The Annals of Statistics
2020, Vol. 48, No. 4, 2303–2322
https://doi.org/10.1214/19-AOS1888

A TEST FOR SEPARABILITY IN COVARIANCE OPERATORS OF RANDOM SURFACES

BY PRAMITA BAGCHI[1] AND HOLGER DETTE[2]

[1] *Department of Statistics, George Mason University*
[2] *Fakultät für Mathematik, Ruhr-Universität Bochum*, holger.dette@rub.de

The assumption of separability is a simplifying and very popular assumption in the analysis of spatiotemporal or hypersurface data structures. It is often made in situations where the covariance structure cannot be easily estimated, for example, because of a small sample size or because of computational storage problems. In this paper we propose a new and very simple test to validate this assumption. Our approach is based on a measure of separability which is zero in the case of separability and positive otherwise. We derive the asymptotic distribution of a corresponding estimate under the null hypothesis and the alternative and develop an asymptotic and a bootstrap test which are very easy to implement. In particular, our approach does neither require projections on subspaces generated by the eigenfunctions of the covariance operator nor distributional assumptions as recently used by (*Ann. Statist.* **45** (2017) 1431–1461) and (*Biometrika* **104** 425–437) to construct tests for separability. We investigate the finite sample performance by means of a simulation study and also provide a comparison with the currently available methodology. Finally, the new procedure is illustrated analyzing a data example.

1. Introduction. Functional *and* multidimensional data is usually called surface data and arises in areas such as medical imaging (see [21, 25]) spectrograms derived from audio signals or geolocalized data (see [3, 19]). In many of these high-dimensional problems, a completely nonparametric estimation of the covariance operator is not possible as the number of available observations is small compared to the dimension. A common approach to obtain reasonable estimates in this context are structural assumptions on the covariance of the underlying process, and in recent years the assumption of separability has become very popular, for example, in the analysis of geostatistical space-time models (see [9, 11], among others). Roughly speaking, this assumption allows us to write the covariance

$$c(s,t,s',t') = \mathbb{E}[X(s,t)X(s',t')]$$

of a (real valued) space-time process $\{X(s,t)\}_{(s,t)\in S\times T}$ as a product of the space and time covariance function, that is,

$$c(s,t,s',t') = c_1(s,s')c_2(t,t'). \tag{1.1}$$

It was pointed out by many authors that the assumption of separability yields a substantial simplification of the estimation problem and thus reduces computational costs in the estimation of the covariance in high dimensional problems (see, e.g., [14, 20]). Despite of its importance, there exist only a few tools to validate the assumption of separability for surface data.

Received December 2018; revised July 2019.
MSC2020 subject classifications. 62G10, 62G20.
Key words and phrases. Functional data, minimum distance, separability, space-time processes, surface data structures.

Many authors developed tests for spatiotemporal data. For example, [8] proposed a test based on the spectral representation, and [15, 17, 18] investigated likelihood ratio tests under the assumption of a normal distribution. Recently, [4] derived the joint distribution of the three statistics appearing in the likelihood ratio test and used this result to derive the asymptotic distribution of the (log) likelihood ratio. These authors also proposed alternative tests, which are based on distances between an estimator of the covariance, under the assumption of separability and an estimator which does not use this assumption. [5] generalized the latter (distance-based) approach to test the assumption of separability for functional time series. To address for serial dependence, they also considered hypotheses of the form (1.1) for lagged covariance operators. More recently, [16] introduced the concept of weak separability, which, roughly speaking, means that the eigenfunctions of the covariance operator c can be written as tensor products of the eigenfunctions of c_1 and c_2. In particular, strong separability as specified by (1.1) is a special case of weak separability, and the latter hypothesis can be tested by checking if the Fourier coefficients calculated with respect to products of basis functions are uncorrelated.

[1] considered the problem of testing for separability in the context of hypersurface data. These authors pointed out that many available methods require the estimation of the full multidimensional covariance structure which can become infeasible for high dimensional data. In order to address this issue, they developed tests based on CLT approximations, as well as bootstrap tests for applications, where replicates from the underlying random process are available. To avoid estimation and storage of the full covariance, finite-dimensional projections of the difference between the covariance operator and a nonparametric separable approximation (using the partial trace operator) were proposed. In particular, they suggested to project onto subspaces generated by the eigenfunctions of the covariance operator estimated under the assumption of separability. However, as pointed in the same references the choice of the number of eigenfunctions onto which one should project is not trivial, and the test might be sensitive with respect to this choice. Moreover, the computational costs increase substantially with the number of eigenfunctions.

In this paper we present an alternative and simple test for the hypothesis of separability in hypersurface data. We consider a similar setup as in [1] and proceed in two steps. Roughly speaking, we derive an *explicit* expression for the minimal distance between the covariance operator and its approximation by a separable covariance operator. It turns out that this minimum vanishes if and only if the covariance operator is separable. Second, we directly estimate the minimal distance (and not the covariance operator itself) from the available data. As a consequence the calculation of the test statistic does neither use an estimate of the full nonseparable covariance operator nor requires the specification of subspaces used for a projection.

In Section 2 we review some basic terminology and discuss the problem of finding a best approximation of the covariance operator by a separable covariance operator. The corresponding minimum distance could also be interpreted as a measure of deviation from separability (it is zero in the case of separability and positive otherwise). In Section 3 we propose an estimator of the minimum distance, prove its consistency and derive its asymptotic distribution under the null hypothesis and alternative. These results are also used to develop an asymptotic and a bootstrap test for the hypothesis of separability, which are—in contrast to the currently available methods—consistent against all alternatives. Moreover, statistical guarantees can be derived under more general and easier to verify moment assumptions than in [1]. Section 4 is devoted to an investigation of the finite sample properties of the new tests and a comparison with alternative tests for this problem which have recently been proposed by [1] and [4]. In particular, we demonstrate that, despite their simplicity, the new procedures have very competitive properties compared to the currently available methodology. Finally, some technical details are deferred to the Supplementary Material [2].

2. Hilbert spaces and a measure of separability. We begin introducing some basic facts about Hilbert spaces, Hilbert–Schmidt operators and tensor products. For more details we refer to the monographs of [7, 24] or [12]. Let H be a real separable Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. The space of bounded linear operators on H is denoted by $S_\infty(H)$ with operator norm

$$|||T|||_\infty := \sup_{\|f\|\le 1} \|Tf\|.$$

A bounded linear operator T is said to be compact if it can be written as

$$T = \sum_{j\ge 1} s_j(T)\langle e_j,\cdot\rangle f_j,$$

where $\{e_j : j \ge 1\}$ and $\{f_j : j \ge 1\}$ are orthonormal sets of H, $\{s_j(T) : j \ge 1\}$ are the singular values of T and the series converges in the operator norm. We say that a compact operator T belongs to the Schatten class of order $p \ge 1$ and write $T \in S_p(H)$ if

$$|||T|||_p = \Bigl(\sum_{j\ge 1} s_j(T)^p\Bigr)^{1/p} < \infty.$$

The Schatten class of order $p \ge 1$ is a Banach space with norm $|||\cdot|||_p$ and with the property that $S_p(H) \subset S_q(H)$ for $p < q$. In particular, we are interested in Schatten classes of order $p = 1$ and 2. A compact operator T is called Hilbert–Schmidt operator if $T \in S_2(H)$ and trace class if $T \in S_1(H)$. The space of Hilbert–Schmidt operators $S_2(H)$ is also a Hilbert space with the Hilbert–Schmidt inner product given by

$$\langle A, B\rangle = \sum_{j\ge 1}\langle Ae_j, Be_j\rangle$$

for each $A, B \in S_2(H)$, where $\{e_j : j \ge 1\}$ is an orthonormal basis (here the inner product does not depend on the choice of the basis).

Let H_1 and H_2 be two real separable Hilbert spaces. For $u \in H_1$ and $v \in H_2$, we define the bilinear form $u \otimes v : H_1 \times H_2 \to \mathbb{R}$ by

$$[u \otimes v](x, y) := \langle u, x\rangle\langle v, y\rangle, \quad (x, y) \in H_1 \times H_2.$$

Let $\mathscr{P}$ be the set of all finite linear combinations of such bilinear forms. An inner product on $\mathscr{P}$ can be defined by the linear extension of $\langle u \otimes v, w \otimes z\rangle = \langle u, w\rangle\langle v, z\rangle$, for $u, w \in H_1$ and $v, z \in H_2$. The completion of $\mathscr{P}$ under the aforementioned inner product is called the tensor product of H_1 and H_2 and denoted as $H_1 \otimes H_2$.

For $C_1 \in S_\infty(H_1)$ and $C_2 \in S_\infty(H_2)$, the tensor product $C_1 \tilde{\otimes} C_2$ is defined as the unique linear operator on $H := H_1 \otimes H_2$ satisfying

$$(C_1 \tilde{\otimes} C_2)(u \otimes v) = C_1u \otimes C_2v, \quad u \in H_1, v \in H_2.$$

In fact $C_1 \tilde{\otimes} C_2 \in S_\infty(H)$ with $|||C_1 \tilde{\otimes} C_2|||_\infty = |||C_1|||_\infty|||C_2|||_\infty$. Moreover, if $C_1 \in S_p(H_1)$ and $C_2 \in S_p(H_2)$ for $p \ge 1$, then $C_1 \tilde{\otimes} C_2 \in S_p(H)$ with $|||C_1 \tilde{\otimes} C_2|||_p = |||C_1|||_p|||C_2|||_p$. For more details we refer to Chapter 8 of [24]. In the sequel we denote the Hilbert–Schmidt inner product on $S_2(H)$ with $H = H_1 \otimes H_2$ as $\langle\cdot,\cdot\rangle_{\mathrm{HS}}$ and that of $S_2(H_1)$ and $S_2(H_2)$ as $\langle\cdot,\cdot\rangle_{S_2(H_1)}$ and $\langle\cdot,\cdot\rangle_{S_2(H_2)}$, respectively.

