\langle v_j, v_\ell\rangle, \quad u_i, u_k \in \mathbb{H}_1, v_j, v_\ell \in \mathbb{H}_2.$$

In the context of our study, $\mathbb{H}_1 = \mathbb{R}^S$ and $\mathbb{H}_2 = L^2([0,1])$. Therefore, the tensor product Hilbert space in our context is $\mathbb{H} = \mathbb{H}_1 \otimes \mathbb{H}_2 = \mathbb{R}^S \otimes L^2([0,1]) = (L^2([0,1]))^S =: L_2^S$, where we omit $[0,1]$ for simplicity. Each $\boldsymbol{X}_n$ is thus an element of a tensor space formed by the tensor product between two real separable Hilbert spaces, $\boldsymbol{X}_n \in \mathbb{H}_1 \otimes \mathbb{H}_2$. We denote by $\mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2)$ the space of Hilbert–Schmidt operators acting on $\mathbb{H}_1 \otimes \mathbb{H}_2$. Note that

$\{u_i \otimes v_j \otimes u_k \otimes v_\ell\}$ is a basis for $\mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2)$. The covariance operator between $\boldsymbol{X}_n$ and $\boldsymbol{X}_{n+h} \in \mathbb{H} = \mathbb{H}_1 \otimes \mathbb{H}_2$, $C^{(h)} = \mathrm{E}[\boldsymbol{X}_n \otimes \boldsymbol{X}_{n+h}] \in \mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2)$, is called separable if

$$C^{(h)} = C_1^{(h)} \widetilde{\otimes} C_2^{(h)}, \tag{4}$$

where $C_1^{(h)}$ is a covariance operator over $\mathbb{H}_1$ and $C_2^{(h)}$ is a covariance operator over $\mathbb{H}_2$. We define $C_1^{(h)} \widetilde{\otimes} C_2^{(h)}$ as a linear operator on $\mathbb{H}_1 \otimes \mathbb{H}_2$ satisfying

$$(C_1^{(h)} \widetilde{\otimes} C_2^{(h)})(u \otimes v) = (C_1^{(h)} u) \otimes (C_2^{(h)} v), \quad \forall u \in \mathbb{H}_1, \forall v \in \mathbb{H}_2.$$

The covariance operator between $\boldsymbol{X}_n$ and $\boldsymbol{X}_{n+h} \in L_2^S$ is in $\mathcal{S}(L_2^S)$, i.e. it is an integral operator with the kernel $c^{(h)}$. Relation (4) is then equivalent to H_0 stated as (2) above.

3. DERIVATION OF THE TEST AND ITS ASYMPTOTIC JUSTIFICATION

To test hypothesis (4), we propose a statistic that quantifies the difference between $\widehat{C}_1^{(h)} \widetilde{\otimes} \widehat{C}_2^{(h)}$ and $\widehat{C}^{(h)}$:

$$\widehat{T} = N \| \widehat{C}_1^{(h)} \widetilde{\otimes} \widehat{C}_2^{(h)} - \widehat{C}^{(h)} \|_{\mathcal{S}}^2, \tag{5}$$

where $\widehat{C}_1^{(h)}, \widehat{C}_2^{(h)}, \widehat{C}^{(h)}$ are estimates defined below, and $\| \cdot \|_{\mathcal{S}}$ is the Hilbert–Schmidt norm. The statistic (5) is a normalized distance between the estimator valid under the restriction imposed by H_0 and a general unrestricted estimator. The term $\widehat{C}_1^{(h)} \widetilde{\otimes} \widehat{C}_2^{(h)}$ is an estimator of the product $c_1^{(h)}(\cdot,\cdot) c_2^{(h)}(\cdot,\cdot)$ in (2) (the autocovariance under separability), whereas $\widehat{C}^{(h)}$ is an estimator of the unrestricted spatiotemporal autocovariance function $c^{(h)}(\cdot,\cdot,\cdot,\cdot)$. While $\widehat{C}^{(h)}$ is not difficult to define, it is not obvious how to define $\widehat{C}_1^{(h)}$ and $\widehat{C}_2^{(h)}$. This section explains how we define the estimators in (5) and what their joint asymptotic distribution is. This will allow us to derive the asymptotic properties of $\widehat{T}$.

The asymptotic null distribution involves the covariance operator of $\widehat{C}_1^{(h)} \widetilde{\otimes} \widehat{C}_2^{(h)} - \widehat{C}^{(h)}$, which we denote by $\mathbf{Q}^{(h)}$. Note that $\mathbf{Q}^{(h)} \in \mathcal{S}(\mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2))$, i.e. it is an operator acting on $\mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2)$. Therefore, it can be expanded using basis functions of the form $\{u_i \otimes v_j \otimes u_k \otimes v_\ell \otimes u_m \otimes v_n \otimes u_p \otimes v_q\}$. In the context of (1), $\mathbf{Q}^{(h)} \in \mathcal{S}(\mathcal{S}(L_2^S))$.

We now define the estimators appearing in (5) and obtain their limiting behavior even in the case where $C^{(h)}$ is not separable. A natural estimator for the general covariance, $C^{(h)}$, is given by

$$\widehat{C}^{(h)} = \frac{1}{N-h} \sum_{n=1}^{N-h} (\boldsymbol{X}_n - \hat{\boldsymbol{\mu}}) \otimes (\boldsymbol{X}_{n+h} - \hat{\boldsymbol{\mu}}) \in \mathcal{S}(L_2^S),$$

where $\boldsymbol{X}_n(t) = [X_{n1}(t), X_{n2}(t), ..., X_{nS}(t)]^{\mathrm{T}}$, $1 \le n \le N$, and $\hat{\boldsymbol{\mu}}(t) = [\hat{\mu}_1(t), \hat{\mu}_2(t), ..., \hat{\mu}_S(t)]^{\mathrm{T}}$ with $\hat{\mu}_s(t) = \frac{1}{N} \sum_{n=1}^{N} X_{ns}(t)$, $1 \le s \le S$. Since centering by the sample mean is asymptotically negligible, we assume, without loss of generality and to ease the notation, that our data are centered. So the estimator takes the form

$$\widehat{C}^{(h)} = \frac{1}{N-h} \sum_{n=1}^{N-h} \boldsymbol{X}_n \otimes \boldsymbol{X}_{n+h}, \tag{6}$$

or equivalently, the kernel of $\widehat{C}^{(h)}$ is

$$\hat{c}^{(h)}(s,t,s',t') = \frac{1}{N-h} \sum_{n=1}^{N-h} X_{ns}(t) X_{n+h,s'}(t').$$

DOI: 10.1111/jtsa.12302

Under H_0, $C^{(h)} = C_1^{(h)} \tilde{\otimes} C_2^{(h)}$ with $C_1^{(h)} \in \mathcal{S}(\mathbb{H}_1) = \mathcal{S}(\mathbb{R}^S)$, $C_2^{(h)} \in \mathcal{S}(\mathbb{H}_2) = \mathcal{S}(L^2([0,1]))$ and $C^{(h)} \in \mathcal{S}(\mathbb{H}) = \mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2) = \mathcal{S}(L_2^S)$. To obtain estimators for $C_1^{(h)}$ and $C_2^{(h)}$, we utilize the trace and the partial trace operators. For any trace-class operator T, see, e.g. Section 13.5 of Horváth and Kokoszka (2012) or Section 4.5 of Hsing and Eubank (2015), its trace is defined by

$$\mathrm{Tr}(T) := \sum_{i=1}^{\infty} \langle Te_i, e_i \rangle,$$

where $(e_i)_{i \geq 1}$ is an orthonormal basis. It is invariant with respect to the basis. The partial-trace operators are defined as

$$\mathrm{Tr}_1(A \tilde{\otimes} B) = \mathrm{Tr}(A)B, \quad A \in \mathbb{H}_1,\ B \in \mathbb{H}_2,$$

and

$$\mathrm{Tr}_2(A \tilde{\otimes} B) = \mathrm{Tr}(B)A, \quad A \in \mathbb{H}_1,\ B \in \mathbb{H}_2.$$

