TEST
https://doi.org/10.1007/s11749-025-00972-y

ORIGINAL PAPER

A kernel-based test for the first-order separability of spatio-temporal point processes

Mohammad Ghorbani[1] · **Nafiseh Vafaei**[1,2] · **Mari Myllymäki**[3]

Received: 16 November 2023 / Accepted: 23 April 2025

Abstract
We present an innovative statistical test designed to assess the first-order separability of a spatio-temporal point process. Our proposed test employs block permutations and a novel test statistic that incorporates a machine learning technique known as the Hilbert–Schmidt independence criterion. To enhance the practicality of the criterion, we apply the kernel trick. The block permutations are designed to maintain the second-order structure of the point pattern, disrupting it only at the block borders. This design enables the application of our test to a general spatio-temporal point process, which may exhibit small-scale clustering or regularity. We investigated the empirical level of the block permutation-based tests with the new and two previously proposed test statistics for clustered and regular point processes, represented in our study by log Gaussian Cox processes and determinantal point processes. By comparing our results with those obtained from a previously proposed permutation-based test, we confirmed the effectiveness of our method in terms of significance level, power, and notably computational cost. We applied the test to real-world datasets, namely the UK's 2001 foot-and-mouth disease epidemic and varicella data from Valencia.

Keywords Block permutation · feature space · Kernel mean embedding · Machine learning · Reproducing kernel Hilbert space · Separability of intensity function · Spatio-temporal point processes.

✉ Mohammad Ghorbani
mohammad.ghorbani@ltu.se

Nafiseh Vafaei
N.Vafaei@uma.ac.ir

Mari Myllymäki
mari.myllymaki@luke.fi

[1] Department of Engineering Sciences and Mathematics, Luleå University of Technology, Luleå, Sweden

[2] Department of Computer and Statistics Sciences, University of Mohaghegh Ardabili, Ardabil, Iran

[3] Natural Resources Institute Finland (Luke), Helsinki, Finland

Published online: 24 May 2025

1 Introduction

Consider a spatio-temporal point process (STPP) X with non-overlapping points as a random countable subset of $\mathbb{R}^2 \times \mathbb{R}$, where each point of X represents the location and time of occurrence of a particular event of interest. In the analysis of spatio-temporal point patterns, the crucial initial step is to model or estimate the intensity function. However, modeling the joint distribution of spatial locations and times in such a process poses challenges in practical applications (Gonzalez et al. 2019; Fuentes-Santos et al. 2018). To overcome these challenges, it is often a beneficial assumption that the point process has a separable spatio-temporal intensity function (Gabriel and Diggle 2009; Møller and Ghorbani 2012). To deduce if one can work under this simplifying assumption, the primary objective of our work is to test the null hypothesis of first-order separability, hereafter referred to as the separability hypothesis, for a given point pattern.

In the context of spatio-temporal marked point processes, some test statistics utilizing the conditional intensity function have been proposed to assess the separability of marks and the spatio-temporal process, as highlighted by Schoenberg (2004), Assunção and Maia (2007) and Diaz-Avaloset al. (2013). It is worth noting that, unlike these papers, our study focuses on the overall intensity function rather than the conditional intensity and tests its separability in space and time. Recently, several tests have been developed for assessing the separability of the intensity function. However, among the available tests in the literature, tests based on the pure permutation approach (Fuentes-Santos et al. 2018; Gonzalez et al. 2019; Ghorbani et al. 2021) work well for Poisson processes only; for non-Poisson processes, the permutation strategy destroys the interaction structure of the observed pattern, making the approach invalid. The same limitation applies to the computationally cheap χ^2-test (Ghorbani et al. 2021). Furthermore, the random reconstruction method proposed by Ghorbani et al. (2021) for non-Poisson processes is computationally very expensive.

To address these limitations, we develop a novel test of separability based on the *Hilbert–Schmidt independence criterion* (HSIC) and *block permutation*, and evaluate its performance in addition to the Poisson case, for regular and clustered point process models, in particular for spatio-temporal log Gaussian Cox Processes (LGCPs) and for spatio-temporal determinantal point processes (STDPPs) (for further insights on STDPPs, refer to the details provided in Vafaei et al. (2023)). The idea of the block permutation strategy, which is used to obtain simulations under the separability hypothesis here, is that it approximately maintains the interaction structure of the point pattern, disrupting it only at the boundaries of the blocks. The other component of the proposed test, namely the test statistic based on the HSIC, is adapted from machine learning literature, where it is used for independence testing (Gretton et al. 2005, 2007; Pfister et al. 2017). The HSIC is based on embedding the distributions derived from the first-order intensity function into the reproducing kernel Hilbert space (RKHS) (Muandet et al. 2017), with the hope that hidden, underlying structures would be better revealed in the reproduced space. While HSIC has primarily been studied in the context of independent and identically distributed (i.i.d.) observations, it has later been developed to test independence of two random processes (Chwialkowski and Gretton 2014), non-i.i.d data (Zhang et al. 2008), and multivariate time series data (Liu et al. 2023). It has also

been applied to quantify spatial dependence in linguistic variables for dialectal analysis (Nguyen and Eisenstein 2017). Similar to its performance with i.i.d. data, HSIC has demonstrated promising results in non-i.i.d. scenarios. However, to our knowledge, the use of such kernel tricks is not common in the context of point processes and the only reference is Rustamov and Klosowski (2020), where they used an approximated maximum mean discrepancy to detect differences in the first-order structure of the spatial point patterns of larynx and lung cancer occurrences in Chorley–Ribble area of Lancashire, England.

The rest of the paper is structured as follows. In Sect. 2, we define the null hypothesis of first-order separability and provide background material on RKHS embedding and HSIC. Section 3 introduces the d-variable HSIC (DHSIC) method for testing first-order separability, including the test statistic and the testing procedure based on the block permutation algorithm. We also describe how the calculated statistic can be used for data analysis and simulations. Section 4 is devoted to three separate simulation studies covering regular, clustered and Poisson processes. The first study investigates the performance of the proposed test for inhomogeneous STDPPs within the scope of the regular point process class. The second study is focused on LGCPs, which are categorized as cluster processes, and the third study on inhomogeneous Poisson processes. We compared the new proposed test to the permutation-based tests of Ghorbani et al. (2021). In Sect. 5, we apply the test to real-world datasets. Specifically, we analyzed the UK's 2001 foot-and-mouth disease data in Cumbria (Keeling et al. 2001; Diggle 2006, 2013; Ghorbani 2013; Møller and Ghorbani 2012) and the cases of varicella-zoster virus registered in Valencia, Spain in 2013 (Iftimi et al. 2015). Finally, Sect. 7 concludes the paper with a brief discussion. Additionally, two appendices are included to provide details related to the simulation study, namely simulation of the STDPPs and choice of parameters for the STDPPs of the simulation study.

2 Assumptions and background

In this section, we provide the necessary background for testing the first-order separability of a STPP (Sect. 2.1) and give a brief overview of the reproducing kernel Hilbert space (RKHS) embedding used in the Hilbert–Schmidt independence criterion (HSIC) (Sects. 2.2 and 2.3).