2.1. *Measuring separability.* We consider random elements X in the Hilbert space H with $\mathbb{E}\|X\|^4 < \infty$ (see Chapter 7 in [13] for more details on Hilbert space valued random variables). Then, the covariance operator of X is defined as $C := \mathbb{E}[(X - \mathbb{E}X) \otimes_o (X - \mathbb{E}X)]$ where for $f, g \in H$ the operator $f \otimes_o g : H \to H$ is defined by

$$(f \otimes_o g)h = \langle h, g\rangle f \quad \text{for all } h \in H.$$

Under the assumption $\mathbb{E}\|X\|^4 < \infty$, we have $C \in S_2(H)$. We also assume $|||C|||_2 \neq 0$ which essentially means the random variable X is nondegenerate. To test separability, we consider the hypothesis

$$H_0 : C = C_1 \tilde{\otimes} C_2 \quad \text{for some } C_1 \in S_2(H_1) \text{ and } C_2 \in S_2(H_2). \tag{2.1}$$

Our approach is based on an approximation of the operator C by a separable operator $C_1 \tilde{\otimes} C_2$ with respect to the norm $|||\cdot|||_2$. Ideally, we are looking for

$$D := \inf\{|||C - C_1 \tilde{\otimes} C_2|||_2^2 \mid C_1 \in S_2(H_1), C_2 \in S_2(H_2)\}, \tag{2.2}$$

such that the hypothesis of separability in (2.1) can be rewritten in terms of the distance D, that is,

$$H_0 : D = 0 \quad \text{versus} \quad H_1 : D > 0. \tag{2.3}$$

It turns out that it is difficult to express D explicitly in terms of the covariance operator C. For this reason we proceed in a slightly different way in two steps. First, we fix an operator $C_1^* \in S_2(H_1)$ and determine

$$D_{C_1^*} := \inf\{|||C - C_1^* \tilde{\otimes} C_2|||_2^2 \mid C_2 \in S_2(H_2)\}, \tag{2.4}$$

that is, we are minimizing $|||C - C_1^* \tilde{\otimes} C_2|||_2^2$ with respect to second factor C_2 of the tensor product. In particular, we will show that the infimum is in fact a minimum and derive an explicit expression for $D_{C_1^*}$ and its minimizer. Instead of working with the distance D in (2.2), we suggest to estimate an appropriate distance from the family

$$\{D_{C_1^*} \mid C_1^* \in S_2(H_1)\}.$$

For this purpose note that for a given covariance operator $C \in S_2(H)$ and $C_1^* \in S_2(H_1)$ the corresponding distance $D_{C_1^*}$ is in general positive. However, we also show in the following that C is separable, that is, $C = C_1 \tilde{\otimes} C_2$, if and only if the corresponding minimum D_{C_1} vanishes. Thus, if we are able to estimate D_{C_1} (for the unknown operator C_1), we can test the hypothesis (2.3) by constructing a test for the hypotheses $H_0 : D_{C_1} = 0$ vs. $H_1 : D_{C_1} > 0$. We explain below that this is in fact possible.

For this purpose we have to introduce some additional notation and have to prove several auxiliary results, which will be proved in Section B.1 of the Supplementary Material. The main statement is given in Theorem 2.1 (whose formulation also requires the new notation). First, consider the bounded linear operator $T_1 : S_2(H) \times S_2(H_1) \mapsto S_2(H_2)$ defined by

$$T_1(A \tilde{\otimes} B, C_1) = \langle A, C_1\rangle_{S_2(H_1)} B \tag{2.5}$$

for all $C_1 \in S_2(H_1)$. Similarly, let $T_2 : S_2(H) \times S_2(H_2) \to S_2(H_1)$ be the bounded linear operator defined by

$$T_2(A \tilde{\otimes} B, C_2) = \langle B, C_2\rangle_{S_2(H_2)} A \tag{2.6}$$

for all $C_2 \in S_2(H_2)$. The proofs of the following two auxiliary results can be found in Sections B.2 and B.3 of the Supplementary Material.

PROPOSITION 2.1. *The operators T_1 and T_2 are well defined, bilinear and continuous with*

$$\langle B, T_1(C, C_1)\rangle_{S_2(H_2)} = \langle C, C_1 \mathbin{\tilde{\otimes}} B\rangle_{\mathrm{HS}}, \tag{2.7}$$

$$\langle A, T_2(C, C_2)\rangle_{S_2(H_1)} = \langle C, A \mathbin{\tilde{\otimes}} C_2\rangle_{\mathrm{HS}} \tag{2.8}$$

for all $A, C_1 \in S_2(H_1)$, $B, C_2 \in S_2(H_2)$ and $C \in S_2(H)$.

While the previous result clarifies the existence of the operators T_1 and T_2, the next proposition provides a property of the operator T_1 which is essential for the proof of the main result in this section.

PROPOSITION 2.2. *For any $C \in S_2(H)$ and $C_1 \in S_2(H_1)$, we have*

$$\langle C, C_1 \mathbin{\tilde{\otimes}} T_1(C, C_1)\rangle_{\mathrm{HS}} = |||T_1(C, C_1)|||_2^2.$$

THEOREM 2.1. *For each $C \in S_2(H)$ and any fixed $C_1^* \in S_2(H_1)$, the distance*

$$D_{C_1^*}(C_2) = |||C - C_1^* \mathbin{\tilde{\otimes}} C_2|||_2 \tag{2.9}$$

is minimized at

$$\tilde{C}_2 = \frac{T_1(C, C_1^*)}{|||C_1^*|||_2^2}. \tag{2.10}$$

Moreover, the minimum distance in (2.9) *is given by*

$$D_{C_1^*} = |||C|||_2^2 - \frac{|||T_1(C, C_1^*)|||_2^2}{|||C_1^*|||_2^2}. \tag{2.11}$$

In particular, $D_{C_1^*}$ is zero if and only if $C = C_1^* \mathbin{\tilde{\otimes}} C_2$ for some $C_2 \in S_2(H_2)$.

REMARK 2.1. By Theorem 2.1 we can construct a test for the hypothesis

$$H_0 : D_{C_1^*} = 0$$

for any given $C_1^* \in S_2(H_1)$ by estimating the quantity in (2.11). If the covariance operator C is not separable, it follows that $D_{C_1^*} > 0$ for all $C_1^* \in S_2(H_1)$. If C is in fact separable (i.e., the null hypothesis is true) such that $C = C_1 \mathbin{\tilde{\otimes}} C_2$ for some C_1 and C_2, we have $D_{C_1} = D_{C_1}(C_2) = 0$. Interestingly, we can obtain C_1 from C up to a multiplicative constant using the operator T_2 defined in (2.6). More precisely, we choose an arbitrary but fixed element $\Psi \in S_2(H_2)$ such that $T_2(C, \Psi) \neq 0$ and note that under the null hypothesis of separability we have $T_2(C, \Psi) = \langle C_2, \Psi\rangle_{S_2(H_2)} C_1$. As the minimum distance in (2.11) is invariant with respect to scalar multiplication of C_1^* it follows for this choice

$$D_0 := D_{C_1} = D_{T_2(C,\Psi)} = |||C|||_2^2 - \frac{|||T_1(C, T_2(C, \Psi))|||_2^2}{|||T_2(C, \Psi)|||_2^2}. \tag{2.12}$$

Note that $D_0 \geq 0$ and D_0 vanishes if and only if C is separable. Thus, we can construct a consistent test of the hypothesis (2.1) via a suitable estimate of the operator C in (2.12). This program is carefully carried out in Section 3.

REMARK 2.2. Note that the representation (2.12) involves only norms of operators and as a consequence, when it comes to estimation, we do not have to store the complete estimate of the covariance kernel. We make this point more precise in Remark 3.4, where we discuss the problem of estimating D_0 in the context of Hilbert–Schmidt integral operators.

2.2. *Hilbert–Schmidt integral operators.* An important case for applications is the set $H := L^2(S \times T)$ of all real-valued square integrable functions defined on $S \times T$, where $S \subset \mathbb{R}^p$, $T \subset \mathbb{R}^q$ are bounded measurable sets. If the covariance operator C of the random variable X is a Hilbert–Schmidt operator, it follows from Theorem 6.11 in [24] that there exists a kernel $c \in L^2((S \times T) \times (S \times T))$ such that C can be characterized as an integral operator, that is,

$$Cf(s,t) = \int_S \int_T c(s,t,s',t') f(s',t')\, ds'\, dt', \quad f \in L^2(S \times T),$$

almost everywhere on $S \times T$. Moreover, the kernel is given by the covariance kernel of X, that is, $c(s,t,s',t') = \operatorname{Cov}[X(s,t), X(s',t')]$, and the space H can be identified with the tensor product of $H_1 = L^2(S)$ and $H_2 = L^2(T)$.

Similarly, if C_1 and C_2 are Hilbert–Schmidt operators on $L^2(S)$ and $L^2(T)$, respectively, there exist symmetric kernels $c_1 \in L^2(S \times S)$ and $c_2 \in L^2(T \times T)$ such that

$$C_1 f(s) = \int_S c_1(s,s') f(s')\, ds', \qquad C_2 g(t) = \int_T c_2(t,t') g(t')\, dt'$$

($f \in H_1$, $g \in H_2$) almost everywhere on S and T, respectively. The following result shows that in this case the operators T_1 and T_2 defined by (2.5) and (2.6), respectively, are also Hilbert–Schmidt integral operators. The proof can be found in Section B.5 of the Supplementary Material and requires that the sets S and T are bounded.

PROPOSITION 2.3. *If C and C_1 are integral operators with continuous kernels $c \in L^2((S \times T) \times (S \times T))$ and $c_1 \in L^2(S \times S)$, then $T_1(C, C_1)$ is an integral operator with kernel given by*

$$k(t,t') = \int_S \int_S c(s,t,s't') c_1(s,s')\, ds\, ds'. \tag{2.13}$$

An analog result is true for the operator T_2.