This means that Tr_1 and Tr_2 are bilinear forms that satisfy $\mathrm{Tr}_1 : \mathbb{H}_1 \otimes \mathbb{H}_2 \to \mathbb{H}_2$ and $\mathrm{Tr}_2 : \mathbb{H}_1 \otimes \mathbb{H}_2 \to \mathbb{H}_1$. In general, the trace of any element of $T \in \mathbb{H}_1 \otimes \mathbb{H}_2$ can be defined using proper basis expansions. More specifically, let $u_1, u_2, \ldots$ be an orthonormal basis for $\mathbb{H}_1$ and $v_1, v_2, \ldots$ an orthonormal basis for $\mathbb{H}_2$. Then a basis for $\mathbb{H}_1 \otimes \mathbb{H}_2$ is given by $\{u_i \otimes v_j : i = 1, 2, \ldots, j = 1, 2, \ldots\}$. Let $T : \mathbb{H}_1 \otimes \mathbb{H}_2 \to \mathbb{H}_1 \otimes \mathbb{H}_2$. Then, the trace of T is defined by

$$\mathrm{Tr}(T) = \sum_{i \geq 1} \sum_{j \geq 1} \langle T(u_i \otimes v_j), u_i \otimes v_j \rangle, \quad \mathrm{Tr} : \mathbb{H}_1 \otimes \mathbb{H}_2 \to \mathbb{R}.$$

If $T = A \tilde{\otimes} B$, the partial-trace operators in terms of a basis are defined as

$$\begin{aligned} \mathrm{Tr}_1(T) &= \mathrm{Tr}_1(A \tilde{\otimes} B) = \mathrm{Tr}(A)B = \sum_{i \geq 1} \langle Au_i, u_i \rangle B \\ &= \sum_{i \geq 1} \langle Au_i, u_i \rangle \sum_{j \geq 1} B_j v_j, \quad \forall A \in \mathbb{H}_1, \quad \forall B \in \mathbb{H}_2, \end{aligned}$$

and

$$\begin{aligned} \mathrm{Tr}_2(T) &= \mathrm{Tr}_2(A \tilde{\otimes} B) = \mathrm{Tr}(B)A = \sum_{j \geq 1} \langle Bv_j, v_j \rangle A \\ &= \sum_{j \geq 1} \langle Bv_j, v_j \rangle \sum_{i \geq 1} A_i u_i, \quad \forall A \in \mathbb{H}_1, \quad \forall B \in \mathbb{H}_2. \end{aligned}$$

In the context of functional panels, let $u_1, u_2, \ldots, u_S$ be an orthonormal basis for $\mathbb{R}^S$ and $v_1, v_2, \ldots$ an orthonormal basis for $L^2([0,1])$. Then a basis for L_2^S is given by $\{u_i \otimes v_j : i = 1, 2, \ldots, S, j = 1, 2, \ldots\}$. Recall that the products $u_i \otimes u_k$, viewed as operators, form a basis for $\mathcal{S}(\mathbb{R}^S)$, that is, a basis for the space of Hilbert–Schmidt operators acting on $\mathbb{R}^S$. Similarly, $\{v_j \otimes v_\ell\}$ is a basis for $\mathcal{S}(L^2([0,1]))$. Finally, $\{u_i \otimes v_j \otimes u_k \otimes v_\ell\}$ is a basis for $\mathcal{S}(L_2^S)$. The basis expansion of $C^{(h)}$ is given by

$$\sum_i \sum_j \sum_k \sum_\ell C_{ijk\ell}^{(h)} u_i \otimes v_j \otimes u_k \otimes v_\ell.$$

DOI: 10.1111/jtsa.12302

Therefore, its trace is given by

$$\mathrm{Tr}(C^{(h)}) = \sum_i \sum_j C^{(h)}_{ijij}.$$

Under the assumption of separability, i.e. $C^{(h)} = C_1^{(h)} \widetilde{\otimes} C_2^{(h)}$, the partial trace with respect to $\mathbb{H}_1$ in terms of a basis is given by

$$\mathrm{Tr}_1(C^{(h)}) = \mathrm{Tr}_1(C_1^{(h)} \widetilde{\otimes} C_2^{(h)}) = \mathrm{Tr}(C_1^{(h)})C_2^{(h)} = \sum_j \sum_\ell \left(\sum_i C^{(h)}_{iji\ell} \right) v_j \otimes v_\ell \quad \text{with} \quad C^{(h)}_{2,j\ell} = \sum_i C^{(h)}_{iji\ell},$$

and with respect to $\mathbb{H}_2$ by

$$\mathrm{Tr}_2(C^{(h)}) = \mathrm{Tr}_1(C_1^{(h)} \widetilde{\otimes} C_2^{(h)}) = \mathrm{Tr}(C_2^{(h)})C_1^{(h)} = \sum_i \sum_k \left(\sum_j C^{(h)}_{ijkj} \right) u_i \otimes u_k \quad \text{with} \quad C^{(h)}_{1,ik} = \sum_j C^{(h)}_{ijkj}.$$

Under the assumption of separability, we define estimators of $C_1^{(h)}$ and $C_2^{(h)}$ as

$$\widehat{C}_1^{(h)} = \frac{1}{\mathrm{Tr}(\widehat{C}^{(h)})} \mathrm{Tr}_2(\widehat{C}^{(h)}) \quad \text{and} \quad \widehat{C}_2^{(h)} = \mathrm{Tr}_1(\widehat{C}^{(h)}), \tag{7}$$

where $\widehat{C}_1^{(h)}$ is an $S \times S$ matrix and $\widehat{C}_2^{(h)}$ is a temporal covariance operator. The intuition behind the above estimators is that $\mathrm{Tr}(C^{(h)})C^{(h)} = \mathrm{Tr}_2(C^{(h)}) \widetilde{\otimes} \mathrm{Tr}_1(C^{(h)})$. Note that the decomposition $C^{(h)} = C_1^{(h)} \widetilde{\otimes} C_2^{(h)}$ is not unique since $C_1^{(h)} \widetilde{\otimes} C_2^{(h)} = (\alpha C_1^{(h)}) \widetilde{\otimes} (\alpha^{-1} C_2^{(h)})$ for any $\alpha \neq 0$; however, the product $C_1^{(h)} \widetilde{\otimes} C_2^{(h)}$ is.

To derive the asymptotic distribution of the test statistic $\widehat{T}$ defined in (5), we must first derive the joint asymptotic distribution of $\widehat{C}^{(h)}, \widehat{C}_1^{(h)}$, and $\widehat{C}_2^{(h)}$. A similar strategy was used in Constantinou *et al.* (2017). However, there the observations were assumed to be independent, and more traditional likelihood methods were used to derive the asymptotic distributions. Here, we take a different approach instead, using the CLT for $\widehat{C}^{(h)}$, and then leveraging a Taylor expansion over Hilbert spaces to obtain the joint asymptotic distribution of $\widehat{C}^{(h)}, \widehat{C}_1^{(h)}, \widehat{C}_2^{(h)}$. In this way, we are able to relax both the independence and Gaussian assumptions from Constantinou *et al.* (2017). The result is provided in Theorem 1. Because of the temporal dependence, the covariance operator of the limit normal distribution is a suitably defined long-run covariance operator. It has a very complex, but explicit and computable, form, which is displayed in Supporting Information, where all theorems that follow are also proven.