2.1 First-order separability

Consider a spatio-temporal point process X as a random countable subset of $\mathbb{R}^2 \times \mathbb{R}$ with intensity function $\rho(u,t)$, $(u,t) \in \mathbb{R}^2 \times \mathbb{R}$ and suppose $\mathbf{x} = \{(u_1,t_1),\ldots,(u_n,t_n)\}$ is a realization of X. Formally, the intensity function ρ is a nonnegative function from $\mathbb{R}^2 \times \mathbb{R} \to [0,\infty)$ that is locally integrable, i.e., $\int_B \rho(u,t)\mathrm{d}u\mathrm{d}t < \infty$ for all bounded Borel sets B of $\mathbb{R}^2 \times \mathbb{R}$ and for any point (u,t). In intuitive terms, $\rho(u,t)d(u,t)$ is the probability that X has a point in an infinitesimally small region around (u,t) of volume $d(u,t)$. We say that the intensity function is separable and denote it by $\rho_{\text{sep}}(u,t)$ if it can be written in the multiplicative

form

$$\rho_{\text{sep}}(u, t) = \rho_1(u)\rho_2(t), \quad (u, t) \in \mathbb{R}^2 \times \mathbb{R},$$

where ρ_1 and ρ_2 are nonnegative measurable functions. First-order separability is commonly considered as a convenient working assumption (Gabriel and Diggle 2009; Møller and Ghorbani 2012; Møller et al. 2021). This is because, within the framework of separability, inferences about the spatio-temporal model can usually be based on the characteristics of lower-dimensional spatial and temporal marginal processes, which are more manageable. However, it is important to acknowledge that the assumption of separability in the intensity function may not always hold true. A test of separability would then be helpful in suggesting suitable explanatory variables for the variations in the intensity. Namely, under separability, spatial and temporal covariates may suffice, whereas otherwise one should also consider spatio-temporal covariates. Therefore, the test for separability can assist in determining the appropriate form of either parametric (Baddeley et al. 2015) or nonparametric intensity functions (Guan 2008; Borrajo et al. 2020).

To investigate the validity of this separability assumption, we recall the concepts of RKHS embedding in Sect. 2.2, a key component of the Hilbert–Schmidt independence criterion in Sect. 2.3. We explore how these concepts can be adapted and employed to evaluate the separability hypothesis in the context of spatio-temporal point processes (see Sect. 3).

2.2 Reproducing kernel Hilbert space embedding

Given a set Ω, a positive definite kernel k on Ω is defined as a symmetric function $k : \Omega \times \Omega \to \mathbb{R}$ that satisfies positive semidefiniteness, expressed by the condition $\sum_{i,j=1} c_i c_j k(x_i, x_j) \geq 0$ for any arbitrary set of points $x_1, \ldots, x_n$ in Ω and real numbers $c_1, \ldots, c_n$. The bivariate kernel k is known as a Mercer kernel, and the resulting matrix $\{k(x_i, x_j)\}_{i,j=1}$ is commonly referred to as the Gram matrix, denoted by G. It is a well-known fact (Aronszajn 1950) that a positive definite kernel k on Ω uniquely defines a Hilbert space $\mathcal{H}_k$ comprising functions defined on Ω. The space $\mathcal{H}_k$ possesses these key properties: i) $k(\cdot, x) \in \mathcal{H}_k$ for any $x \in \Omega$; ii) the set of all linear combinations of the kernels in Ω, i.e., $\text{Span}\{k(\cdot, x); x \in \Omega\}$, is dense in $\mathcal{H}_k$; and iii) $\langle f, k(\cdot, x)\rangle = f(x)$ for any $x \in \Omega$ and $f \in \mathcal{H}_k$ (the reproducing property), where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathcal{H}_k$ (Ghojogh et al. 2023). The Hilbert space $\mathcal{H}_k$ is commonly referred to as an RKHS associated with the kernel k.

In kernel methods, the space Ω represents the domain where data resides, and a positive definite kernel k is specifically tailored for Ω. The corresponding RKHS $\mathcal{H}_k$ is served as a feature space, and a nonlinear mapping, often referred to as a feature mapping, is defined to map the data space Ω into the feature space $\mathcal{H}_k$. This mapping, denoted as $\psi : \Omega \to \mathcal{H}_k$, assigns each data point x to the function $k(\cdot, x)$. In this context, $k(\cdot, x) \in \mathcal{H}_k$ should be understood as a function of the first argument, with x held fixed. In short, the general idea of the kernel method is to transform real data $x \in \Omega$ into a function in RKHS $\mathcal{H}_k$ where a similarity measure is employed instead of a

traditional distance metric. This functional representation of data enables the extraction of various nonlinear features inherent within the data. When the entire dataset is mapped in this way, it is anticipated that latent structures in the dataset which may not be readily apparent in the original space Ω, will become more evident with the help of an appropriate feature map ψ in $\mathcal{H}_k$. The mapping $\psi : \Omega \to \mathcal{H}_k$, given by $\psi(x) = k(\cdot, x)$, is known as the canonical feature map of k, i.e., $k(x, x') = \langle \psi(x), \psi(x') \rangle$ (Berlinet and Thomas-Agnan 2004). This extension of the inner product between two feature maps to a kernel function is often referred to as the *kernel trick*, as it elegantly eliminates the need for explicitly defining the feature maps and provides one of the essential elements in kernel methods. More generally, given two linear combinations of the feature vectors, say $f = \sum_{i=1}^{n} \alpha_i \psi(x_i)$ and $g = \sum_{j=1}^{n} \beta_j \psi(x_j)$, the inner product between f and g is given by $\langle f, g \rangle = \boldsymbol{\alpha}^T G_x \boldsymbol{\beta}$, where α_is and β_is are nonzero real numbers, $\boldsymbol{\alpha}^T = (\alpha_1, \ldots, \alpha_n)$, $\boldsymbol{\beta}^T = (\beta_1, \ldots, \beta_n)$, and G_x is the Gram matrix of the feature vectors.

Consider a measurable space denoted by $(\mathcal{X}, \mathcal{B}_\mathcal{X})$. Let X be a random variable taking values in $\mathcal{X}$ with a distribution P. We assume that the expected value of $\sqrt{k(X, X)}$ is finite, i.e., $\mathbb{E}_P(\sqrt{k(X, X)}) < \infty$, where the expectation is taken with respect to P. The kernel mean embedding (KME) of the distribution P into the Hilbert space $\mathcal{H}_k$ is a mapping $\mu^P(\cdot) : \mathcal{X} \to \mathcal{H}_k$ defined by

$$\mu^P(\cdot) = \int_{\mathcal{X}} k(\cdot, x) \mathrm{d}P(x) = \mathbb{E}_P[k(\cdot, X)] \tag{1}$$

, and its existence is guaranteed by the assumption $\mathbb{E}_P(\|k(\cdot, X)\|) = \mathbb{E}_P(\sqrt{k(X, X)}) < \infty$ (see Lemma 3.1 in Muandet et al. (2017)). For more details on kernel mean embedding, see Schneider et al. (2016). Given a sample $\{x_i\}_{i=1}^n$ from X, the moment estimator of μ^P is given by $\hat{\mu}^P(\cdot) = \frac{1}{n} \sum_{i=1}^{n} k(\cdot, x_i)$.