Using the explicit formula for T_1 described in Proposition 2.3 the minimum distance in Theorem 2.1 can be expressed in terms of the corresponding kernels of the operators, that is,

$$D_{C_1} = \int_T \int_T \int_S \int_S c^2(s,t,s',t')\, ds\, ds'\, dt\, dt' - \frac{\int_T \int_T [\int_S \int_S c(s,t,s't') c_1(s,s')\, ds\, ds']^2\, dt\, dt'}{\int_S \int_S c_1^2(s,s')\, ds\, ds'}.$$

3. Estimation and asymptotic properties. Formally, we estimate the minimum distance given in (2.11) by plugging in estimators for C and C_1 based on a sample $X_1, X_2, \ldots, X_N$. The covariance operator C is estimated by

$$\widehat{C}_N := \frac{1}{N} \sum_{i=1}^{N} [(X_i - \overline{X}) \otimes_o (X_i - \overline{X})]. \tag{3.1}$$

As pointed out in Remark 2.1 it is sufficient to estimate the operator C_1 up to a multiplicative constant, due to the self-normalizing form of the second term of the minimum distance D_{C_1}. Let Ψ be any fixed element of $S_2(H_2)$; recall that under the null hypothesis of separability $H_0: C = C_1 \widetilde{\otimes} C_2$ we have $T_2(C, \Psi) = \langle C_2, \Psi \rangle_{S_2(H_2)} C_1$. Observing the representation (2.12), we suggest to estimate (a multiple of) the operator C_1 by

$$\widehat{C}_{1N} = T_2(\widehat{C}_N, \Psi). \tag{3.2}$$

With this choice we obtain the test statistic

$$\widehat{D}_N = |||\widehat{C}_N|||_2^2 - \frac{|||T_1(\widehat{C}_N, T_2(\widehat{C}_N, \Psi))|||_2^2}{|||T_2(\widehat{C}_N, \Psi)|||_2^2}. \tag{3.3}$$

As this representation only involves norms, we do not have to store the complete estimate of the covariance kernel (see Remark 3.4 for a more detailed discussion of this property).

3.1. *Weak convergence.* The main results of this section provide the asymptotic properties of the statistic $\widehat{D}_N$ under the null hypothesis of separability and the alternative.

THEOREM 3.1. *Assume that* $\mathbb{E}\|X\|_2^4 < \infty$ *and the null hypothesis of separability holds. Then, we have*

$$\begin{aligned} N\widehat{D}_N \xrightarrow{d} & \left\| \mathscr{G} - \frac{T_2(\mathscr{G}, \Psi) \widetilde{\otimes} T_1(C, T_2(C, \Psi))}{|||T_2(C, \Psi)|||_2^2} \right\|_2^2 \\ & - \frac{|||T_1(\mathscr{G}, T_2(C, \Psi)) - T_1(C, T_2(\mathscr{G}, \Psi))|||_2^2}{|||T_2(C, \Psi)|||_2^2} \\ = & \left\| \mathscr{G} - \frac{T_2(\mathscr{G}, \Psi) \widetilde{\otimes} C_2}{\langle C_2, \Psi\rangle_{S_2(H_2)}} \right\|_2^2 \\ & - \left\| \frac{T_1(\mathscr{G}, C_1)}{|||C_1|||_2} - \frac{\langle C_1, T_2(\mathscr{G}, \Psi)\rangle_{S_2(H_1)} C_2}{\langle C_2, \Psi\rangle_{S_2(H_2)} |||C_1|||_2} \right\|_2^2, \end{aligned} \tag{3.4}$$

where $\mathscr{G}$ *is a centered Gaussian process with covariance operator*

$$\Gamma := \lim_{N\to\infty} \operatorname{Var}(\sqrt{N}\widehat{C}_N) = \operatorname{Var}(X_1 \otimes_o X_1). \tag{3.5}$$

PROOF. The equality in (3.4) follows by a direct calculation using (2.5). For the proof of the first part, define the mapping $\phi : S_2(H) \mapsto \mathbb{R}$ by

$$\phi(A) = |||A|||_2^2 \big|\big|\big| T_2(A, \Psi)\big|\big|\big|_2^2 - \big|\big|\big| T_1\big(A, T_2(A, \Psi)\big)\big|\big|\big|_2^2.$$

By Proposition 5 in [6] the random variable $\sqrt{N}(\widehat{C}_N - C)$ converges in distribution to a centered Gaussian random element $\mathscr{G}$ with variance (3.5) in $S_2(H)$ with respect to Hilbert–Schmidt topology, and we will derive the asymptotic distribution of $\phi(\widehat{C}_N) - \phi(C)$ using von Mises calculus as described in Section 20.1 in [23]. For this purpose we determine the derivatives of the map $\phi_{C,G} : t \mapsto \phi(C + tG)$ for any fixed $G \in S_2(H)$. Note that $\phi_{C,G}(t)$ is a polynomial in t. More precisely, we have

$$\begin{aligned} \phi(C + tG) = & |||C + tG|||_2^2 \big|\big|\big| T_2(C + tG, \Psi)\big|\big|\big|_2^2 \\ & - \big|\big|\big| T_1\big(C + tG, T_2(C + tG, \Psi)\big)\big|\big|\big|_2^2 \\ = & \big(a_0 + a_1 t + a_2 t^2\big)\big(c_0 + c_1 t + c_2 t^2\big) \\ & - \big(b_0 + b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4\big), \end{aligned} \tag{3.6}$$

where

$$\begin{aligned} & a_0 = |||C|||_2^2, \qquad a_1 = 2\langle C, G\rangle_{\mathrm{HS}}, \qquad a_2 = |||G|||_2^2, \\ & c_0 = \big|\big|\big| T_2(C, \Psi)\big|\big|\big|_2^2, \qquad c_1 = 2\big\langle T_2(C, \Psi), T_2(G, \Psi)\big\rangle_{S_2(H_1)}, \\ & c_2 = \big|\big|\big| T_2(G, \Psi)\big|\big|\big|_2^2 \end{aligned}$$

and

$$\begin{aligned}
b_0 &= |||T_1(C, T_2(C, \Psi))|||_2^2, \qquad b_4 = |||T_1(G, T_2(G, \Psi))|||_2^2,\\
b_1 &= 2\big[\langle T_1(C, T_2(C, \Psi)), T_1(C, T_2(G, \Psi))\rangle_{S_2(H_2)}\\
&\quad + \langle T_1(C, T_2(C, \Psi)), T_1(G, T_2(C, \Psi))\rangle_{S_2(H_2)}\big],\\
b_2 &= \big[2\langle T_1(C, T_2(C, \Psi)), T_1(G, T_2(G, \Psi))\rangle_{S_2(H_2)} + |||T_1(C, T_2(G, \Psi))|||_2^2\\
&\quad + |||T_1(G, T_2(C, \Psi))|||_2^2 + 2\langle T_1(G, T_2(C, \Psi)), T_1(C, T_2(G, \Psi))\rangle_{S_2(H_2)}\big],\\
b_3 &= 2\big[\langle T_1(G, T_2(G, \Psi)), T_1(C, T_2(G, \Psi))\rangle_{S_2(H_2)}\\
&\quad + \langle T_1(G, T_2(G, \Psi)), T_1(G, T_2(C, \Psi))\rangle_{S_2(H_2)}\big].
\end{aligned}$$

Now, under the null hypothesis of separability we have for the quantity in (2.12) $D_{C_1} = 0$ and $T_2(C, \Psi) = \langle C_2, \Psi\rangle_{S_2(H_2)} C_1$ which implies for the constant term in the polynomial $\phi(C + tG)$

$$\phi(C + tG)|_{t=0} = \phi(C) = a_0 c_0 - b_0 = 0. \tag{3.7}$$

Similarly, using the fact that $C = C_1 \widetilde{\otimes} C_2$ and $\frac{|||T_1(C, T_2(C,\Psi))|||_2^2}{|||T_2(C,\Psi)|||_2^2} = |||C|||_2^2$, it follows that

$$\begin{aligned}
&\langle T_1(C, T_2(C, \Psi)), T_1(C, T_2(G, \Psi))\rangle_{S_2(H_2)}\\
&\quad = \langle C_2, \Psi\rangle_{S_2(H_2)} \langle T_1(C, C_1), T_1(C, T_2(G, \Psi))\rangle_{S_2(H_2)}\\
&\quad = \langle C_2, \Psi\rangle_{S_2(H_2)} \langle C_1, C_1\rangle_{S_2(H_1)} \langle C_2, T_1(C, T_2(G, \Psi))\rangle_{S_2(H_2)}\\
&\quad = \langle C_2, \Psi\rangle_{S_2(H_2)} \langle C_1, C_1\rangle_{S_2(H_1)} \langle C_2, C_2\rangle_{S_2(H_2)} \langle C_1, T_2(G, \Psi)\rangle_{S_2(H_1)}\\
&\quad = |||C|||_2^2 \langle C_2, \Psi\rangle_{S_2(H_2)} \langle C_1, T_2(G, \Psi)\rangle_{S_2(H_1)}\\
&\quad = |||C|||_2^2 \langle T_2(C, \Psi), T_2(G, \Psi)\rangle_{S_2(H_1)}
\end{aligned}$$

and

$$\begin{aligned}
&\langle T_1(C, T_2(C, \Psi)), T_1(G, T_2(C, \Psi))\rangle_{S_2(H_2)}\\
&\quad = \langle C_2, \Psi\rangle_{S_2(H_2)}^2 \langle T_1(C, C_1), T_1(G, C_1)\rangle_{S_2(H_2)}\\
&\quad = \langle C_1, C_1\rangle_{S_2(H_1)} \langle C_2, T_1(G, C_1)\rangle_{S_2(H_2)} \langle C_2, \Psi\rangle_{S_2(H_2)}^2\\
&\quad = \langle C_2, T_1(G, C_1)\rangle_{S_2(H_2)} |||T_2(C, \Psi)|||_2^2\\
&\quad = \langle G, C_1 \widetilde{\otimes} C_2\rangle_{\mathrm{HS}} |||T_2(C, \Psi)|||_2^2 = \langle G, C\rangle_{\mathrm{HS}} |||T_2(C, \Psi)|||_2^2,
\end{aligned}$$

which implies for the linear term in the polynomial $\phi(C + tG)$

$$\frac{d}{dt}\phi(C + tG)|_{t=0} = a_1 c_0 + a_0 c_1 - b_1 = 0 \tag{3.8}$$

(under the null hypothesis). Next, we look at the second derivative and note the identities

$$\begin{aligned}
&|||C|||_2^2 |||T_2(G, \Psi)|||_2^2\\
&\quad = \langle C_1, C_1\rangle_{S_2(H_1)} \langle C_2, C_2\rangle_{S_2(H_2)} \langle T_2(G, \Psi), T_2(G, \Psi)\rangle_{S_2(H_1)}\\
&\quad = \langle C_1, C_1\rangle_{S_2(H_1)} \langle T_2(G, \Psi) \widetilde{\otimes} C_2, T_2(G, \Psi) \widetilde{\otimes} C_2\rangle_{\mathrm{HS}}
\end{aligned}$$