Recall that we are interested in testing

$$H_0 : C^{(h)} = C_1^{(h)} \widetilde{\otimes} C_2^{(h)} \quad \text{vs.} \quad H_A : C^{(h)} \neq C_1^{(h)} \widetilde{\otimes} C_2^{(h)}.$$

In the following theorems, notice that Theorems 1 and 2 hold without the assumption of separability, i.e. they hold under H_0 and under H_A. These two theorems are used to establish the behavior of our test statistic under both the null, Theorem 3, and the alternative, Theorem 4. Under the alternative, both $C_1^{(h)}$ and $C_2^{(h)}$ are still defined as partial traces of C; it is just that their tensor product no longer recovers the original $C^{(h)}$. Before we state our theoretical results, we mention the asymptotic distribution of $\widehat{C}^{(h)}$, which is the key to proof Theorem 1. It follows from Theorem 3 of Kokoszka and Reimherr (2013a) that, under Assumption 1

$$\sqrt{N}(\widehat{C}^{(h)} - C^{(h)}) \xrightarrow{\mathcal{L}} N(0, \mathbf{\Gamma}^{(h)}),$$

DOI: 10.1111/jtsa.12302

where $\boldsymbol{\Gamma}^{(h)}$ is given by

$$\boldsymbol{\Gamma}^{(h)} = \mathbf{R}_0^{(h)} + \sum_{i=1}^{\infty}[\mathbf{R}_i^{(h)} + (\mathbf{R}_i^{(h)})^*] \quad \text{with } \mathbf{R}_i^{(h)} = \mathrm{E}[(\boldsymbol{X}_1 \otimes \boldsymbol{X}_{1+h} - C^{(h)}) \otimes (\boldsymbol{X}_{1+i} \otimes \boldsymbol{X}_{1+i+h} - C^{(h)})]. \tag{8}$$

Here, $(\mathbf{R}_i^{(h)})^*$ denotes the adjoint of $\mathbf{R}_i^{(h)}$. Since we have the asymptotic distribution of $\widehat{C}^{(h)}$, and recalling that $\widehat{C}_1^{(h)}$ and $\widehat{C}_2^{(h)}$ are functions of $\widehat{C}^{(h)}$ from 7, we can use the Delta method to prove the following theorem; details of the proof of Theorem 1 are given in Section A of Supporting Information.

Theorem 1. Under Assumption 1, one can explicitly define a long-run covariance operator $\boldsymbol{W}^{(h)}$ such that

$$\sqrt{N}\begin{pmatrix} \widehat{C}_1^{(h)} - C_1^{(h)} \\ \widehat{C}_2^{(h)} - C_2^{(h)} \\ \widehat{C}^{(h)} - C^{(h)} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathrm{N}(\mathbf{0}, \boldsymbol{W}^{(h)}).$$

The definition of $\boldsymbol{W}^{(h)}$ is given in (A.2) of Supporting Information.

Armed with Theorem 1, we can derive the asymptotic distribution of $\widehat{C}_1^{(h)}\widetilde{\otimes}\widehat{C}_2^{(h)} - \widehat{C}^{(h)}$.

Theorem 2. Under Assumption 1

$$\sqrt{N}((\widehat{C}_1^{(h)}\widetilde{\otimes}\widehat{C}_2^{(h)} - \widehat{C}^{(h)}) - (C_1^{(h)}\widetilde{\otimes}C_2^{(h)} - C^{(h)})) \xrightarrow{\mathcal{L}} N(0, \boldsymbol{Q}^{(h)}).$$

The covariance operator $\boldsymbol{Q}^{(h)} \in \mathcal{S}(\mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2))$ is defined in (A.7) of Supporting Information.

As a corollary, we obtain the asymptotic distribution of $\widehat{C}_1^{(h)}\widetilde{\otimes}\widehat{C}_2^{(h)} - \widehat{C}^{(h)}$ under H_0.

Corollary 1. Suppose Assumption 1 holds. Then, under H_0

$$\sqrt{N}(\widehat{C}_1^{(h)}\widetilde{\otimes}\widehat{C}_2^{(h)} - \widehat{C}^{(h)}) \xrightarrow{\mathcal{L}} N(0, \boldsymbol{Q}^{(h)}),$$

where the covariance operator $\boldsymbol{Q}^{(h)}$ is the same as in Theorem 2.

As noted above, in the context of (1), $\mathbf{Q}^{(h)} \in \mathcal{S}(\mathcal{S}(L_2^S))$, i.e. it is a Hilbert–Schmidt operator acting on a space of Hilbert–Schmidt operators over L_2^S. The following result is a direct consequence of Theorem 2. While the weighted chi-square expansion is standard, to compute the weights the operator $\mathbf{Q}^{(h)}$ must be estimated, so $\boldsymbol{W}^{(h)}$ must be estimated. Formula (A.2) defining $\boldsymbol{W}^{(h)}$ is new and nontrivial.

Theorem 3. Suppose Assumption 1 holds. Let $\mathbf{Q}^{(h)}$ be the covariance operator appearing in Theorem 2, whose eigenvalues are $\gamma_1, \gamma_2, \ldots$. Then, under H_0, as $N \to \infty$

$$\widehat{T} \xrightarrow{\mathcal{L}} \sum_{r=1}^{\infty} \gamma_r Z_r^2,$$

where the Z_r are i.i.d. standard normal.

DOI: 10.1111/jtsa.12302

To describe the behavior of the test statistic under the alternative, some specific form of the alternative must be assumed, as the violation of (4) can take many forms. A natural approach corresponding to a fixed alternative to $C_1^{(h)} \widetilde{\otimes} C_2^{(h)} - C^{(h)} = 0$ is to assume that

$$C_1^{(h)} \widetilde{\otimes} C_2^{(h)} - C^{(h)} =: \Delta \neq 0. \tag{9}$$

Theorem 4. Suppose Assumption 1 holds. If (9) holds, then

$$\widehat{T} = N\|\Delta\|^2 + O_P(N^{1/2}) \xrightarrow{\mathcal{P}} \infty.$$

In our applications, $\boldsymbol{X}_n \in \mathbb{H}_1 \otimes \mathbb{H}_2$, where $\mathbb{H}_1 = \mathbb{R}^S$ and $\mathbb{H}_2 = L^2([0,1])$. Therefore, in practice, we must first project these random elements onto a truncated basis by using a dimension reduction procedure. Note that $\mathbb{H}_1 = \mathbb{R}^S$ is already finite. However, if the number of coordinates in the panel is large, then a dimension reduction in $\mathbb{H}_1 = \mathbb{R}^S$ is also recommended. Here we present the general case where we use dimension reduction in both $\mathbb{H}_1 = \mathbb{R}^S$ and $\mathbb{H}_2 = L^2([0,1])$. The truncated basis is of the form $\hat{u}_k \otimes \hat{v}_j$ with $1 \leq k \leq K,\ 1 \leq j \leq J$ where $K < S$ and $J < \infty$. In our implementation, $\hat{u}_k$ and $\hat{v}_j$ are the empirical principal components. We can approximate each $\boldsymbol{X}_n \in \mathbb{H}_1 \otimes \mathbb{H}_2$ by a $K \times J$ random matrix $\mathbf{Z}_n \in \mathbb{R}^{K\times J}$, where $\mathbf{Z}_n(k,j) = \langle \boldsymbol{X}_n, \hat{u}_k \otimes \hat{v}_j \rangle$, $1 \leq k \leq K,\ 1 \leq j \leq J$. Therefore, from now on, we work with observations in the form of random $K \times J$ matrices defined as

$$\mathbf{Z}_n = [z_{kj;n},\ 1 \leq k \leq K,\ 1 \leq j \leq J],$$

where $z_{kj;n} = \langle \boldsymbol{X}_n, \hat{u}_k \otimes \hat{v}_j \rangle$. Let $\widehat{T}_F$ be the truncated test statistic $\widehat{T}$, i.e.