An important property of this embedding is that if k is a *characteristic kernel* (i.e., the mean embedding μ^P captures all the information of the distribution P), then the embedding is injective (Fukumizu et al. 2007; Muandet et al. 2017). In other words, for two probability distributions P and Q over $\mathcal{X}$, $\mu^P = \mu^Q$ if and only if $P = Q$, which implies that P and Q are identical. This property is particularly useful in the context of two-sample hypothesis testing, as the problem of testing whether $P = Q$ can be reduced to testing whether $\mu^P = \mu^Q$, albeit in infinite-dimensional space $\mathcal{H}_k$. To further illustrate this for readers unfamiliar with the kernel trick, note that the characteristic function of the probability distribution P can be seen as a special case of the kernel mean embedding μ^P, where $k(u, X) = \exp(iuX)$, with i denoting the imaginary unit. With this choice of k and using the one-to-one correspondence between cumulative distribution functions and characteristic functions (Ahsanullah 2017), testing whether two distributions P and Q are equal becomes equivalent to testing whether $\mu^P = \mu^Q$.

2.3 Hilbert–Schmidt Independence Criterion

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X : \Omega \to \mathcal{X}$ and $Y : \Omega \to \mathcal{Y}$ be two random variables with distributions P and Q, respectively. In essence, the HSIC is a nonparametric test used to determine whether the random variables X and Y are independent, given observed realizations $\{x_i\}_{i=1}^n$ for X and $\{y_i\}_{i=1}^n$ for Y. To be explicit, we observe pairs (x_i, y_i), where each x_i is a realization of the random variable X and each y_i is a realization of the random variable Y. Therefore, the dataset consists of n such pairs $\{(x_i, y_i)\}_{i=1}^n$, where each pair represents an instance of the joint realization of (X, Y). Recall that two random variables X and Y are independent if and only if their joint distribution is identical to the product of their respective marginal distributions. It is worth noting that the same notation X is used for a STPP as well as for a random variable; however, it will always be clear from the context which X is being referred to. The concept behind HSIC involves embedding both the joint distribution and the corresponding product form of the marginal distributions into a suitable RKHS, and subsequently assessing whether these embedded elements are equivalent.

Further, let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ be two continuous, bounded, and positive semidefinite kernels, and denote by $\mathcal{H}_k$ and $\mathcal{H}_l$ their corresponding RKHSs. Furthermore, suppose $\mu : \mathcal{X} \times \mathcal{Y} \to \mathcal{H}_k \otimes \mathcal{H}_l$ is the mean embedding function for the joint distribution associated with the characteristic kernel $k \otimes l$, where $\otimes$ denotes tensor product. Following Pfister et al. (2017) and considering the above assumptions, the distance DHSIC is defined as follows:

Definition 1 Consider an RKHS $\mathcal{H}_k$ of functions $\mathcal{X} \to \mathbb{R}$ with kernel k, and an RKHS $\mathcal{H}_l$ of functions $\mathcal{Y} \to \mathbb{R}$ with kernel l. The statistical functional

$$\text{DHSIC}(P(X, Y)) = \|\mu^{P(X,Y)} - \mu^{P(X)\otimes P(Y)}\|^2_{\mathcal{H}_k \otimes \mathcal{H}_l} \tag{2}$$

is called the d-variable (here $d = 2$) Hilbert–Schmidt independence criterion (DHSIC), where $P(X, Y)$ is the joint distribution of (X, Y), and $P(X)$ and $P(Y)$ are marginal distributions of X and Y.

Essentially, the DHSIC measures the squared distance between the joint distribution and the product of the marginals after embedding in RKHSs, providing a measure of the dissimilarity between the joint distribution and the product of the marginals in the RKHS framework. The DHSIC criterion can be readily extended to any dimension d; for further details, refer to Pfister et al. (2017). If both kernels k and l are characteristic, then under the null hypothesis of independence, $\text{DHSIC}(P(X, Y)) = 0$.

Note that while positive semidefinite kernels can be integrated into the DHSIC framework for i.i.d. data, the literature suggests that characteristic kernels are preferable for use with non-i.i.d. data, e.g., point patterns considered in this paper, see, e.g., Liu and Ruan (2024,Pages2-3) and Chwialkowski and Gretton (2014, Lemma1). This recommendation arises from the fact that characteristic kernels can capture all modes of dependence between random variables X and Y. Although non-characteristic kernels can also be employed within the DHSIC framework, they may fail to detect certain forms of dependence. Notably, it is possible for $\text{DHSIC}(P(X, Y)) = 0$ even when X

and Y are not independent. In contrast, for characteristic kernels, $\text{DHSIC}(P(X, Y)) = 0$ if and only if X and Y are independent.

To obtain an estimator of DHSIC, it is helpful to express it in terms of the individual kernels k and l. Based on the individual kernels, and following Gretton et al. (2007), the DHSIC is given by

$$\begin{aligned}\text{DHSIC}(P(X, Y)) = \mathbb{E}_{XX'YY'}\left[k(X, X')l(Y, Y')\right] + \mathbb{E}_{XX'}\left[k(X, X')\right]\mathbb{E}_{YY'}\left[l(Y, Y')\right] \\ - 2\mathbb{E}_{XY}\Big[\mathbb{E}_{X'}\left[k(X, X')\right]\mathbb{E}_{Y'}\left[l(Y, Y')\right]\Big],\end{aligned} \tag{3}$$

where $\mathbb{E}_{XX'YY'}$ represents the expectation with respect to pairs (X, Y) and (X', Y') drawn from the joint distribution $P(X, Y)$. Similarly, $\mathbb{E}_{XX'}$ denotes the expectation taken over copies of X and X' drawn from $P(X)$, while $\mathbb{E}_{XY}$ signifies the expectation with respect to the pair (X, Y) drawn from the joint distribution $P(X, Y)$.

According to Pfister et al. (2017) and Gretton et al. (2007), an estimator of (3) is given by

$$\begin{aligned}\widehat{\text{DHSIC}}(P(X, Y)) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}k(x_i, x_j)l(y_i, y_j) \\ + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}k(x_i, x_j)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}l(y_i, y_j) \\ - \frac{2}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}k(x_i, x_j)\frac{1}{n}\sum_{j=1}^{n}l(y_i, y_j)\right],\end{aligned} \tag{4}$$

which can be easily evaluated after specifying kernels k and l, and observing realizations $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$.