$$= \frac{\langle C_2, \Psi\rangle^2_{S_2(H_2)} \langle C_1, C_1\rangle^2_{S_2(H_1)} \langle T_2(G,\Psi) \widetilde{\otimes} C_2, T_2(G,\Psi) \widetilde{\otimes} C_2\rangle_{\mathrm{HS}}}{|||T_2(C,\Psi)|||_2^2}$$

$$= \frac{\langle T_2(G,\Psi) \widetilde{\otimes} T_1(C, T_2(C,\Psi)), T_2(G,\Psi) \widetilde{\otimes} T_1(C, T_2(C,\Psi))\rangle_{\mathrm{HS}}}{|||T_2(C,\Psi)|||_2^2},$$

$$\begin{aligned}
&\langle T_1(C, T_2(G,\Psi)), T_1(G, T_2(C,\Psi))\rangle_{S_2(H_2)} \\
&= \langle C_2 \langle C_1, T_2(G,\Psi)\rangle_{S_2(H_1)}, T_1(G, C_1)\langle C_2, \Psi\rangle_{S_2(H_2)}\rangle_{S_2(H_2)} \\
&= \langle C_1, T_2(G,\Psi)\rangle_{S_2(H_1)} \langle C_2, \Psi\rangle_{S_2(H_2)} \langle C_2, T_1(G, C_1)\rangle_{S_2(H_2)} \\
&= \langle \langle C_2, \Psi\rangle_{S_2(H_2)} C_1, T_2(G,\Psi)\rangle_{S_2(H_1)} \langle C_1 \widetilde{\otimes} C_2, G\rangle_{\mathrm{HS}} \\
&= \langle T_2(C,\Psi), T_2(G,\Psi)\rangle_{S_2(H_1)} \langle C, G\rangle_{\mathrm{HS}},
\end{aligned}$$

and

$$\begin{aligned}
&\langle T_1(C, T_2(C,\Psi)), T_1(G, T_2(G,\Psi))\rangle_{S_2(H_2)} \\
&= \langle C_2, \Psi\rangle_{S_2(H_2)} \langle C_1, C_1\rangle_{S_2(H_1)} \langle C_2, T_1(G, T_2(G,\Psi))\rangle_{S_2(H_2)} \\
&= \langle C_2, \Psi\rangle_{S_2(H_2)} \langle C_1, C_1\rangle_{S_2(H_1)} \langle G, T_2(G,\Psi) \widetilde{\otimes} C_2\rangle_{\mathrm{HS}} \\
&= \langle G, T_2(G,\Psi) \widetilde{\otimes} (\langle C_2, \Psi\rangle_{S_2(H_2)} \langle C_1, C_1\rangle_{S_2(H_1)} C_2)\rangle_{\mathrm{HS}} \\
&= \langle G, T_2(G,\Psi) \widetilde{\otimes} T_1(C, T_2(C,\Psi))\rangle_{\mathrm{HS}}
\end{aligned}$$

(here we make extensive use of Proposition 2.1). This gives for the quadratic term in the polynomial $\phi(C + tG)$

$$\begin{aligned}
&\frac{1}{2}\frac{d^2}{d^2 t}\phi(C + tG)|_{t=0} \\
&= a_0 c_2 + a_1 c_1 + a_2 c_0 - b_2 \\
&= \left|\left|\left| G |||T_2(C,\Psi)|||_2 - \frac{T_2(G,\Psi) \widetilde{\otimes} T_1(C, T_2(C,\Psi))}{|||T_2(C,\Psi)|||_2} \right|\right|\right|_2^2 \\
&\quad - |||T_1(G, T_2(C,\Psi)) - T_1(C, T_2(G,\Psi))|||_2^2.
\end{aligned} \tag{3.9}$$

Finally, taking $G := \sqrt{N}(\widehat{C}_N - C)$, $t = 1/\sqrt{N}$ and expanding $\phi(C + tG)$ in powers of t, we obtain (note that $\phi(C) = 0$ under the null hypothesis and that the terms corresponding to t^3 and t^4 in (3.6) are at least of order $N^{-3/2}$)

$$\begin{aligned}
N\phi(\widehat{C}_N) &= \frac{1}{2}\frac{d^2}{d^2 t}\phi(C + t\sqrt{N}(\widehat{C}_N - C))|_{t=0} + o_p(1) \\
&\xrightarrow{d} \left|\left|\left| \mathscr{G} |||T_2(C,\Psi)|||_2 - \frac{T_2(\mathscr{G},\Psi) \widetilde{\otimes} T_1(C, T_2(C,\Psi))}{|||T_2(C,\Psi)|||_2} \right|\right|\right|_2^2 \\
&\quad - |||T_1(\mathscr{G}, T_2(C,\Psi)) - T_1(C, T_2(\mathscr{G},\Psi))|||_2^2.
\end{aligned}$$

Therefore, Theorem 3.1 follows from Slutsky's Lemma noting that

$$\widehat{D}_N = \frac{\phi(\widehat{C}_N) - \phi(C)}{|||T_2(\widehat{C}_N,\Psi)|||_2^2} = \frac{\phi(\widehat{C}_N)}{|||T_2(\widehat{C}_N,\Psi)|||_2^2}.$$

□

In the following let $q_{1-\alpha}$ be the $100\alpha\%$ quantile of the limiting random variable in Theorem 3.1, then an asymptotic level α test for the hypothesis in (2.1) is obtained by rejecting the null hypothesis of separability, whenever

$$N\widehat{D}_N > q_{1-\alpha}. \tag{3.10}$$

The next result provides the asymptotic distribution under the alternative and yields as a consequence the consistency of this test.

THEOREM 3.2. *If $\mathbb{E}\|X\|_2^4 < \infty$, then the statistic $\sqrt{N}(\widehat{D}_N - D_0)$ converges in distribution to a centered normal distribution with variance*

$$\nu^2 := 4\langle\Gamma(A-B), (A-B)\rangle_{\mathrm{HS}}, \tag{3.11}$$

where

$$A = C - \frac{T_2(C,\Psi)\,\widetilde{\otimes}\,T_1(C,T_2(C,\Psi))}{|||T_2(C,\Psi)|||_2^2},$$

$$B = \frac{1}{|||T_2(C,\Psi)|||_2^2}\Big[T_2\big(C,T_1\big(C,T_2(C,\Psi)\big)\big)\,\widetilde{\otimes}\,\Psi - \frac{|||T_1(C,T_2(C,\Psi))|||_2^2}{|||T_2(C,\Psi)|||_2^2}T_2(C,\Psi)\,\widetilde{\otimes}\,\Psi\Big]$$

and the centering term D_0 is defined in (2.12).

PROOF. Observing (2.12) and (3.3) we write

$$\sqrt{N}(\widehat{D}_N - D_0) = \sqrt{N}\bigg(|||\widehat{C}_N|||_2^2 - \frac{|||T_1(\widehat{C}_N,T_2(\widehat{C}_N,\Psi))|||^2}{|||T_2(\widehat{C}_N,\Psi)|||_2^2} - |||C|||_2^2 + \frac{|||T_1(C,T_2(C,\Psi))|||^2}{|||T_2(C,\Psi)|||_2^2}\bigg) \tag{3.12}$$

and note that $\widehat{D}_N$ and D_0 are functions of the random variables

$$G_N = \big(|||\widehat{C}_N|||_2^2, \big|\big|\big|T_1\big(\widehat{C}_N,T_2(\widehat{C}_N,\Psi)\big)\big|\big|\big|^2, \big|\big|\big|T_2(\widehat{C}_N,\Psi)\big|\big|\big|_2^2\big)^T, \tag{3.13}$$

$$G = \big(|||C|||_2^2, \big|\big|\big|T_1\big(C,T_2(C,\Psi)\big)\big|\big|\big|^2, \big|\big|\big|T_2(C,\Psi)\big|\big|\big|_2^2\big)^T, \tag{3.14}$$

respectively. Therefore, we first investigate the weak convergence of the vector $\sqrt{N}(G_N - G)$. For this purpose we note that for $K, L \in S_2(H)$, the identity

$$|||K|||_2^2 - |||L|||_2^2 = |||K-L|||_2^2 + 2\langle K-L, L\rangle_{\mathrm{HS}}$$

holds and introduce the decomposition

$$\sqrt{N}(G_N - G) = H_N^{(1)} + H_N^{(2)},$$

where the random variables $H_N^{(1)}$ and $H_N^{(2)}$ are defined by

$$\begin{aligned} H_N^{(1)} &= \sqrt{N}\big(|||\widehat{C}_N - C|||_2^2, \big|\big|\big|T_1\big(\widehat{C}_N,T_2(\widehat{C}_N,\Psi)\big) - T_1\big(C,T_2(C,\Psi)\big)\big|\big|\big|_2^2, \\ &\qquad \big|\big|\big|T_2(\widehat{C}_N,\Psi) - T_2(C,\Psi)\big|\big|\big|_2^2\big)^T, \\ H_N^{(2)} &= 2\sqrt{N}\big(\langle\widehat{C}_N - C, C\rangle_{\mathrm{HS}}, \\ &\qquad \big\langle T_1\big(\widehat{C}_N,T_2(\widehat{C}_N,\Psi)\big) - T_1\big(C,T_2(C,\Psi)\big), T_1\big(C,T_2(C,\Psi)\big)\big\rangle_{\mathrm{HS}}, \\ &\qquad \big\langle T_2(\widehat{C}_N,\Psi) - T_2(C,\Psi), T_2(C,\Psi)\big\rangle_{\mathrm{HS}}\big)^T. \end{aligned}$$