$$\widehat{T}_F = N\|\widehat{C}_{1,K}^{(h)} \widetilde{\otimes} \widehat{C}_{2,J}^{(h)} - \widehat{C}_{KJ}^{(h)}\|_F^2,$$

where $\widehat{C}_{1,K}^{(h)}$ is a $K\times K$ matrix, $\widehat{C}_{2,J}^{(h)}$ is a $J\times J$ matrix, $\widehat{C}_{KJ}^{(h)}$ is a fourth-order array of dimension $K\times J\times K\times J$, and $\|\cdot\|_F$ is the Frobenius norm, which is the Hilbert–Schmidt norm in finite dimensions. Finally, let $\mathbf{Q}_{KJ}^{(h)}$ be the truncated covariance operator $\mathbf{Q}^{(h)}$, i.e. $\mathbf{Q}_{KJ}^{(h)}$ is the asymptotic covariance operator in the convergence

$$\sqrt{N}((\widehat{C}_{1,K}^{(h)} \widetilde{\otimes} \widehat{C}_{2,J}^{(h)} - \widehat{C}_{KJ}^{(h)}) - (C_{1,K}^{(h)} \widetilde{\otimes} C_{2,J}^{(h)} - C_{KJ}^{(h)})) \xrightarrow{\mathcal{L}} N(0, \boldsymbol{Q}_{KJ}^{(h)}).$$

Note that $\mathbf{Q}_{KJ}^{(h)}$ is an array of order eight with finite dimensions, $\mathbf{Q}_{KJ}^{(h)} \in \mathbb{R}^{K\times J\times K\times J\times K\times J\times K\times J}$. More details are given in Remark A.2 in Supporting Information. As a finite array, it has only a finite number of eigenvalues, which we denote $\gamma_1^\dagger, \gamma_2^\dagger, \dots, \gamma_R^\dagger$. The arguments leading to Theorem 3 show that under H_0, as $N \to \infty$

$$\widehat{T}_F \xrightarrow{\mathcal{L}} \sum_{r=1}^{R} \gamma_r^\dagger Z_r^2, \tag{10}$$

where the Z_r are i.i.d. standard normal. The asymptotic argument needed to establish (10) relies on the bounds $\|\hat{u}_k - u_k\| = O_P(N^{-1/2})$ and $\|\hat{v}_j - v_j\| = O_P(N^{-1/2})$, which hold under Assumption 1. It is similar to the technique used in the proof of Theorem 4 in Constantinou *et al.* (2017), so it is omitted.

4. DETAILS OF IMPLEMENTATION

Recall that we assume that all functions have been rescaled so that their domain is the unit interval $[0,1]$, and that they have mean zero. The testing procedure consists of dimension reduction in time and, for large panels,

DOI: 10.1111/jtsa.12302

a further dimension reduction in coordinates. After reducing the dimension, our "observations" are of the form of $K \times J$ matrices, which we use to compute the estimators we need to perform our test. The remainder of this section explains the details in an algorithmic form. The reader will notice that most steps have obvious variants, for example, different weights and bandwidths can be used in Step 6, and different percentages of explained variance, rather than 85%, which is merely a rule of thumb, may work better in different scenarios. Procedure 4.1 describes the exact implementation used in Sections 5 and 6.

Procedure 4.1.

1. [*Pool across s to get estimated temporal functional principal components (FPCs).*] Under the assumption of separability, i.e., under the H_0 stated in Section 2, the optimal functions used for temporal dimension reduction are the same for each member (coordinate) of the panel; information can then be pooled across the coordinates to get better estimates of these functions. In other words, under separability, we can use simultaneously all the $N \times S$ functions to compute the temporal FPCs $\hat{v}_1, \dots, \hat{v}_J$ as the eigenfunctions of the covariance function

$$\hat{c}_2(t, t') = \frac{1}{NS} \sum_{n=1}^{N} \sum_{s=1}^{S} X_{ns}(t) X_{ns}(t').$$

2. Approximate each curve $X_{ns}(t)$ by

$$X_{ns}^{(J)}(t) = \sum_{j=1}^{J} \xi_{nsj} \hat{v}_j(t),$$

where $\xi_{nsj} = \langle X_{ns}(t), \hat{v}_j(t) \rangle$. Construct $S \times J$ matrices $\boldsymbol{\Xi}_n$ defined as

$$\boldsymbol{\Xi}_n = [\xi_{nsj},\ 1 \le s \le S,\ 1 \le j \le J],$$

where J is chosen large enough so that the first J FPCs explain at least 85% of the variance. This is the FPC analysis carried out on the pooled (across coordinates) sample.

3. [*Pool across time to get panel PCs.*] Under the assumption of separability, the panel principal components are the same for each time. In other words, the panel PCs are the principal components of the following covariance matrix:

$$\hat{c}_1(s, s') = \frac{\sum_n^N \int X_{ns}(t) X_{ns'}(t)\, \mathrm{d}t}{N\, \mathrm{tr}(\widehat{C})}.$$

However, since we have already reduced the dimension of the observed functions, the panel PCs $\hat{u}_1, \dots, \hat{u}_K$ are the principal components of the covariance matrix

$$\tilde{c}_1(s, s') = \frac{1}{NJ} \sum_{n=1}^{N} \sum_{j=1}^{J} \frac{\xi_{nsj} \xi_{ns'j}}{\lambda_j}.$$

4. Approximate each row $\xi_{n \cdot j} = (\xi_{n1j}, \xi_{n2j}, \dots, \xi_{nSj})$ of the $\boldsymbol{\Xi}_n$ matrices by

$$\xi_{n \cdot j}^{(K)} = \sum_{k=1}^{K} z_{kj;n} \hat{u}_k, \quad z_{kj;n} = \langle \xi_{n \cdot j}, \hat{u}_k \rangle.$$

DOI: 10.1111/jtsa.12302

Construct the $K \times J$ matrices $\mathbf{Z}_n = [z_{kj;n},\ 1 \le k \le K,\ 1 \le j \le J]$, where K is chosen large enough so that the first K eigenvalues explain at least 85% of the variance. This is a multivariate principal components analysis (PCA) on the pooled (across time) variance-adjusted sample.

If the number of panel coordinates is small, then a multivariate dimension reduction is not necessary, so one can skip Steps 3 and 4 and use the $\boldsymbol{\Xi}_n$ matrices instead of the $\mathbf{Z}_n$ matrices, and replace K with S in the following steps: The dimension reduction steps reduce the computational time and the memory requirements by reducing the matrix size the 4D and 8D covariance tensors.

5. Approximate covariance (6) by the fourth-order array of dimensions $K \times J \times K \times J$

$$\widehat{C}^{(h)}_{KJ} = \frac{1}{N-h} \sum_{n=1}^{N-h} \mathbf{Z}_n \otimes \mathbf{Z}_{n+h}.$$

Approximate $\widehat{C}^{(h)}_1$ and $\widehat{C}^{(h)}_2$ in (7) by

$$\widehat{C}^{(h)}_{1,K}(k,k') = \frac{\sum_{j=1}^{J} \widehat{C}^{(h)}_{KJ}(k,j,k',j)}{\sum_{k=1}^{K} \sum_{j=1}^{J} C^{(h)}_{KJ}(k,j,k,j)} \quad \text{and} \quad \widehat{C}^{(h)}_{2,J}(j,j') = \sum_{k=1}^{K} \widehat{C}^{(h)}_{KJ}(k,j,k,j'),$$

where $\widehat{C}^{(h)}_{1,K}$ is a $K \times K$ matrix and $\widehat{C}^{(h)}_{2,J}$ is a $J \times J$ matrix.