3 Test of the first-order separability using Hilbert–Schmidt independence criterion

Let $\mathbf{x} = \{(u_1, t_1), \ldots, (u_n, t_n)\}$ represent a realization of a space-time point process X. We denote the spatial component process of X by X_{space} and the temporal component process by X_{time}. The spatial component process consists of all spatial locations with times t in the temporal domain $T \subset \mathbb{R}$, i.e., $X_{\text{space}} = \{u : (u, t) \in X,\ t \in T\}$, while the temporal component process contains all time points associated with locations u in the spatial window $W \subset \mathbb{R}^2$, i.e., $X_{\text{time}} = \{t : (u, t) \in X,\ u \in W\}$.

To align with the general framework (see Sects. 2.2 and 2.3), we treat X_{space} as equivalent to X, and X_{time} as equivalent to Y. To test the separability of the intensity function of a STPP X, we define the space-time probability density function (normalized intensity function) of the events for both the general (non-separable) and separable

cases, respectively, as

$$P(u,t) = \frac{\rho(u,t)}{\int_W \int_T \rho(u,t)\mathrm{d}u\mathrm{d}t}$$

and

$$\begin{aligned} P_{\text{sep}}(u,t) &= \frac{\rho_{\text{sep}}(u,t)}{\int_W \int_T \rho_{\text{sep}}(u,t)\mathrm{d}u\mathrm{d}t} \\ &= \frac{\rho_1(u)\rho_2(t)}{\int_W \rho_1(u)\mathrm{d}u \int_T \rho_2(t)\mathrm{d}t} = P_s(u)P_t(t), \end{aligned} \tag{5}$$

where $P_s(u)$ and $P_t(t)$ are the probability density functions in space and time, respectively (Baddeley et al. (2015), Page 197). The test of the null hypothesis

$$H_0 : \rho(u,t) = \rho_{\text{sep}}(u,t)$$

is now equivalent to the test of equality of the induced distributions, i.e., $P(u,t) = P_s(u)P_t(t)$. In other words, the test of the separability hypothesis for a STPP X reduces to testing the independence of the induced marginal probability distributions in space and time, see (5). If the intensity function is separable and the kernel $k(\cdot,(u,t))$ is multiplicative, i.e., $k(\cdot,(u,t)) = k_s(\cdot,u) \times k_t(\cdot,t)$, then the corresponding kernel mean embedding is also separable. Namely, using (1), we have

$$\begin{aligned} \mu^{P_{\text{sep}}} &= \int_{W\times T} k(\cdot,(u,t))\mathrm{d}P_{\text{sep}}(u,t) \\ &= \int_W k_s(\cdot,u)\mathrm{d}P_s(u) \times \int_T k_t(\cdot,t)\mathrm{d}P_t(t) = \mu^{P_s} \times \mu^{P_t}. \end{aligned}$$

Therefore, based on the DHSIC criterion in Definition 1, we introduce the statistical functional of separability for a STPP as

$$\text{DHSIC}(P(X_{\text{space}}, X_{\text{time}})) = \|\mu^P - \mu^{P_{\text{sep}}}\|^2_{\mathcal{H}_k\otimes\mathcal{H}_l}. \tag{6}$$

Having observed a realization $\mathbf{x} = \{(u_1,t_1),\ldots,(u_n,t_n)\}$ of a spatio-temporal point process (STPP) X, the estimation of DHSIC in (6) is carried out using the equation (4), which provides the estimate in terms of kernels. We use this estimate as our test statistic and call it DHSIC *test*.

There are various kernel functions commonly used in machine learning, such as linear, Gaussian, Laplacian, sigmoid, polynomial, cosine, and chi-squared kernels. In this study, we employed the characteristic Gaussian kernel

$$k(\mathbf{x},\mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|_2^2}{2\sigma^2}\right), \qquad \mathbf{x},\mathbf{y} \in \mathbb{R}^n \tag{7}$$

where $\|\cdot\|_2$ represents the ℓ_2 norm, and σ^2 is the bandwidth parameter. The median heuristic approach, as described in Pfister et al. (2017), is used to select the bandwidth σ. This approach requires that the median of $\{\|x_i - x_j\|^2 : i < j\}$ equals $2\sigma^2$. Pfister et al. (2017) claim that this heuristic approach performs quite well in many practical applications. For more details, see Pfister et al. (2017, Section 4.4). Thus, we employed the Gaussian kernel with the selected bandwidth σ for kernels k and l in (4) to estimate (6).

3.1 Test procedure

Since the distribution of the DHSIC test statistic is unknown, we employ simulation-based inference, a common approach in spatial and spatio-temporal statistics. One well-established method is the permutation test (Diggle 2013; Ghorbani et al. 2021). A remarkable advantage of the permutation test is that it does not require any distributional assumptions or any kind of model specification for the separability structure. The permutation test relies on the principle that, under the null hypothesis, the shuffled datasets should display a similar pattern to the observed data; otherwise, they would deviate from it. This notion is rooted in the concept of permutation invariance, where the joint distribution of the permuted data for each permutation is expected to be identical to the joint distribution of the observed data. However, in the case of non-Poisson patterns, simply randomly shuffling the data, i.e., random permutations of t_i holding u_i fixed, can disrupt the underlying interaction structure of the patterns. Consequently, the standard permutation test is inappropriate. To overcome this limitation, we rely on block permutations in our separability test described in Algorithm 3.1.

To generate approximate simulations under the null hypothesis, we employ the block permutation method. Here, the data is split to J blocks, which are randomly permuted. The idea is that this approach maintains the interaction structure within each block and breaks it only at the boundaries. Thus, the method is expected to approximately simulate data that aligns with the null hypothesis. Specifically, we partition the time interval T into J disjoint sub-intervals $T_1, \cdots, T_J$ (with approximately equal number of points), which splits the observed point pattern into J segments residing in $W \times T_1, \ldots, W \times T_J$. To achieve equal number of points within each sub-interval, we use the empirical quantiles of the observed times $t_i, i = 1, \ldots, n$. The time blocks are then permuted while maintaining structure within each block as in the original data, and holding the locations x_i fixed. The block permutation scheme is illustrated with an example in the Supplementary materials S1.

In our simulation study and real-world applications, we utilize the block permutation approach in conjunction with the test statistic (6) to test the null hypothesis of separability. Additionally, we employ it alongside the S- and χ^2-test statistics (see equations (8) and (9) in Sect. 4 and Ghorbani et al. (2021)). For each of these test statistics, large values are considered significant. Thus, in the Monte Carlo test, the observed and simulated statistics are ranked so that the largest statistic obtains rank 1. If the rank of the observed test statistic is among the most extreme statistics, there is evidence against the null hypothesis. This extremeness is conveyed in the Monte Carlo p value (see Algorithm 3.1).

The minimal p value based on $J!$ permutations (including the identity permutation for the data) is $1/(J!)$. Therefore, for $J = 3$, it is only possible to reach a p value $1/(J!) \approx 0.17$, and for $J = 4, 5, 6$, the smallest possible p values are approximately $0.042, 0.008, 0.001$, respectively. As $\alpha(J!)$ should be chosen as an integer (e.g., Barnard (1963)), for $J = 3$ and $J = 4$ one can use the significance levels of 0.17 and 0.042, respectively. In this paper, we mainly used $J = 5, 6, 7$, considering all $J!$ permutations and a significance level of 0.05. We note that for sufficiently large J, one can work with a smaller number of block permutations than $J!$, i.e., not going through all the possible permutations but taking a sample from them. To preserve the interaction structure of the given point pattern, the size of the blocks should not be too small in comparison with the range of interaction.