Using the linearity of T_1 and T_2, we further write

$$
\begin{aligned}
&T_1(\widehat{C}_N, T_2(\widehat{C}_N, \Psi)) - T_1(C, T_2(C, \Psi)) \\
&\quad = T_1(\widehat{C}_N, T_2(\widehat{C}_N, \Psi)) - T_1(C, T_2(\widehat{C}_N, \Psi)) \\
&\qquad + T_1(C, T_2(\widehat{C}_N, \Psi)) - T_1(C, T_2(C, \Psi)) \\
&\quad = T_1(\widehat{C}_N - C, T_2(\widehat{C}_N, \Psi)) + T_1(C, T_2(\widehat{C}_N - C, \Psi))
\end{aligned}
$$

and obtain the representation

$$
\begin{aligned}
H_N^{(1)} &= \frac{1}{\sqrt{N}} \begin{pmatrix} \|\sqrt{N}(\widehat{C}_N - C)\|_2^2 \\ \|T_1(\sqrt{N}(\widehat{C}_N - C), T_2(\widehat{C}_N, \Psi)) + T_1(C, T_2(\sqrt{N}(\widehat{C}_N - C), \Psi))\|_2^2 \\ \|T_2(\sqrt{N}(\widehat{C}_N - C), \Psi)\|_2^2 \end{pmatrix} \\
&=: \frac{1}{\sqrt{N}} F_1(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N), \\
H_N^{(2)} &= 2 \begin{pmatrix} \langle \sqrt{N}(\widehat{C}_N - C), C\rangle_{\mathrm{HS}} \\ \langle T_1(\sqrt{N}(\widehat{C}_N - C), T_2(\widehat{C}_N, \Psi)) + T_1(C, T_2(\sqrt{N}(\widehat{C}_N - C), \Psi)), T_1(C, T_2(C, \Psi))\rangle_{\mathrm{HS}} \\ \langle T_2(\sqrt{N}(\widehat{C}_N - C), \Psi), T_2(C, \Psi)\rangle_{\mathrm{HS}} \end{pmatrix} \\
&=: F_2(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N),
\end{aligned}
$$

where the last equations define the functions F_1 and F_2 in an obvious manner. Note that $F := (F_1, F_2) : S_2(H) \times S_2(H) \mapsto \mathbb{R}^6$ is composition of continuous functions and hence continuous. By Proposition 5 in [6], the random variable $\sqrt{N}(\widehat{C}_N - C)$ converges in distribution to a centered Gaussian random element $\mathscr{G}$ with variance (3.5) in $S_2(H)$ with respect to Hilbert–Schmidt topology. Therefore, using continuous mapping arguments, we have $F(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N) \xrightarrow{d} F(\mathscr{G}, C)$, and consequently

$$
\begin{aligned}
\sqrt{N}(G_N - G) &= \frac{1}{\sqrt{N}} F_1(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N) + F_2(\sqrt{N}(\widehat{C}_N - C), \widehat{C}_N) \\
&\xrightarrow{d} F_2(\mathscr{G}, C).
\end{aligned}
$$

We write

$$
F_2(\mathscr{G}, C) = 2 \begin{pmatrix} \langle \mathscr{G}, C\rangle_{\mathrm{HS}} \\ \langle T_1(\mathscr{G}, T_2(C, \Psi)) + T_1(C, T_2(\mathscr{G}, \Psi)), T_1(C, T_2(C, \Psi))\rangle_{\mathrm{HS}} \\ \langle T_2(\mathscr{G}, \Psi), T_2(C, \Psi)\rangle_{\mathrm{HS}} \end{pmatrix},
$$

which can be further simplified as

$$
\begin{aligned}
&F_2(\mathscr{G}, C) \\
&\quad = 2 \begin{pmatrix} \langle \mathscr{G}, C\rangle_{\mathrm{HS}} \\ \langle \mathscr{G}, T_2(C, \Psi) \widetilde{\otimes} T_1(C, T_2(C, \Psi))\rangle_{\mathrm{HS}} + \langle C, T_2(\mathscr{G}, \Psi) \widetilde{\otimes} T_1(C, T_2(C, \Psi))\rangle_{\mathrm{HS}} \\ \langle \mathscr{G}, T_2(C, \Psi) \widetilde{\otimes} \Psi\rangle_{\mathrm{HS}} \end{pmatrix}.
\end{aligned}
\tag{3.15}
$$

By Proposition 2.1 $T_2(\mathscr{G}, \Psi)$ is a Gaussian process in $S_2(H_2)$. This fact along with Lemma B.3 in the Supplementary Material imply that $F_2(\mathscr{G}, C)$ is a normal distributed random vector with mean zero and covariance matrix, say Σ. By (3.12),

$$
\sqrt{N}(\widehat{D}_N - D_0) = \sqrt{N}(f(G_N) - f(G)),
$$

where the function $f : \mathbb{R}^3 \mapsto \mathbb{R}$ is defined by $f(x, y, z) = x - y/z$ and G_N and G are defined in (3.13) and (3.14), respectively. Therefore, using the delta method and the fact that

$$\mathbb{P}(|||T_2(\widehat{C}_N, \Psi)|||_2^2 > 0) \to \mathbb{P}(|||T_2(C, \Psi)|||_2^2 > 0) = 1$$

as $|||C|||_2 \neq 0$, we finally obtain

$$\sqrt{N}(\widehat{D}_N - D_0) \xrightarrow{d} N(0, (\nabla f(G))^T \Sigma(\nabla f(G))) \tag{3.16}$$

as $n \to \infty$ where $\nabla f(x, y, z) = (1, -1/z, y/z^2)^T$ denotes the gradient of the function f. Finally, the proof of the representation (3.11) of the limiting variance is given in Section B.6 of the Supplementary Material. □

REMARK 3.1. If the null hypothesis is true, that is, $C = C_1 \widetilde{\otimes} C_2$, the variance ν^2 in Theorem 3.2 becomes zero. Indeed, under the null hypothesis of separability we have

$$\begin{aligned} T_2(C, \Psi) \widetilde{\otimes} T_1(C, T_2(C, \Psi)) &= \langle C_2, \Psi\rangle^2 C_1 \widetilde{\otimes} T_1(C, C_1) \\ &= \langle C_2, \Psi\rangle^2 \langle C_1, C_1\rangle C_1 \widetilde{\otimes} C_2 \\ &= |||\langle C_2, \Psi\rangle C_1|||_2^2 C = |||T_2(C, \Psi)|||_2^2 C, \end{aligned}$$

which implies $A = 0$ for the quantity A in Theorem 3.2. Similarly,

$$\begin{aligned} &\frac{|||T_1(C, T_2(C, \Psi))|||_2^2}{|||T_2(C, \Psi)|||_2^2} T_2(C, \Psi) \\ &= \frac{|||T_1(C, C_1)|||_2^2 \langle C_2, \Psi\rangle^2}{|||T_2(C, \Psi)|||_2^2} T_2(C, \Psi) \\ &= \frac{|||\langle C_1, C_1\rangle C_2|||_2^2 \langle C_2, \Psi\rangle^2}{|||\langle C_2, \Psi\rangle C_1|||_2^2} T_2(C, \Psi) = \frac{|||C_2|||_2^2 |||C_1|||_2^4 \langle C_2, \Psi\rangle^2}{|||C_1|||_2^2 \langle C_2, \Psi\rangle^2} T_2(C, \Psi) \\ &= |||C_2|||_2^2 |||C_1|||_2^2 T_2(C, \Psi) = \langle C_2, C_2\rangle \langle C_1, C_1\rangle \langle C_2, \Psi\rangle C_1 \\ &= \langle C_1, C_1\rangle \langle C_2, \Psi\rangle T_2(C, C_2) = T_2(C, C_2\langle C_1, C_1\rangle)\langle C_2, \Psi\rangle \\ &= T_2(C, T_1(C, C_1))\langle C_2, \Psi\rangle \\ &= T_2(C, T_1(C, \langle C_2, \Psi\rangle C_1)) = T_2(C, T_1(C, T_2(C, \Psi))) \end{aligned}$$

and consequently the quantity B in Theorem 3.2 also vanishes. Therefore, under the null hypothesis $\sqrt{N}\widehat{D}_N \xrightarrow{p} 0$ (which is also a consequence of Theorem 3.1).

REMARK 3.2. A sufficient condition for Theorems 3.1 and 3.2 to hold is $\mathbb{E}\|X\|_2^4 < \infty$. As indicated in Remark 2.2(1) of [1], this is a weaker assumption than Condition 2.1 in their paper, which assumes $\sum_{j=1}^{\infty} (\mathbb{E}[\langle X, e_j\rangle^4])^{1/4} < \infty$, for some orthonormal basis $(e_j)_{j\geq 1}$ of H. These authors used this assumption to prove weak convergence under the trace-norm topology which is required to establish Theorem 2.3 in [1]. In contrast, the proof of Theorem 3.2 here only requires weak convergence under the Hilbert–Schmidt topology which defines a weaker topology.

REMARK 3.3. Note that the asymptotic distribution depends (under the null hypothesis and alternative) on the operator Ψ. Under the assumptions of Theorem 3.2, we obtain an approximation of the power of the test (3.10) by

$$\mathbb{P}(N\widehat{D}_N > q_{1-\alpha}) = \mathbb{P}\Big(\sqrt{N}(\widehat{D}_N - D_0) > \frac{q_{1-\alpha}}{\sqrt{N}} - \sqrt{N}D_0\Big)$$
$$\approx 1 - \Phi\Big(\frac{q_{1-\alpha}}{\sqrt{N}\nu} - \frac{\sqrt{N}D_0}{\nu}\Big),$$

where Φ is the standard normal distribution function and ν^2 is defined by (3.11). Under the alternative D_0 is positive. Therefore the rejection probability converges to 1 with increasing sample size N and, consequently, the proposed test is consistent.

Moreover, if N is sufficiently large, the power is a decreasing function of the variance ν^2 in (3.11). As this quantity depends on the operator Ψ, it is desirable to choose Ψ such that υ^2 is minimal. The solution of this optimization problem depends on the unknown covariance operator C, and it seems to be intractable to obtain it explicitly. However, we will demonstrate in Section 4 that for finite sample sizes the resulting tests are not very sensitive with respect to the choice of the operator Ψ.