6. Calculate the estimators $\widehat{\mathbf{R}}^{(h)}_{0,KJ}, \widehat{\mathbf{R}}^{(h)}_{i,KJ}, (\widehat{\mathbf{R}}^{(h)}_{i,KJ})^* \in \mathbb{R}^{K\times J\times K\times J\times K\times J\times K\times J}$, by using

$$\begin{aligned}
\widehat{\mathbf{R}}^{(h)}_{0,KJ} &= \frac{1}{N-h} \sum_{n=1}^{N-h} [(\mathbf{Z}_n \otimes \mathbf{Z}_{n+h} - \widehat{C}^{(h)}_{KJ}) \otimes (\mathbf{Z}_n \otimes \mathbf{Z}_{n+h} - \widehat{C}^{(h)}_{KJ})],\\
\widehat{\mathbf{R}}^{(h)}_{i,KJ} &= \frac{1}{N-i-h} \sum_{n=1}^{N-i-h} [(\mathbf{Z}_n \otimes \mathbf{Z}_{n+h} - \widehat{C}^{(h)}_{KJ}) \otimes (\mathbf{Z}_{n+i} \otimes \mathbf{Z}_{n+i+h} - \widehat{C}^{(h)}_{KJ})],\\
(\widehat{\mathbf{R}}^{(h)}_{i,KJ})^* &= \frac{1}{N-i-h} \sum_{n=1}^{N-i-h} [(\mathbf{Z}_{n+i} \otimes \mathbf{Z}_{n+i+h} - \widehat{C}^{(h)}_{KJ}) \otimes (\mathbf{Z}_n \otimes \mathbf{Z}_{n+h} - \widehat{C}^{(h)}_{KJ})].
\end{aligned} \tag{11}$$

7. Calculate the estimator $\widehat{\boldsymbol{\Gamma}}^{(h)}_{KJ} \in \mathbb{R}^{K\times J\times K\times J\times K\times J\times K\times J}$, by using the following Bartlett-type estimator:

$$\widehat{\boldsymbol{\Gamma}}^{(h)}_{KJ} = \widehat{\mathbf{R}}^{(h)}_{0,KJ} + \sum_{i=1}^{N-h-1} \omega_i (\widehat{\mathbf{R}}^{(h)}_{i,KJ} + (\widehat{\mathbf{R}}^{(h)}_{i,KJ})^*), \tag{12}$$

where $\widehat{\mathbf{R}}^{(h)}_{0,KJ}, \widehat{\mathbf{R}}^{(h)}_{i,KJ}, (\widehat{\mathbf{R}}^{(h)}_{i,KJ})^*$ are defined in 11 and ω_i are the Bartlett's weights, i.e.

$$\omega_i = \begin{cases} 1 - \frac{i}{1+q}, & \text{if } i \le q,\\ 0, & \text{otherwise,} \end{cases}$$

with i being the number of lags and q the bandwidth, which is assumed to be a function of the sample size, i.e. $q = q(N)$. In our simulations in Section 5, we use the formula $q \approx 1.1447(\frac{N}{4})^{1/3}$ (Horváth and Kokoszka 2012, Chapter 16).

DOI: 10.1111/jtsa.12302

Note that the estimators $\widehat{\mathbf{R}}^{(h)}_{0,KJ}$, $\widehat{\mathbf{R}}^{(h)}_{i,KJ}$, $(\widehat{\mathbf{R}}^{(h)}_{i,KJ})^*$, and $\widehat{\mathbf{\Gamma}}^{(h)}_{KJ}$ defined in Steps 6 and 7 are the truncated analogs of the estimators $\widehat{\mathbf{R}}^{(h)}_{0}$, $\widehat{\mathbf{R}}^{(h)}_{i}$, $(\widehat{\mathbf{R}}^{(h)}_{i})^*$, and $\widehat{\mathbf{\Gamma}}^{(h)}$, which can be obtained by simply changing $\boldsymbol{Z_n}$ with $\boldsymbol{X_n}$ in 11.

8. Estimate the arrays $\mathbf{W}^{(h)}_{KJ}$ (the truncated analog of $\mathbf{W}^{(h)}$) and $\mathbf{Q}^{(h)}_{KJ}$ defined in Section 3. Details are given in Remark A.2 in Supporting Information.

9. Calculate the p-value using the limit distribution specified in (10).

Step 2 can be easily implemented using `R` function `pca.fd`, and Step 3 by using `R` function `prcomp`. The matrix $\mathbf{Q}^{(h)}_{KJ}$ can be computed using the `R` package `tensorA` by van den Boogaart (2007).

5. A SIMULATION STUDY

The purpose of this section is to provide information on the performance of our test procedure in finite samples. We first comment on the performance of existing tests. Constantinou *et al.* (2017) derived several separability tests based on the assumption of independent $\boldsymbol{X}_n$. For the functional panels that exhibit temporal dependence (we define them below), the empirical sizes are close to zero; the tests of Constantinou *et al.* (2017) are too conservative to be usable unless we have independent replications of the spatiotemporal structure. Aston *et al.* (2017) proposed three tests, also for independent $\boldsymbol{X}_n$. In the presence of temporal dependence, their tests are not useable either; they can severely over-reject, and the empirical size can approach 50% at the nominal level of 5%. We give some specific numbers at the end of this section.

For our empirical study, we simulate functional panels as the moving average process

$$X_{ns}(t) = \sum_{s'=1}^{S} \Psi_{ss'}[e_{ns'}(t) + e_{n-1s'}(t)],$$

which is an 1-dependent functional time series. Direct verification shows that it is separable as long the $e_{ns}(t)$ are separable. We generate $e_{ns}(t)$ as Gaussian processes with the following covariance function, which is a modified version of Example 2 of Cressie and Huang (1999):

$$\sigma_{ss'}(t,t') = \frac{\sigma^2}{(a|t-t'|+1)^{1/2}} \exp\left(-\frac{b^2[|s-s'|/(S-1)]^2}{(a|t-t'|+1)^c}\right). \tag{13}$$

In this covariance function, a and b are nonnegative scaling parameters of time and space respectively, and $\sigma^2 > 0$ is an overall scale parameter. The most important parameter is the separability parameter c, which takes values in $[0,1]$. If $c = 0$, the covariance function is separable, otherwise it is not. We set $a = 3$, $b = 2$, and $\sigma^2 = 1$. To simulate the functions, we use $T = 50$ time points equally spaced on $[0,1]$, and $S \in \{4, 6, 8, 10, 12, 14\}$ coordinates in the panel. The MA coefficients are taken as follows:

$$\Psi_{ss'} = \exp\left(-\frac{25(s-s')^2}{(S-1)^2}\right).$$

Notice that in the covariance above, the differences in the coordinates of the panel, i.e. $|s-s'|$, are rescaled to be within the interval $[0,1]$, i.e. we use $|s-s'|/(S-1)$.

We set

$$c = 0 \text{ under } H_0; \quad c = 1 \text{ under } H_A.$$

We consider two different cases: the first one with dimension reduction only in time, and the second one with dimension reduction in both time and coordinates. For each case, we study two different scenarios. The first

DOI: 10.1111/jtsa.12302

scenario is under the null hypothesis (separability), and the second scenario under the alternative hypothesis. We consider different numbers of temporal FPCs, J, in the first case, and different numbers of coordinate PCs, K, and temporal FPCs, J, in the second case. We will also consider different values for the series length N. All empirical rejection rates are based on 1000 replications, so their standard deviation (SD) is about 0.7% for size (we use the nominal significance level of 5%) and about 2% for power.

5.1. Case 1: dimension reduction in time only

We examine the effect of the series length N and the number of principal components J on the empirical size (Table I) and power (Table II) for $S \in \{4, 6, 8\}$. Each table reports the rejection rates in percent. In parentheses, the proportion of variance explained by the J PCs is given.

From Table I, we can see that the size of our test is robust to the number of the principal components used. This is a very desirable property, as in all procedures of FDA there is some uncertainty about the optimal number of FPCs that should be used. While still within two standard errors of the nominal size, the empirical size becomes inflated for $S = 8$. We recommend dimension reduction in panel coordinates if $S \geq 10$. In Table II, we see that the empirical power increases as N and J increase. The power increase with N is expected; its increase with J reflects the fact that projections on larger subspaces better capture a departure from H_0. However, J cannot be chosen too large so as not to increase the dimensionality of the problem, which negatively affects the empirical size.