Algorithm 1 The separability test based on block permutations

1: Partition the time interval T into disjoint sub-intervals $T_1, \cdots, T_J$.
2: Permute T_js, $j = 1, \ldots, J$, with $J!$ possible ways.
3: Calculate the chosen test statistic for the observed pattern (the identity permutation).
4: For each permutation $i = 1$ to $i = J! - 1$
5: Assign the blockwise permuted times to the points whose locations are kept fixed.
6: Calculate the chosen test statistic for the permuted pattern.
7: Calculate the rank r_{obs} of the observed test statistic among the simulated statistics.
8: Obtain the Monte Carlo p value by $r_{\text{obs}}/J!$.

4 Simulation study

We investigated the performance of the DHSIC separability test based on the estimator of the statistical functional (6) and Algorithm 3.1 (block permutations) for STDPPs studied in Vafaei et al. (2023) (see Sect. 4.2), for LGCPs with the study setup adapted from Ghorbani et al. (2021) (see Sect. 4.3), as well as for spatio-temporal Poisson point processes (Sect. 4.4). The STDPPs serve as an example of regular spatio-temporal point processes, while the LGCPs are examples of clustered processes. We tested all hypotheses both with the block permutations with different numbers of blocks as well as with the pure permutations. This was done for completeness and having pure permutations as a reference strategy, even though the pure permutation-based methods were shown to fail for non-Poisson processes in Ghorbani et al. (2021). Regarding the Poisson processes, we tested both permutation strategies to understand how much of power may be lost with using the block permutation strategy instead of the pure permutations. We compared the new DHSIC statistic to the S- and χ^2-statistics from the literature, as well as different bandwidth selection rules as described in Sect. 4.1.

The significance level of each test was set to 0.05. A thousand (1000) simulations were generated from each considered STDPP, LGCP, and Poisson model. The block permutation-based tests used $J!$ permutations for $J < 7$ and 1999 block permutations for $J \geq 7$ (Sect. 4.2.3 only). The number of pure permutations was 1999 in all cases.

4.1 Test statistics

We compared our new DHSIC test statistic to the two test statistics suggested by Ghorbani et al. (2021), namely the S- and χ^2-statistics. The S-statistic is defined as

$$S(u,t) = \frac{\hat{\rho}(u,t)}{\hat{\rho}_1(u)\hat{\rho}_2(t)} = \frac{\hat{\rho}(u,t)}{\hat{\rho}_{\text{space}}(u)\hat{\rho}_{\text{time}}(t)/n}, \quad (u,t) \in W \times T, \tag{8}$$

where $\hat{\rho}(u,t) > 0$ is the spatio-temporal intensity estimate, and $\hat{\rho}_{\text{space}}(u) > 0$ and $\hat{\rho}_{\text{time}}(t) > 0$ are the intensity estimates for the corresponding spatial and temporal components, respectively. Following Ghorbani et al. (2021), the spatial bandwidths for estimating the intensities—whether spatio-temporal or spatial—in the test statistic (8) were determined using the `bw.diggle()` function from the R package spatstat (Baddeley et al. 2015). The aim was to minimize the mean-square error criterion (Diggle 1985). For the temporal bandwidth, we followed the method proposed by Sheather and Jones (1991), which involves pilot estimation of derivatives to minimize the mean integrated square error criterion. The `bw.SJ()` function from the R package base (R Core Team 2023) was utilized to calculate this bandwidth. To test the separability hypothesis, as done by Ghorbani et al. (2021) for Poisson processes, we applied the global extreme rank length (ERL) envelope (Myllymäki et al. 2017; Myllymäki and Mrkvička 2023) to the empirical S-statistic and the S-statistics computed from simulations (here permutations or block permutations). Further, to ensure a fair comparison between the S-statistic and the DHSIC test, we employed the median heuristic bandwidths proposed by Pfister et al. (2017) for the DHSIC statistic (see Sect. 3) also for the S-statistic. Likewise, we computed the DHSIC statistic with both sets of bandwidths. In the subsequent sections, we denote the test statistics as DHSIC_D and S_D when the spatial bandwidth is determined with `bw.diggle()` and the temporal bandwidth with `bw.SJ()`. Conversely, when the median heuristic approach is used for bandwidth selection, we denote them as DHSIC_H and S_H.

The χ^2-test statistic is defined as

$$\chi^2 = \sum_{i=1}^{J^\chi} \sum_{j=1}^{I^\chi} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}, \tag{9}$$

where the observation window $W \times T$ is divided into $I^\chi \times J^\chi$ disjoint subsets (or cells), and n_{ij} and $e_{ij} = \left(\sum_{i=1}^{I^\chi} n_{ij}\right)\left(\sum_{j=1}^{J^\chi} n_{ij}\right)\Big/ n$ are, respectively, the number of observed and expected points under the null hypothesis in the subsets $W_i \times T_j$, $i = 1, \ldots, I^\chi$, $j = 1, \ldots, J^\chi$. We chose J^χ and I^χ in such a way that the expected number of points in each cell is at least five. The χ^2-test follows the asymptotic χ^2 distribution with $(I^\chi - 1)(J^\chi - 1)$ degrees of freedom under H_0, being asymptotically exact under the Poisson assumption. We also applied the same test statistic (9) within Algorithm 3.1, incorporating the block permutation strategy.

4.2 Spatio-temporal determinantal point processes

We investigated the empirical rejection rates of the DHSIC, S-based global envelope, and χ^2 statistic (9)-based tests for STDPPs with the models explained in Sect. 4.2.1. We considered both separable and non-separable intensity functions. The separable model was used to evaluate the empirical significance levels of the tests, whereas the non-separable model was employed for power assessment. For each inhomogeneous STDPP model, we generated 1000 realizations in the unit cube, $W \times T = [0, 1]^3$, using the thinning procedure described in Appendix A. The DHSIC test with Gaussian kernel was then applied to each simulated point pattern, obtained by splitting the pattern into disjoint $J = 5, 6$ and 7 blocks and permuting the resulting J blocks, and hence computing the Monte Carlo test based on the DHSIC test statistic as described in Sect. 3.1. For these choices of J, we considered the mean number of points in each block adequately large.

We first conducted experiments using the models described in Sect. 4.2.1. The results are reported in Sect. 4.2.2. Additionally, we replicated these experiments with doubled intensity of the simulated point processes for model (i). The findings from these experiments are summarized in Sect. 4.2.3.