3.2. *Hilbert–Schmidt integral operators.* In the remaining part of this section, we concentrate on the case where X is a random element in $H = L^2(S \times T)$ and $S \subset \mathbb{R}^p$ and $T \subset \mathbb{R}^q$ are bounded measurable sets. In this particular scenario we choose Ψ also to be an integral operator generated by a kernel ψ. With this choice, using the explicit formula for the operator T_1 described in Proposition 2.3, the minimum distance can be expressed in terms of the corresponding kernels, that is,

$$\begin{aligned} D_0 &= D\big(T_2(C,\Psi)\big) \\ &= \int_T\int_T\int_S\int_S c^2(s,t,s',t')\,ds\,ds'\,dt\,dt' \\ &\quad - \frac{\int_T\int_T[\int_S\int_S c(s,t,s't')\tilde{c}_1(s,s')\,ds\,ds']^2\,dt\,dt'}{\int_S\int_S \tilde{c}_1^2(s,s')\,ds\,ds'}, \end{aligned} \tag{3.17}$$

where $\tilde{c}_1$ denotes the kernel corresponding to the operator $T_2(C,\Psi)$, that is,

$$\tilde{c}_1(s,s') = \int_T\int_T c(s,t,s',t')\psi(t,t')\,dt\,dt'.$$

In this case the estimator $\widehat{C}_N$ defined in (3.1) is induced by the kernel

$$\hat{c}_N(s,t,s',t') = \frac{1}{N}\sum_{i=1}^N \big(X_i(s,t) - \overline{X}(s,t)\big)\big(X_i(s',t') - \overline{X}(s',t')\big),$$

and the estimator $\widehat{C}_{1N} = T_2(\widehat{C}_N,\Psi)$ defined in (3.2) is induced by the kernel

$$\hat{c}_{1N}(s,s') = \int_T\int_T \hat{c}_N(s,t,s',t')\psi(t,t')\,dt\,dt'.$$

The estimator $\widehat{D}_N$ of D_0 is calculated by plugging in $\hat{c}_N$ and $\hat{c}_{1N}$ to the expression in (3.17).

REMARK 3.4. (a) A natural choice for Ψ is an operator with constant kernel, that is, $\psi(t,t') \equiv 1$ which gives $\int_T\int_T c(s,t,s',t')\,dt\,dt'$ for the kernel of the operator $T_1(C,\Psi)$. This operator has to be distinguished form partial trace which is defined as the integral operator with kernel $\int_T c(s,t,s',t)\,dt$, and was used by [1].

(b) Although the proposed estimator is based on the norm of the complete covariance kernel c, numerically we do not need to store the complete covariance kernel. For example, we obtain for the first term of the statistic $\widehat{D}_N$ the representation

$$\begin{aligned}
&|||\widehat{C}_N|||_2^2 \\
&\quad = \frac{1}{N^2}\int_T\int_S\int_T\int_S\left(\sum_{i=1}^{N}\big(X_i(s,t)-\overline{X}(s,t)\big)\big(X_i(s',t')-\overline{X}(s',t')\big)\right)^2 ds\,dt\,ds'\,dt' \\
&\quad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\int_T\int_S\big(X_i(s,t)-\overline{X}(s,t)\big)\big(X_j(s,t)-\overline{X}(s,t)\big)\,ds\,dt\right]^2.
\end{aligned}$$

All other terms of the estimator in (3.3) can be represented similarly using simple matrix operations on the data matrix without storing the full or marginal covariance kernels.

3.3. *Bootstrap test for separability.* An obvious method for testing the hypothesis of separability is based on the quantiles of the limiting random variable given in Theorem 3.1. For this purpose one can estimate the limiting covariance operator Γ from the data and simulate a centered Gaussian process $\mathscr{G}$ with covariance operator Γ. The limiting distribution can then be calculated as function of the simulated Gaussian processes. The simulated $100(1-\alpha)\%$ quantile is finally compared to $N\widehat{D}_N$ to test the null hypothesis of separability which gives the test (3.10). It turns out that this approach provides a very powerful test for the hypothesis of separability (see the empirical results in Section 4).

As this method requires the estimation of the covariance kernel Γ, we also propose a bootstrap test. The simplest method would be to approximate the limiting distribution of $N\widehat{D}_N$ by the distribution of the statistic $\{N\hat{D}_N^* - N\widehat{D}_N\}$, where $\hat{D}_N^*$ is the test statistic calculated from a bootstrap sample drawn from $X_1, \dots, X_N$ with replacement.

However, this procedure fails to give good power under the alternative. This observation can be explained by studying the test statistic a little more closely. In general, we can write

$$\begin{aligned}
\widehat{D}_N - D_0 &= |||\widehat{C}_N|||_2^2 - \frac{|||T_1(\widehat{C}_N, T_2(\widehat{C}_N,\Psi))|||_2^2}{|||T_2(\widehat{C}_N,\Psi)|||_2^2} \\
&\qquad - |||C|||_2^2 + \frac{|||T_1(C, T_2(C,\Psi))|||_2^2}{|||T_2(C,\Psi)|||_2^2} \\
&= A_{1,N} + A_{2,N},
\end{aligned}$$

where the statistics $A_{1,N}$ and $A_{2,N}$ are given by

$$\begin{aligned}
A_{1,N} &= |||\widehat{C}_N - C|||_2^2 \\
&\qquad - \frac{|||T_1(\widehat{C}_N - C, T_2(\widehat{C}_N,\Psi))|||_2^2 + |||T_1(C, T_2(\widehat{C}_N - C,\Psi))|||_2^2}{|||T_2(\widehat{C}_N,\Psi)|||_2^2} \qquad (3.18)\\
&\qquad + \frac{|||T_1(C,(T_2(C,\Psi)))|||_2^2}{|||T_2(\widehat{C}_N,\Psi)|||_2^2 |||T_2(C,\Psi)|||_2^2}|||T_2(\widehat{C}_N - C,\Psi)|||_2^2, \\
A_{2,N} &= 2\langle \widehat{C}_N - C, C\rangle_{\mathrm{HS}} \\
&\qquad - \frac{2\langle T_1(\widehat{C}_N - C, T_2(\widehat{C}_N,\Psi)), T_1(C, T_2(\widehat{C}_N,\Psi))\rangle}{|||T_2(\widehat{C}_N,\Psi)|||_2^2} \qquad (3.19)
\end{aligned}$$

$$- \frac{2\langle T_1(C, T_2(\widehat{C}_N - C, \Psi)), T_1(C, T_2(C, \Psi))\rangle_{S_2(H_2)}}{\|\|T_2(\widehat{C}_N, \Psi)\|\|_2^2}$$
$$+ \frac{2\|\|T_1(C, (T_2(C, \Psi)))\|\|_2^2}{\|\|T_2(\widehat{C}_N, \Psi)\|\|_2^2 \|\|T_2(C, \Psi)\|\|_2^2} \langle T_2(\widehat{C}_N - C, \Psi), T_2(C, \Psi)\rangle_{S_2(H_1)},$$

respectively. If the true underlying covariance operator C is separable, then $A_{2,N} = 0$ and, hence, only the first term contributes to the limiting null distribution. Now, note that a similar decomposition for the bootstrap statistic gives

$$D_N^* - \widehat{D}_N = A_{1,N}^* + A_{2,N}^*,$$

where $A_{1,N}^*$ and $A_{2,N}^*$ are defined similarly as in (3.18) and (3.19) replacing $\widehat{C}_N$ by its bootstrap analogue $\widehat{C}_N^*$ and C by $\widehat{C}_N$, respectively. The first term $NA_{1,N}^*$ can be shown to approximate the limiting distribution of $NA_{1,N}$ which is the desired null limiting distribution. However, the estimate $\widehat{C}_N$ is in general not separable. As consequence $NA_{2,N}^*$ is not zero, and a simple bootstrap using the quantile of the distribution of $N(\hat{D}_N^* - \widehat{D}_N)$ will result in a test with very low power.

To avoid this problem, instead of using the quantile of $N(\hat{D}_N^* - \widehat{D}_N)$ we propose to use the quantile of the distribution of $NA_{1,N}^*$. This quantile can be estimated by the empirical quantile from the bootstrap sample $NA_{1,N}^{1*}, \dots, NA_{1,N}^{B*}$ (here, $NA_{1,N}^{b*}$ is the corresponding statistic calculated from the bth bootstrap sample, for $b = 1, \dots, B$).

4. Finite sample properties. In this section we study the finite sample properties of a family of tests for the hypothesis of separability described in Section 3.3 by means of a small simulation study. We also compare the new tests with the tests proposed by [1] and [4] and illustrate potential applications in a data example. For this purpose we have implemented the asymptotic test (3.10) based on simulated quantiles of the random variable appearing in (3.4), the new bootstrap test as described in Section 3.3, the asymptotic and studentized empirical bootstrap test described in [1] and the weighted χ^2 test based on the test statistic $\widehat{T}_F$ as described in Theorem 3 of [4]. The new tests depend on the choice of the operator Ψ, and we will demonstrate that they are not very sensitive with respect to this choice. Both tests proposed by [1] require the specification of the eigensubspace, which was chosen to be $I_k = \{(i, j) : i = 1, \dots, k; j = 1, \dots, k\}$ for $k = 2, 3, 4$, and p-values are obtained by the asymptotic distribution based test and empirical studentized bootstrap. We use the R package "covsep" (see [22] for details) to implement their method. For the tests proposed by [4], we choose the procedure based on the statistic $\widehat{T}_F$ as in a simulation study it turned out to be the most powerful procedure among the four methods proposed in this paper. The test requires the specifications of the number of spatial and temporal principal components which are taken to be equal and the number is chosen to be 2, 3 and 4.