5.2. Case 2: dimension reduction in both time and panel coordinates

The general setting is the same as in Section 5.1, but we consider larger panels, $S \in \{10, 12, 14\}$, and reduce their dimension to $K \in \{2, 3, 4\}$ coordinates. The proportion of variance explained is now computed as

$$\mathrm{CPV}(J, K) = \frac{\sum_{j=1}^{J} \lambda_j}{\sum_{j=1}^{\infty} \lambda_j} \times \frac{\sum_{k=1}^{K} \mu_k}{\sum_{k=1}^{S} \mu_k}, \tag{14}$$

where the $\lambda_1, \lambda_2, \ldots$, and $\mu_1, \mu_2, \ldots, \mu_S$ are respectively the estimated eigenvalues of the time and panel PCAs.

Table I. Rejection rates under H_0 ($c = 0$) at the nominal 5% level

	$N = 100$			$N = 150$			$N = 200$		
	$J = 2$	$J = 3$	$J = 4$	$J = 2$	$J = 3$	$J = 4$	$J = 2$	$J = 3$	$J = 4$
$S = 4$	5.5	6.4	5.0	5.9	5.7	5.3	6.5	5.1	5.5
	(87%)	(90%)	(94%)	(85%)	(90%)	(92%)	(87%)	(90%)	(92%)
$S = 6$	5.6	5.9	5.3	6.2	5.3	4.7	5.6	6.2	5.1
	(85%)	(91%)	(93%)	(85%)	(91%)	(93%)	(86%)	(91%)	(92%)
$S = 8$	5.4	6.0	7.5	4.8	5.8	6.6	6.0	5.7	6.1
	(87%)	(89%)	(94%)	(86%)	(91%)	(94%)	(85%)	(89%)	(93%)

J is the number of temporal PCs. The explained variance of the temporal functional principal component analysis (FPCA) is given in parentheses.

Table II. Empirical power ($c = 1$)

	$N = 100$			$N = 150$			$N = 200$		
	$J = 2$	$J = 3$	$J = 4$	$J = 2$	$J = 3$	$J = 4$	$J = 2$	$J = 3$	$J = 4$
$S = 4$	67.6	90.6	95.1	91.9	99.3	99.8	98.2	100	100
	(86%)	(90%)	(94%)	(87%)	(90%)	(93%)	(87%)	(92%)	(94%)
$S = 6$	54.5	79.7	89.0	80.7	97.9	99.3	94.5	99.7	100
	(88%)	(91%)	(94%)	(85%)	(91%)	(94%)	(88%)	(92%)	(94%)
$S = 8$	45.2	74.9	85.2	75.1	96.8	98.7	91.5	99.9	100
	(89%)	(91%)	(94%)	(89%)	(92%)	(94%)	(88%)	(92%)	(94%)

J is the number of temporal PCs. The explained variance of the temporal FPCA is given in parentheses.

DOI: 10.1111/jtsa.12302

Table III. Rejection rates under H_0 ($c = 0$)

		$N = 100$			$N = 150$			$N = 200$		
		$J = 2$	$J = 3$	$J = 4$	$J = 2$	$J = 3$	$J = 4$	$J = 2$	$J = 3$	$J = 4$
$S = 10$	$K = 2$	6.4 (80%)	6.2 (84%)	6.1 (90%)	6.1 (80%)	5.2 (85%)	4.3 (88%)	5.8 (80%)	5.5 (84%)	5.6 (88%)
	$K = 3$	6.1 (84%)	4.8 (88%)	5.6 (94%)	5.0 (83%)	5.5 (89%)	4.7 (92%)	5.3 (85%)	6.1 (89%)	5.0 (92%)
	$K = 4$	5.9 (84%)	6.3 (90%)	4.7 (92%)	5.2 (85%)	5.8 (90%)	5.3 (92%)	6.1 (84%)	6.1 (90%)	5.8 (92%)
$S = 12$	$K = 2$	6.3 (83%)	6.4 (88%)	6.1 (90%)	5.6 (83%)	6.0 (86%)	6.2 (89%)	4.6 (80%)	6.3 (87%)	6.4 (88%)
	$K = 3$	6.1 (87%)	5.9 (91%)	5.1 (93%)	5.0 (87%)	5.6 (90%)	6.0 (92%)	6.1 (85%)	4.8 (90%)	6.1 (93%)
	$K = 4$	6.0 (87%)	5.4 (91%)	5.0 (93%)	6.3 (85%)	6.1 (90%)	6.0 (93%)	5.0 (86%)	6.5 (90%)	5.7 (93%)
$S = 14$	$K = 2$	6.4 (82%)	5.2 (87%)	4.5 (89%)	6.2 (82%)	5.8 (87%)	5.0 (88%)	5.6 (82%)	6.6 (86%)	5.3 (89%)
	$K = 3$	6.0 (85%)	5.2 (90%)	4.7 (92%)	4.4 (83%)	6.2 (88%)	6.0 (92%)	4.2 (84%)	5.6 (90%)	6.2 (93%)
	$K = 4$	6.0 (85%)	5.1 (90%)	4.6 (93%)	6.3 (86%)	5.5 (90%)	5.7 (93%)	6.5 (87%)	5.7 (89%)	5.6 (91%)

K is the reduced panel dimension and J the number of temporal PCs. The explained variance of the dimension reduction is given in parentheses.

Table IV. Empirical power ($c = 1$)

		$N = 100$			$N = 150$			$N = 200$		
		$J = 2$	$J = 3$	$J = 4$	$J = 2$	$J = 3$	$J = 4$	$J = 2$	$J = 3$	$J = 4$
$S = 10$	$K = 2$	31.3 (85%)	49.2 (85%)	60.8 (83%)	50.5 (81%)	78.0 (85%)	85.9 (84%)	67.3 (81%)	92.4 (82%)	96.4 (84%)
	$K = 3$	43.0 (87%)	73.3 (90%)	81.7 (93%)	72.3 (85%)	95.2 (91%)	98.2 (93%)	90.6 (86%)	99.5 (91%)	100 (93%)
	$K = 4$	42.2 (85%)	73.5 (92%)	84.7 (93%)	71.1 (87%)	96.6 (92%)	98.5 (94%)	90.6 (88%)	99.8 (90%)	100 (92%)
$S = 12$	$K = 2$	30.8 (82%)	49.7 (83%)	57.4 (85%)	46.6 (81%)	76.5 (83%)	87.2 (84%)	67.7 (82%)	91.8 (83%)	95.7 (86%)
	$K = 3$	42.9 (89%)	72.5 (91%)	82.8 (94%)	67.1 (88%)	94.8 (92%)	98.8 (93%)	89.5 (87%)	99.5 (93%)	99.9 (93%)
	$K = 4$	43.3 (87%)	72.0 (92%)	82.9 (94%)	71.1 (86%)	95.9 (92%)	97.8 (94%)	89.0 (86%)	99.6 (91%)	100 (93%)
$S = 14$	$K = 2$	27.7 (86%)	46.2 (84%)	55.0 (84%)	47.7 (82%)	74.9 (83%)	82.8 (84%)	69.0 (81%)	90.6 (84%)	94.0 (87%)
	$K = 3$	39.2 (87%)	66.6 (92%)	81.3 (93%)	67.5 (89%)	91.0 (91%)	93.4 (93%)	88.1 (88%)	94.4 (90%)	94.1 (93%)
	$K = 4$	43.7 (87%)	70.4 (92%)	78.9 (94%)	70.5 (88%)	91.1 (91%)	93.7 (94%)	88.2 (88%)	94.4 (93%)	95.7 (94%)

K and J are as in Table III. The explained variance of the dimension reduction is given in parentheses.