4.2.1 Models

We considered the separable intensity function

$$\rho_{\text{sep}}(u, t) = at(1 + \cos(x + y)) \tag{10}$$

and the non-separable intensity function

$$\rho(u, t) = a(1 + \cos(5(x + y + t)). \tag{11}$$

For STDPPs with these separable and non-separable intensities, we considered the following three spectral densities:

(i) the separable spectral density

$$\varphi_{sep}(\omega, \tau) = \frac{2\pi\rho\alpha_s^2\alpha_t\sigma_s^2\sigma_t^2}{(1 + 4\pi^2\alpha_t^2\tau^2)} \exp\left(-\pi^2\alpha_s^2\|\omega\|^2\right) \tag{12}$$

with the variance parameters $\sigma_s^2 = \sigma_t^2 = 1$;

(ii) the spectral density

$$\varphi_\epsilon(\omega, \tau) = \gamma(\alpha_s^2\alpha_t^2 + \alpha_t^2|\omega|^2 + \alpha_s^2\tau^2 + \epsilon|\omega|^2\tau^2)^{-\nu}, \tag{13}$$

with $\epsilon = 1$ and $\nu = 2$; and

(iii) the same spectral density (13) with $\epsilon = 0$ and $\nu = 2$.

The models (i) and (ii) are separable spectral densities, while the spectral density (iii) is non-separable. In model (i), α_t and α_s are temporal and spatial range parameters, respectively, while in models (ii) and (iii), α_t^{-1} and α_s^{-1} serve as temporal and spatial range parameters, respectively. See Vafaei et al. (2023) for more detailed explanation of the parameters in models (i)-(iii), and their roles in defining these STDPP models.

We fixed $a = 100$ in (10) and (11) such that the maximum of each intensity function equaled $\rho_{\max} = 2a = 200$. For the parameters of the spectral densities, we considered different parameter value combinations. The choice of these parameters is explained in detail in Appendix B. Shortly, the choices were done to fulfill the existence condition of a STDPP, as explained in details in Vafaei et al. (2023). Given the intensity and temporal range parameters, the maximal spatial range parameter was specified accordingly. The parameter values are displayed in the legends of Figs. 1 and 2, which present the results, and are also replicated in the columns of Tables S1–S9 of Supplementary material S3. Supplementary material S2 presents examples of simulated point patterns from the models (i)-(iii).

4.2.2 Results

Figure 1 shows the empirical significance levels of the tests for the separable intensity (10). The three panels correspond to spectral densities (i), (ii), and (iii). Detailed numerical results can be found in the upper parts of Tables S1–S9 in the Supplementary materials. The empirical rejection rates in the case of separable intensity should range from 0.037 to 0.064 with a probability of 0.95, given by the 2.5% and 97.5% quantiles of the binomial distribution with parameters 1000 and 0.05. We can observe the following:

- For model (i), empirical significance levels of all tests were close to the nominal level of 0.05, except DHSIC_D, which was slightly conservative.
- For models (ii) and (iii), all tests were conservative, with the DHSIC_D test demonstrating notably more severe conservativeness than the other tests. The S-based tests were less conservative than the other tests for model (iii) when block permutations were used.
- The bandwidth choice influenced the empirical significance level of DHSIC, with only DHSIC_H yielding levels close to the nominal level. On the other hand, there was no clear difference between S_H and S_D.
- All tests using pure permutation, regardless of the bandwidth selection method, exhibited severe conservatism for models (ii) and (iii), whereas they were either weakly conservative or close to the nominal level for model (i).
- The empirical significance levels either remained largely unchanged or moved closer to the nominal level when switching to the *block permutation* method. This improvement was particularly pronounced in spectral densities (ii) and (iii) and with the DHSIC_D test for spectral density (i).
- When comparing the tests across different numbers of blocks ($J = 5, 6, 7$), no prominent trends were observed in the empirical significance levels.
- We observed no trends in the empirical significance levels with respect to different parameter values.

Figure 2 shows the empirical rejection rates (power) when the model had the non-separable intensity (11) and the spectral densities (i)–(iii). Detailed numerical results can be found in the upper parts of Tables S1–S9 in the Supplementary materials. We can observe the following:

- The power of DHSIC_H was higher than that of DHSIC_D.
- In contrast, the power of S_H was lower than that of S_D.
- Even though the asymptotic χ^2-test is, in principle, not valid for STDPPs, it had higher power than its block permutation version in all considered cases.
- The DHSIC_H test outperformed all other tests for all J values in the case of spectral densities (ii) and (iii) and performed comparably to S_D for spectral density (i).
- The power of the DHSIC tests and χ^2 tests tended to increase slightly with increasing J, but there was no clear difference between the choices of J values for the S-based tests.

In summary, for STDPPs the DHSIC test with heuristic (H) bandwidth demonstrated strong overall performance, achieving empirical significance levels close to 5% along with the highest power among the tests considered.

4.2.3 Higher intensity

With the separable and non-separable intensities and the spectral density model (i), we conducted further experiments by increasing $\rho_{\max}$ from 200 to 400, resulting in a higher mean number of points. The model parameters were again chosen following the guidelines described in Appendix B. For this higher number of points, we also explored the test performance for a higher number of blocks $J = 8, 9$ and 10, while ensuring that each block contained adequately large number of points. The χ^2 tests were now performed with $3 \times 3 \times 3$ cells. With this number of cells, the number of points in each cell is almost the same as with $3 \times 2 \times 2$ cells and $\rho_{\max} = 200$, making the comparison meaningful. The parameters, as well as the results of this experiment, are summarized in Fig. 3. Detailed numerical results can be found in Table S10 in the Supplementary materials. The results are as follows:

- The empirical significance levels were satisfactory for all tests here, too, except for S_D, which was liberal, in particular for $J = 5, 6, 7$.
- DHSIC_H, S_H, and χ^2 tests demonstrated empirical significance levels very close to the nominal 5%, especially for smaller values of α_s, outperforming the other tests.
- The DHSIC_H test had higher power than all other tests across all J.
- The DHSIC_D was more conservative than the DHSIC_H test, whereas the reverse order was observed for S_D and S_H.
- Again the asymptotic χ^2-test exhibited higher power than its block permutation version.
- As intensity increased, the power of the test improved, and the empirical significance levels approached the nominal level 5%, almost for all tests, in particular for DHSIC_H.