4.1. *Simulation studies.* The data are generated from a zero-mean Gaussian and a t-distribution with five degrees of freedom with the spatiotemporal covariance kernel

$$c(s, t, s', t') = \frac{\sigma^2}{(a|t - t'|^{2\alpha} + 1)^\tau} \exp\left(-\frac{c\|s - s'\|^{2\gamma}}{(a|t - t'|^{2\alpha} + 1)^{\beta\gamma}}\right), \tag{4.1}$$

introduced by [10]. In this covariance function a and c are nonnegative scaling parameters of time and space, respectively; α and γ are smoothness parameters which take values in the interval (0, 1]; β is the separability parameter which varies in the interval [0, 1]; $\sigma^2 > 0$ is the point-wise variance; and $\tau \geq \beta d/2$, where d is the spatial dimension. If $\beta = 0$, the covariance is separable, and the space-time interaction becomes stronger with increasing values of β. We

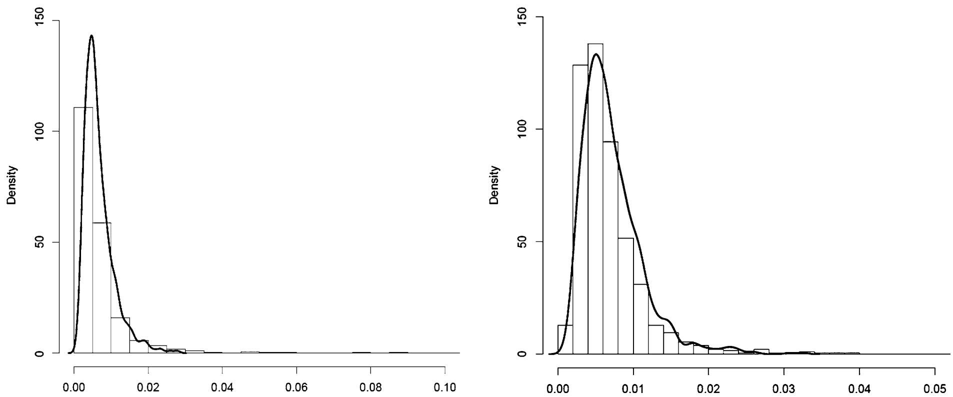

FIG. 1. *The histogram of simulated values of $N\widehat{D}_N$ under H_0 along with the simulated density of the limiting random variable in* (3.4). *The left panel shows the distribution for $N = 100$, and the right panel is for $N = 1000$ observations.*

fix $\gamma = 1$, $\alpha = 1/2$, $\sigma^2 = 1$, $a = 1$, $c = 1$ and $\tau = 1$ in the following discussion and choose different values for the parameter β specifying the level of separability. Further simulation results for a different covariance kernel can be found in Section A of the Supplementary Material.

We generate data at 100 equally spaced time points in [0, 1] and 11 space points on the grid $[0, 1] \times [0, 1]$. The integrals are approximated by an average of the function values at grid points. The nominal significance level is taken to be 5%, and empirical rejection region are computed by 1000 Monte-Carlo replications and 1000 bootstrap samples. In order to estimate the quantiles, the asymptotic test (3.10) we use 1000 simulation runs.

Figure 1 illustrates the convergence of the test statistic under the null hypothesis (see Theorem 3.1) for sample size $N = 100$ and $N = 1000$. We have plotted an histogram of simulated values of the statistic $N\widehat{D}_N$, where the data has been generated from a t-distribution with $five$ degrees of freedom and covariance kernel given in (4.1) with $\beta = 0$. The simulated density of the limiting random variable defined in (3.4) is overlayed in the same figure. To simulate the limit distribution, we used a plug-in-estimator of the covariance operator Γ based on a sample of size 1000. The kernel in the statistic $\widehat{D}_N$ is taken to be constant. We observe a rather good approximation by the bootstrap procedure.

In Figure 2 we investigate the effect of different choices of the operator Ψ on the performance of the methods proposed in this paper. For this purpose we consider three integral operators with the following kernels

$$\psi_1(t, t') \equiv 1, \qquad \psi_2(t, t') = |t - t'|, \qquad \psi_3(t, t') = \exp(-\pi(t^2 + t'^2)).$$

The plots show the empirical rejection probabilities for the tests at 5% level (indicated by a horizontal line in the figure), where the sample size is $N = 100$. We observe that both tests are very robust with respect to the choice of the kernel ψ and that the level is well approximated. Further simulation results which are not presented for the sake of brevity show a similar picture. The power increases consistently as we move away from separability in all cases under consideration. The empirical power is approximately 1 for $\beta = 1$ which corresponds to the case of extreme nonseparability. The asymptotic test (M2) performs better than the bootstrap test (M1). However, the bootstrap test is computationally more efficient and significantly faster than the asymptotic test. For a more detailed discussion of the computational issues, we refer to the Supplementary Material.

In Figure 3 and Figure 4 we compare the power of the new procedures with the tests proposed by [1] and [4]. We observe that the test of [1] based on the asymptotic distribution does

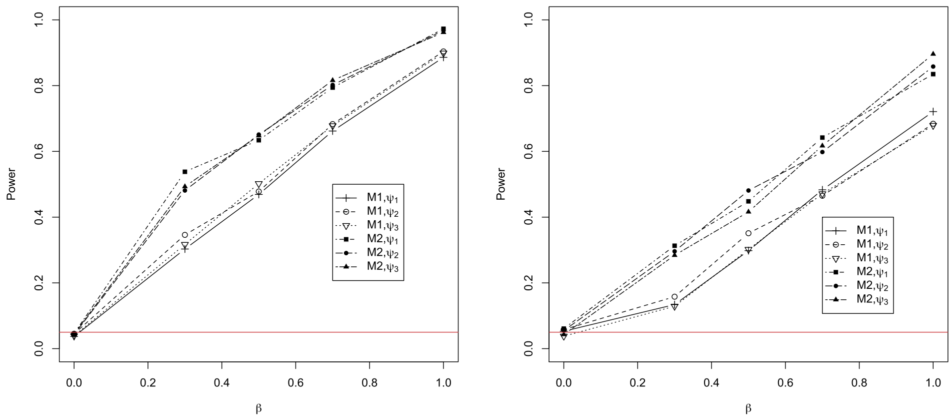

FIG. 2. *Empirical rejection probabilities of the bootstrap test* ($M1$) *and the asymptotic test* (3.10) ($M2$) *proposed in this paper for different choices of kernels* ψ (*level* 5%, *indicated by the horizontal line*). *The data are generated from a zero mean Gaussian distribution* (*left part*) *and zero-mean* t*-distribution with five degrees of freedom* (*right part*). *The covariance kernel is given by* (4.1), *and the sample sizes are* $N = 100$. *The case* $\beta = 0$ *corresponds to the null hypothesis of separability.*

not keep the nominal level and its performance deteriorates with an increasing values of number of eigensubspaces. It performs better for the Gaussian case with $N = 500$ observations, but it still does not provide an accurate approximation of the nominal level.

All other procedures yield rather similar results under the null hypothesis, and in general the nominal level is very well approximated by all tests under consideration. On the other hand, under the alternative we observe more differences. The asymptotic test (3.10) proposed in this paper yields the best power. The power of the bootstrap test of [1] is increasing with the number k of eigensubspaces used in the procedure. This improvement is achieved at the expense of the computing time. A similar observation can be made for the test of [4] with respect to the number of spatial and temporal principal components (see Table 3 in the Supplementary Material for a more detailed discussion of the computation time of the different tests).

The results of the bootstrap test proposed in this paper and the test of [1] are similar if the latter is used with $k = 2$ subspaces, but the test of [1] is more powerful for $k = 4$. The bootstrap test proposed in this paper has more power than the test of [4] with $L = J = 2$ spatial and temporal principal components. For the choice $L = J = 4$ the test [4] shows a slightly better performance.

In general, an improvement in power is always achieved at a cost of computational time, and we refer to the Supplementary Material for a more detailed discussion.

4.2. *Application to real data*. We apply our new methods to the acoustic phonetic data discussed in [1]. This data set has been compiled in the Phonetics Laboratory of the University of Oxford between 2012–2013. It consists of natural speech recordings of five languages: French, Italian, Portuguese, American Spanish and Castilian Spanish. The speakers utter the numbers one to 10 in their native language. The data set consists of a sample of 219 recordings by 23 speakers. More information about this data and related project can be found on the website http://www.phon.ox.ac.uk/ancient_sounds. We use the preprocessed data used in [1]. In that paper the data was transformed to a smoothed log-spectogram through a short-time Fourier transformation using a Gaussian window function with window-size 10 milliseconds. The log-spectograms were demeaned separately for each language. We employed the bootstrap test with $B = 1000$ (M1), and the test based on the simulated quantiles of the asymptotic distribution of $N\widehat{D}_N$ as appeared in Theorem 3.1 (M2) on the dataset for three choices

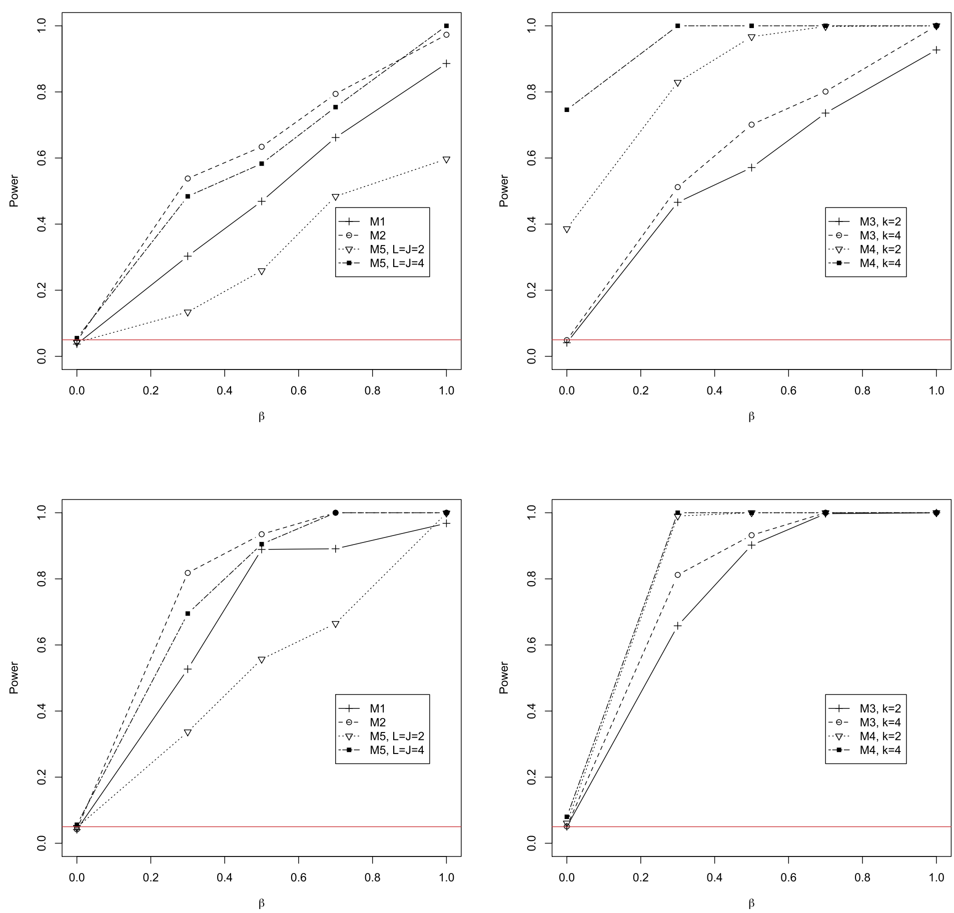

FIG. 3. *Empirical rejection probabilities of different methods for testing the hypothesis of separability* (*level* 5%, *indicated by the horizontal line*). *M*1: *the bootstrap test proposed in this paper using kernel* ψ_1, *M*2: *the asymptotic test* (3.10) *proposed in this paper using the kernel* ψ_1, *M*3: *the empirical bootstrap test* (*studentized*) *proposed by* [1] *M*4: *the asymptotic test proposed by* [1], *M*5: *the test proposed in* [4]. *The data are generated from a Gaussian distribution with zero mean and covariance kernel* (4.1). *The case* $\beta = 0$ *corresponds to the null hypothesis of separability. Upper part*: $N = 100$; *lower part*: $N = 500$.

of kernels as mentioned in Section 4.1. The results are presented in Table 1. The hypothesis of separability is clearly rejected.