Tables III and IV show that the reduction of the panel dimension does not negatively affect the properties of the tests. The conclusions are the same as in Section 5.1. Either approach leads to a test with well-controlled size, which does not depend on J (J, K) as long as the proportion of explained variance remains within the generally recommended range of 85–95%. If $J = 2$ or $K = 2$ are used, this requirement is generally not met, resulting in a size distortion, which is however acceptable and decreases with N.

As noted at the beginning of this section, the tests of Constantinou *et al.* (2017) are too conservative; they almost never reject under the null for all scenarios considered in this section. The tests of Aston *et al.* (2017) reject too often under the null. For example, in the settings considered in Table III, the rejection rates for their asymptotic test, Gaussian parametric bootstrap test, and Gaussian parametric bootstrap test using Hilbert–Schmidt distance range 19.0–49.4%, 14.6–32.2%, and 38.1–44.9% respectively. By contrast, the test derived in this paper, in its both versions and under all reasonable choices of tuning parameters, has precise empirical size at the standard 5% nominal level and useful power.

DOI: 10.1111/jtsa.12302

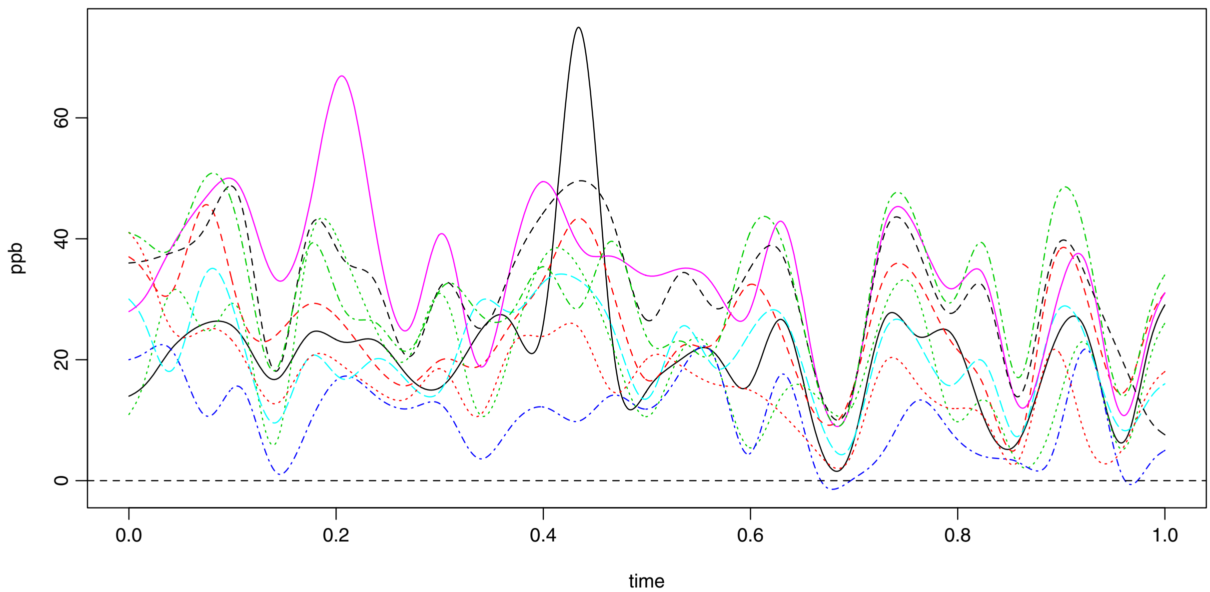

Figure 1. Maximum 1-h nitrogen dioxide curves for December 2012 at the nine locations [Color figure can be viewed at wileyonlinelibrary.com]

In Section B of Supporting Information, we show the results of other simulations that study the effect of different covariance functions, the magnitude of the departure from H_0, and the lag h. They do not modify the general conclusion that the test is reasonably well calibrated and has useful power.

6. APPLICATIONS TO POLLUTION AND STOCK MARKET DATA

We begin by applying our method to air quality data studied by Constantinou *et al.* (2017) under the assumption that the monthly curves are i.i.d. These curves, however, form a time series, so it is important to check whether a test that accounts for the temporal dependence leads to the same or a different conclusion.

The Environmental Protection Agency (EPA) collects massive amounts of air quality data which are available through its website http://www3.epa.gov/airdata/ad_data_daily.html. The records consist of data for six common pollutants, collected by outdoor monitors in hundreds of locations across the United States. The number and frequency of the observations vary greatly by location, but some locations have as many as three decades worth of daily measurements. We focus on nitrogen dioxide, a common pollutant emitted by combustion engines and power stations.

We consider nine locations along the east coast that have relatively complete records since 2000: Allentown, Baltimore, Boston, Harrisburg, Lancaster, New York City, Philadelphia, Pittsburgh, and Washington D.C. We use the data for the years 2000–2012. Each functional observation $X_{ns}(t)$ consists of the daily maximum 1-h nitrogen dioxide concentration measured in ppb (parts per billion) for day t, month n ($N = 156$), and location s. We thus have a panel of $S = 9$ functional time series (one at every location), $X_{ns}(t), s = 1, 2, \dots, 9, n = 1, 2, \dots, 156$. Figure 1 shows the data for the nine locations for December 2012. Before the application of the test, the curves were deseasonalized by removing the monthly mean from each curve.

We applied both versions of Procedure 4.1 (dimension in time only and double dimension reduction). Requiring 85–95% of explained variance yielded the values $J, K = 2, 3, 4$, similar as in our simulated data example. For all possible combinations, we obtained p-values smaller than 10E−−4. This indicates a nonseparable covariance function and confirms the conclusion obtained by Constantinou *et al.* (2017); nonseparability is an intrinsic feature of pollution data, and simplifying the covariance structure by assuming separability may lead to incorrect conclusions.

DOI: 10.1111/jtsa.12302

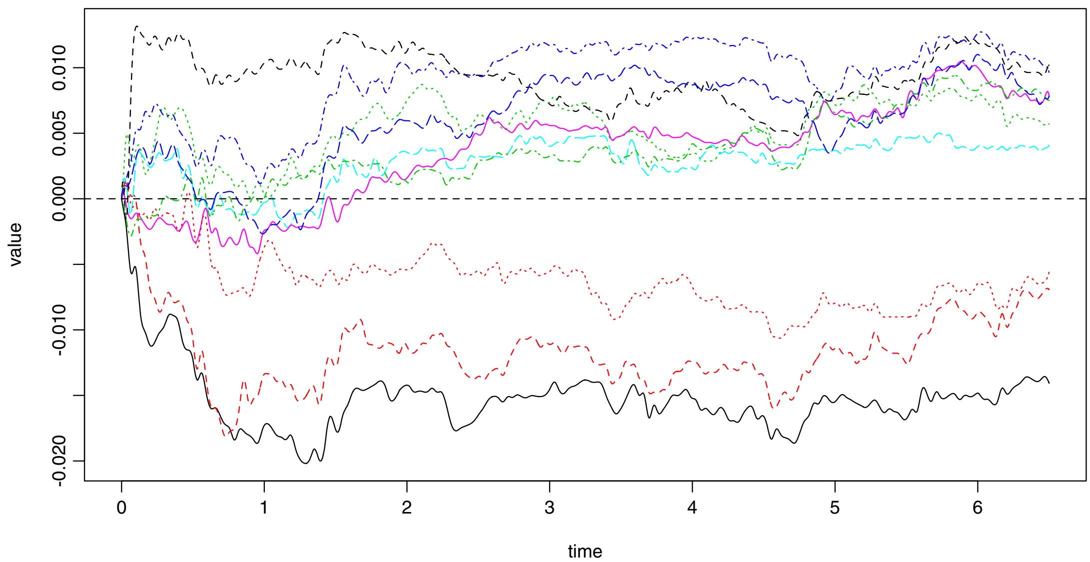

Figure 2. Cumulative intraday return curves for the 10 companies for April 2, 2007 [Color figure can be viewed at wileyonlinelibrary.com]

We now turn to an application to a stock portfolio. Cumulative intraday returns have recently been studied in several papers, including Kokoszka and Reimherr (2013b), Kokoszka *et al.* (2015), and Lucca and Moench (2015). If $P_n(t)$ is the price of a stock at minute t of the trading day n, then the cumulative intraday return curve on day n is defined by

$$R_n(t) = \log(P_n(t)) - \log(P_n(0)),$$

where time 0 corresponds to the opening of the market (9:30 EST for the NYSE). Horváth *et al.* (2014) did not find evidence against temporal stationarity of such time series. The work of Kokoszka and Reimherr (2013b) shows that cumulative intradaily returns do not form an i.i.d. sequence. (This can be readily verified by computing the autocorrelation function (ACF) of squared scores.) Figure 2 shows the curves R_n for 10 companies on April 2, 2007. This portfolio of $S = 10$ stocks produces a panel of functional time series studied in this paper. We selected ten U.S. blue chip companies, and want to determine whether the resulting panel can be assumed to have a separable covariance function. The answer is yes, as we now explain.