Note that as the number of points increases, the small bandwidth selected by Diggle's method results in a kernel intensity estimate $\hat{\rho}(u, t)$ that captures local fluctuations that

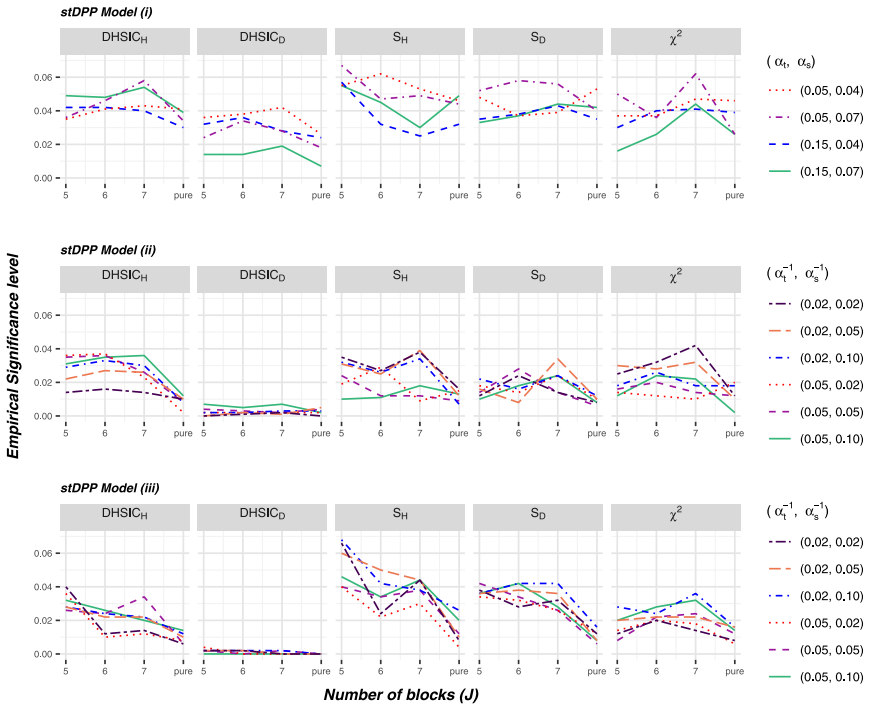

Fig. 1 Rejection rates of the different tests (columns) among 1000 replicates for the inhomogeneous STDPP with separable intensity function (10) and spectral densities (i), (ii) and (iii) (rows) with different scale parameters (α_t, α_s) or $(\alpha_t^{-1}, \alpha_s^{-1})$ as specified in the row-specific legends (see Sect. 4.2.1 for details). All tests were performed using block permutation with different numbers of blocks $J = 5, 6, 7$, and also with pure permutation with 1999 permutations, (labeled as **'pure'** on the x-axis), except for the χ^2-test where 'pure' corresponds to the asymptotic test. The indices H and D for DHSIC and S statistics refer to the median heuristic and the Diggle bandwidth selection methods, respectively. All χ^2-tests were computed based on $3 \times 2 \times 2$ cells and all S-statistics based on $25 \times 25 \times 20$ grids. The simulations of each model had on average about 75 points

should be rather interpreted as second-order behavior than true variation in the first-order intensity. Since the separability statistic S is highly sensitive to such deviations, these small-scale irregularities driven by high intensity and small bandwidth can mimic signals of non-separability, resulting in an inflated type I error rate (i.e., liberality). This likely explain why S_D values in the top row, fourth panel of Fig. 3 look different from the corresponding panel in Fig. 1 for $J = 5, 6, 7$. Recall that the block permutation tests were conducted using $J!$ permutations for $J < 7$, and 1999 block permutations for $J \geq 7$. This was done for computational efficiency, because we observed that the test results did not significantly change when larger values of permutations were used instead. It should be noted that running $J!$ permutations for larger values of J, the computations may take a considerable amount of time. Thus, as a practical recommendation, we suggest utilizing only a subset of all possible permutations for higher values of J.

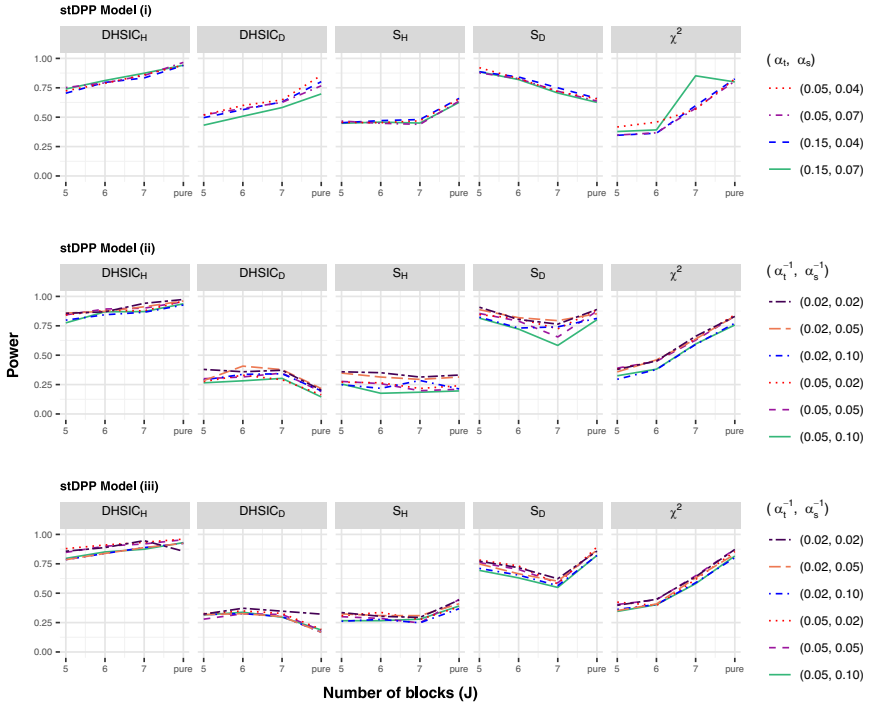

Fig. 2 Rejection rates of the separability tests as in Fig. 1 but with the non-separable intensity functions (ρ) (11). The simulations of each model had on average about 100 points

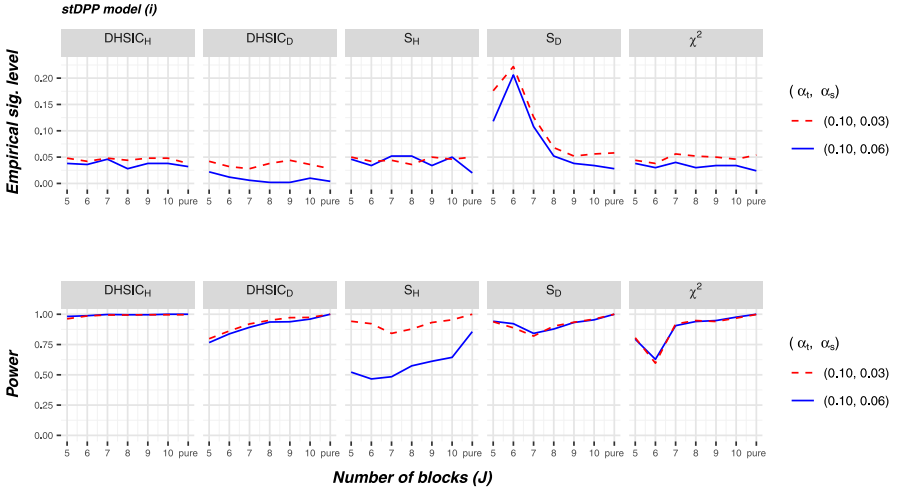

Fig. 3 Rejection rates of the separability tests as in Fig. 1 (top row) and Fig. 2 (bottom row) but for higher intensity of STDPP model (i) with different scale parameters (α_t, α_s) as specified in the row-specific legends and with all χ^2-tests computed based on $3 \times 3 \times 3$ cells. The simulations of each model had on average about 170 points

4.3 Log Gaussian Cox processes

To facilitate comparisons for the clustered LGCP, we adopted the simulation study setup of Ghorbani et al. (2021). For each parameter combination of the LGCP models described in Sect. 4.3.1 (see also Ghorbani et al. (2021)), we simulated 1000 realizations of the LGCP on the domain $[0, 1]^2 \times [0, 1]$. For each realization, we tested the null hypothesis of separability using the DHSIC, S and χ^2 statistics-based tests, employing both pure and block permutations, and considering Diggle and median heuristic bandwidth selection methods.