Acknowledgments. The authors are grateful to Juan Cuesta Albertos, Shahin Tavakoli for very helpful discussions and to Martina Stein who typed parts of this paper with considerable technical expertise. The authors would also like to thank three anonymous referees and the Associate Editor for their constructive comments.

This work was supported in part by the Collaborative Research Center "Statistical modelling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1, C1) of the German Research Foundation (DFG).

SUPPLEMENTARY MATERIAL

Supplement to "A test for separability in covariance operators of random surfaces" (DOI: 10.1214/19-AOS1888SUPP; .pdf). We provide additional simulation results and technical details for Sections 2 and 3 in the supplement.

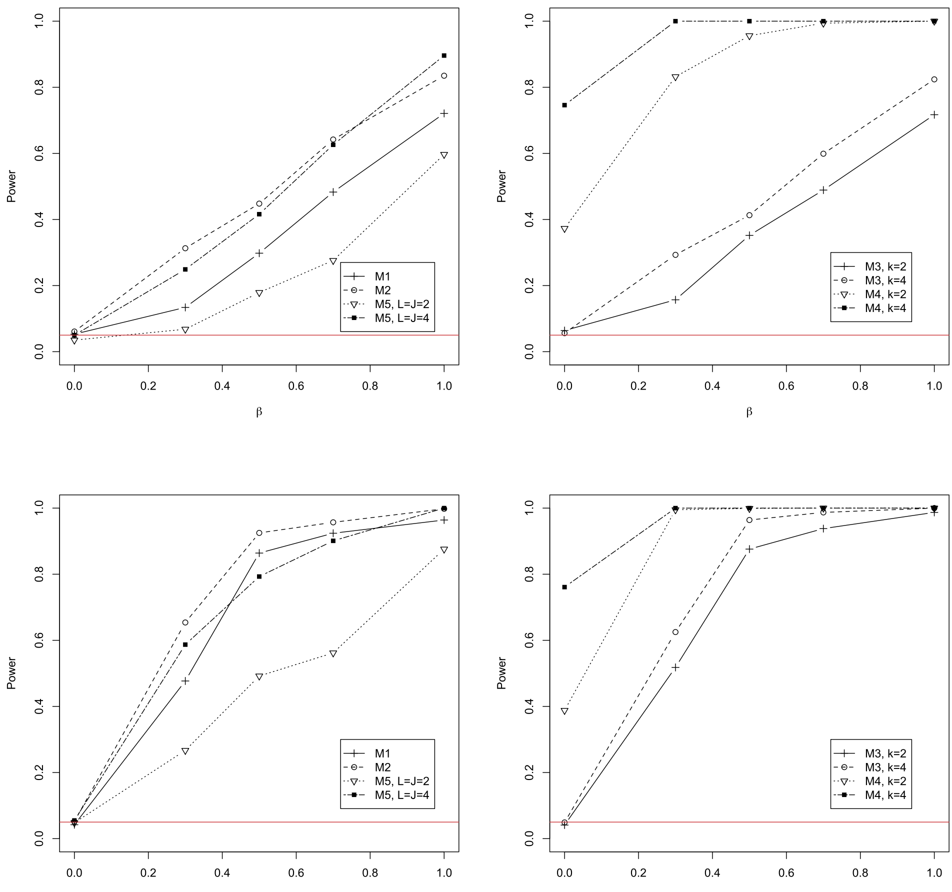

FIG. 4. *Empirical rejection probabilities of different methods for testing the hypothesis of separability* (*level* 5%, *indicated by the horizontal line*). *M*1: *the bootstrap test proposed in this paper using kernel* ψ_1, *M*2: *the asymptotic test* (3.10) *proposed in this paper using the kernel* ψ_1, *M*3: *the empirical bootstrap test* (*studentized*) *proposed by* [1] *M*4: *the asymptotic test proposed by* [1], *M*5: *the test proposed in* [4]. *The data are generated from a* t-*distribution with* $five$ *degrees of freedom and covariance kernel* (4.1). *The case* $\beta = 0$ *corresponds to the null hypothesis of separability. Upper part*: $N = 100$; *lower part*: $N = 500$.

TABLE 1
P-values of the tests (*with different kernels*) *for the phonetic acoustic data*

Languages	N	M1 (ψ_1)	M1 (ψ_2)	M1 (ψ_3)	M2 (ψ_1)	M2 (ψ_2)	M2 (ψ_3)
French	60	0.003	0.005	0.002	0.002	<0.001	<0.001
Italian	50	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Portuguese	25	<0.001	0.002	<0.001	<0.001	<0.001	<0.001
American Spanish	46	0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Castilian Spanish	38	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

REFERENCES

[1] ASTON, J. A. D., PIGOLI, D. and TAVAKOLI, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *Ann. Statist.* **45** 1431–1461. MR3670184 https://doi.org/10.1214/16-AOS1495

[2] BAGCHI, P. and DETTE, H. (2020). Supplement to “A test for separability in covariance operators of random surfaces.” https://doi.org/10.1214/19-AOS1888SUPP.

[3] BAR-HEN, A., BEL, L. and CHEDDADI, R. (2008). Spatio-temporal functional regression on paleoecological data. In *Functional and Operatorial Statistics*. *Contrib. Statist*. 53–56. Physica-Verlag/Springer, Heidelberg. MR2490328

[4] CONSTANTINOU, P., KOKOSZKA, P. and REIMHERR, M. (2017). Testing separability of space-time functional processes. *Biometrika* **104** 425–437. MR3698263 https://doi.org/10.1093/biomet/asx013

[5] CONSTANTINOU, P., KOKOSZKA, P. and REIMHERR, M. (2018). Testing separability of functional time series. *J. Time Series Anal*. **39** 731–747. MR3849524 https://doi.org/10.1111/jtsa.12302

[6] DAUXOIS, J., POUSSE, A. and ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal*. **12** 136–154. MR0650934 https://doi.org/10.1016/0047-259X(82)90088-4

[7] DUNFORD, N. and SCHWARTZ, J. T. (1988). *Linear Operators*, *Vols*. *I*, *II*, *III*. *Wiley Classics Library*. Wiley, New York.

[8] FUENTES, M. (2006). Testing for separability of spatial-temporal covariance functions. *J. Statist. Plann. Inference* **136** 447–466. MR2211349 https://doi.org/10.1016/j.jspi.2004.07.004

[9] GENTON, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics* **18** 681–695. MR2408938 https://doi.org/10.1002/env.854

[10] GNEITING, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc*. **97** 590–600. MR1941475 https://doi.org/10.1198/016214502760047113

[11] GNEITING, T., GENTON, M. G. and GUTTORP, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. In *Statistical Methods for Spatio-Temporal Systems* (B. Finkenstadt, L. Held and V. Isham, eds.). *Monographs on Statistics and Applied Probability* **107**. CRC Press/CRC, Boca Raton, FL. MR2307967

[12] GOHBERG, I., GOLDBERG, S. and KAASHOEK, M. A. (1990). *Classes of Linear Operators*, *Vol*. *I*. *Operator Theory*: *Advances and Applications* **49**. Birkhäuser, Basel.

[13] HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis*, *with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3379106 https://doi.org/10.1002/9781118762547

[14] HUIZENGA, H. M., DE MUNCK, J. C., WALDORF, L. J. and GRASMAN, R. P. P. P. (2002). Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model. *IEEE Trans*. *Biomed*. *Eng*. **49** 533–539.

[15] LU, N. and ZIMMERMAN, D. L. (2005). The likelihood ratio test for a separable covariance matrix. *Statist. Probab. Lett*. **73** 449–457. MR2187860 https://doi.org/10.1016/j.spl.2005.04.020

[16] LYNCH, B. and CHEN, K. (2018). A test of weak separability for multi-way functional data, with application to brain connectivity studies. *Biometrika* **105** 815–831. MR3877867 https://doi.org/10.1093/biomet/asy048

[17] MITCHELL, M. W., GENTON, M. G. and GUMPERTZ, M. L. (2005). Testing for separability of space-time covariances. *Environmetrics* **16** 819–831. MR2216653 https://doi.org/10.1002/env.737

[18] MITCHELL, M. W., GENTON, M. G. and GUMPERTZ, M. L. (2006). A likelihood ratio test for separability of covariances. *J. Multivariate Anal*. **97** 1025–1043. MR2276147 https://doi.org/10.1016/j.jmva.2005.07.005

[19] RABINER, L. R. and SCHAFER, R. W. (1978). *Digital Processing of Speech Signals* 100. Prentice-Hall, Englewood Cliffs, NJ.

[20] ROUGIER, J. (2017). A representation theorem for stochastic processes with separable covariance functions, and its implications for emulation. Preprint. Available at arXiv:1702.05599.

[21] SKUP, M. (2010). Longitudinal fMRI analysis: A review of methods. *Stat*. *Interface* **3** 235–252. MR2659514 https://doi.org/10.4310/SII.2010.v3.n2.a10

[22] TAVAKOLI, S. (2016). covsep: Tests for determining if the covariance structure of 2-dimensional data is separable. R package version 1.0.0.

[23] VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[24] WEIDMANN, J. (1980). *Linear Operators in Hilbert Spaces*. *Graduate Texts in Mathematics* **68**. Springer, New York–Berlin. MR0566954

[25] WORSLEY, K. J., MARRETT, S., NEELIN, P., VANDAL, A. C., FRISTON, K. J. and EVANS, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Hum. Brain Mapp*. **4** 58–73.