We consider stock values, recorded every minute, from October 10, 2001 to April 2, 2007 (1378 trading days) for the following 10 companies: Bank of America (BOA), Citi Bank, Coca Cola, Chevron Corporation (CVX), Walt Disney Company (DIS), International Business Machines (IBM), McDonald's Corporation (MCD), Microsoft Corporation (MSFT), Walmart Stores (WMT) and Exxon Mobil Corporation Common (XOM). On each trading day, there are 390 discrete observations. There is an outlier on August 26, 2004 for Bank of America, which is due to a stock split. That day is discarded from further analysis, so the sample size is $N = 1377$.

We now discuss the results of applying Procedure 4.1. Using dimension reduction in time only, we obtained p-values 0.234 for $J = 2$ (CPV = 92%) and 0.220 for $J = 3$ (CPV= 95%). Using the double dimension reduction, we obtained the following values:

	p-Value	CPV
$K = 2, J = 2$	0.272	45%
$K = 3, J = 3$	0.217	62%
$K = 4, J = 4$	0.224	67%
$K = 6, J = 4$	0.223	80%
$K = 7, J = 4$	0.221	85%

These remarkably similar *p*-values indicate that panels of cumulative intraday return curves can in some cases be assumed to have a separable covariance function. This could be useful for portfolio managers, as it indicates that they can exploit separability of the data for more efficient modeling.

7. CONCLUSION

We conclude by noting that, in practice, it is important to ensure that the time series forming the panel are at comparable scales. This has been the case in our data examples, and will be the case if the series are measurements of the same quantity and are generated as a single group. If some of the series are much more variable than the others, they may bias the test, and should perhaps be considered separately.

ACKNOWLEDGMENTS

This work was partially supported by the United States National Science Foundation grants at Colorado State University and Penn State University.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Ang A, Bekaert G. 2002. Regime switches in interest rates. *Journal of Business and Economic Statistics* **20**: 163–182.
Aston JAD, Pigoli D, Tavakoli S. 2017. Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics* **45**: 1431–1461.
Bagchi P, Dette H. 2017. A test for separability in covariance operators of random surfaces. Technical Report, Ruhr-Universität, Bochum, arXiv:1710.08388.
Bosq D. 2000. *Linear Processes in Function Spaces*. New York: Springer.
Bowsher CG, Meeks R. 2008. The dynamics of economic functions: modeling and forecasting the yield curve. *Journal of the American Statistical Association* **103**: 1419–1437.
Constantinou P, Kokoszka P, Reimherr M. 2017. Testing separability of space–time functional processes. *Biometrika* **104**: 425–437.
Cressie NAC, Huang H-C. 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* **94**: 1330–1339.
Cressie NAC, Wikle CK. 2015. *Statistics for Spatio-Temporal Data*. Hoboken: John Wiley & Sons.
French J, Kokoszka P, Stoev S, Hall L. 2016. Quantifying the risk of extreme heat waves over North America using climate model forecasts. Technical Report, Colorado State University.
Fuentes M. 2006. Testing for separability of spatial–temporal covariance functions. *Journal of Statistical Planning and Inference* **136**: 447–466.
Gromenko O, Kokoszka P, Reimherr M. 2016. Detection of change in the spatiotemporal mean function. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**: 29–50.
Gromenko O, Kokoszka P, Sojka J. 2017. Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *The Annals of Applied Statistics* **11**: 898–918.
Hays S, Shen H, Huang JZ. 2012. Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics* **6**: 870–894.
Hoff PD. 2011. Separable covariance arrays via the Tucker product, with applications to multivariate relational data. *Bayesian Analysis* **6**: 179–196.
Hörmann S, Kokoszka P. 2010. Weakly dependent functional data. *The Annals of Statistics* **38**: 1845–1884.
Horváth L, Kokoszka P. 2012. *Inference for Functional Data with Applications*. New York: Springer.
Horváth L, Kokoszka P, Rice G. 2014. Testing stationarity of functional time series. *Journal of Econometrics* **179**: 66–82.
Hsing T, Eubank R. 2015. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Hoboken: John Wiley & Sons.
Kokoszka P, Reimherr M. 2013a. Asymptotic normality of the principal components of functional time series. *Stochastic Processes and their Applications* **123**: 1546–1562.
Kokoszka P, Reimherr M. 2013b. Predictability of shapes of intraday price curves. *The Econometrics Journal* **16**: 285–308.

DOI: 10.1111/jtsa.12302

Kokoszka P, Reimherr M. 2017. *Introduction to Functional Data Analysis*. Boca Raton: CRC Press.
Kokoszka P, Miao H, Zhang X. 2015. Functional dynamic factor model for intraday price curves. *Journal of Financial Econometrics* **13**: 456–477.
Kokoszka P, Reimherr M, Wölfing N. 2016. A randomness test for functional panels. *Journal of Multivariate Analysis* **151**: 37–53.
Kowal DR, Matteson DS, Ruppert D. 2017. A Bayesian multivariate functional dynamic linear model. *Journal of the American Statistical Association* **112**: 733–744.
Liu C, Ray S, Hooker G. 2017. Functional principal components analysis of spatially correlated data. *Statistics and Computing* **27**: 1639–1654.
Lucca DO, Moench E. 2015. The pre-FOMC announcement drift. *The Journal of Finance* **70**: 329–371.
Mitchell MW, Genton MG, Gumpertz ML. 2005. Testing for separability of space–time covariances. *Environmetrics* **16**: 819–831.
Mitchell MW, Genton MG, Gumpertz ML. 2006. A likelihood ratio test for separability of covariances. *Journal of Multivariate Analysis* **97**: 1025–1043.
Paul D, Peng J. 2011. Principal components analysis for sparsely observed correlated functional data using a kernel smoothing approach. *Electronic Journal of Statistics* **5**: 1960–2003.
Shang HL, Hyndman RJ. 2017. Grouped functional time series forecasting: an application to age-specific mortality rates. *Journal of Computational and Graphical Statistics* **26**: 330–343.
Shao X, Wu WB. 2007. Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics* **35**: 1773–1801.
Sun Y, Li B, Genton MG. 2012. Geostatistics for large datasets. In *Advances and Challenges in Space–time Modelling of Natural Events*, Porcu E, Montero JM, Schlather M (eds.). Berlin: Springer; 55–77 (chap. 3).
Tupper LL, Matteson DS, Anderson CL, Zephyr L. 2017. Band depth clustering for nonstationary time series and wind speed behavior. *Technometrics*. Published online: 06 Dec 2017.
van den Boogaart KG. 2007. tensorA: Advanced tensors arithmetic with named indices. R package version 0.31. http://CRAN R-project. org/package= tensorA.

DOI: 10.1111/jtsa.12302