4.3.1 Models

Following Ghorbani et al. (2021), we considered a spatio-temporal stationary log Gaussian Cox process X driven by a nonnegative stochastic process

$$\Lambda(u, t) = \exp(m(u, t) + Z(u, t)), \quad (u, t) \in \mathbb{R}^2 \times \mathbb{R},$$

where $m(u, t)$ is a non-random/deterministic trend, and $Z(u, t)$ is a Gaussian random field with mean 0, variance $\sigma^2(u, t)$ and covariance function $C((u_1, t_1), (u_2, t_2))$. The intensity function of the LGCP is

$$\rho(u, t) = \exp(m(u, t) + \sigma^2(u, t)/2).$$

Under the assumption that variance $\sigma^2(u, t) = C((u, t), (u, t)) = \sigma^2$ is constant, $\rho(u, t)$ is separable if the mean function $m(u, t)$ takes an additive structure, i.e.,

$$m(u, t) = \log \rho_1(u) + \log \rho_2(t) - \sigma^2/2 = m_1(u) + m_2(t),$$

where $m_1(u)$ and $m_2(t)$ are two functions that describe the trend of the field in space and time, respectively. In this simulation study, we consider the following mean function

$$m(u, t) = \beta_0 + \beta_1(x - t) + \gamma' xt.$$

Here, $\gamma' = 0$ corresponds to the hypothesis of separability, while $\gamma' \neq 0$ implies that the intensity function is non-separable. In addition, a higher value of γ' indicates a greater degree of non-separability in the intensity function. We additionally set

$$Z(u, t) = \sigma_1 Z_s(u) + \sigma_2 Z_t(t) + \gamma'' Z_{st}(u, t), \qquad u \in \mathbb{R}^2 \times \mathbb{R},$$

for parameters $\sigma_1, \sigma_2 > 0$, and $\gamma'' \geq 0$, where Z_s, Z_t, and Z_{st} are independent Gaussian random fields with mean zero and covariance functions $C_1(u_1, u_2) = \exp(-\|u_1 - u_2\|^2/\phi_1)$, $C_2(t_1, t_2) = \exp(-|t_1 - t_2|)/\phi_2)$, and $C_3((u_1, t_1), (u_2, t_2)) = C_1(u_1, u_2)C_2(t_1, t_2)$, respectively, with correlation parameters $\phi_1, \phi_2 > 0$. The parameter γ'' governs the second-order separability in terms of the pair correlation function, as outlined in (Møller and Ghorbani 2012). Specifically, when $\gamma'' = 0$, the spatio-temporal covariance function takes an additive form, resulting in a separable pair

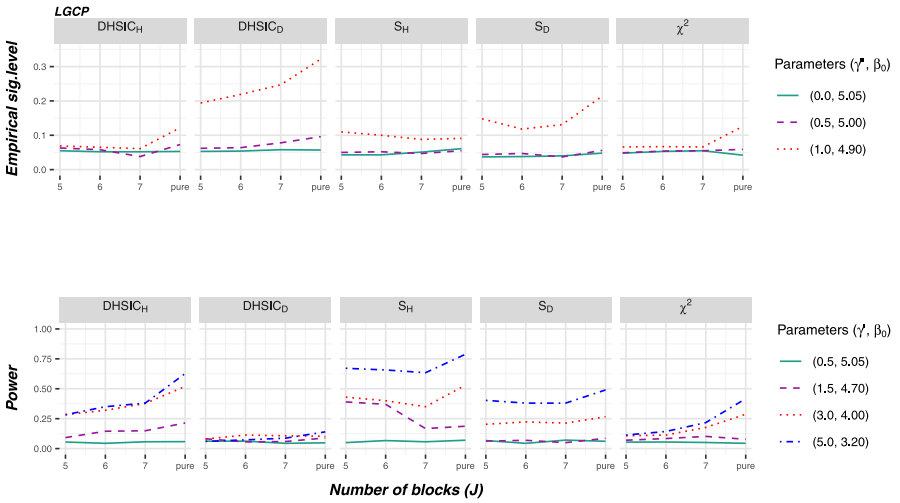

Fig. 4 Empirical significance levels ($\gamma' = 0$) for LGCP with different values of the parameters of (γ'', β_0) (top row) and power ($\gamma' > 0$) with $\gamma'' = 0$ and different values of β_0 across different tests (columns). Parameter values are specified in the row-specific legends (see Sect. 4.3 for details). All tests were performed using block permutation with different number of blocks $J = 5, 6, 7$, and also with pure permutation with 1999 permutations, except for the χ^2-test 'pure' corresponds to the asymptotic test. The indices H and D for DHSIC and S statistics refer to the median heuristic and the Diggle bandwidth selection methods, respectively. All χ^2-tests were computed based on $3 \times 2 \times 2$ cells and all S-statistics based on $25 \times 25 \times 20$ grids. The simulations of each model had on average about 200 points

correlation function. In contrast, when $\gamma'' > 0$, the covariance function becomes non-separable, leading the process to exhibit a non-separable second-order structure. Setting $\gamma' = \gamma'' = 0$ implies that the realization of the underlying random intensity $\Lambda(u, t)$ has multiplicative form. This, in turn, indicates that conditional on $\Lambda(u, t)$, for $\gamma' = \gamma'' = 0$, a realization of a separable inhomogeneous Poisson process is obtained.

We explored the performance of the tests for the case $\gamma' = 0$ (empirical significance levels) as well as for $\gamma' > 0$ (power). For testing the empirical significance levels with $\gamma' = 0$, we considered $\gamma'' = 0, 0.5, 1$. For the power comparison, γ' took values 0.5, 1.5, 3 and 5, while γ'' was fixed to 0. Regarding other parameters, we fixed $\beta_1 = 0.25$, $\sigma_1 = \sigma_2 = 0.5$, $\phi_1 = 0.06$ and $\phi_2 = 0.05$. Note that the parameter values were chosen such that the expected number of the points of each realization was around 200.

4.3.2 Results

In Fig. 4, the upper row shows the empirical significance levels for the tests, while the lower row presents the power results. Detailed numerical results can be found in Tables S11–S13 in the Supplementary materials. We observed the following about the empirical significance levels:

- For $\gamma'' = 0$, all tests achieved the nominal level of 5%. However, for $\gamma'' > 0$, DHSIC$_D$, S_H, and S_D exhibited liberal behavior, which became more pronounced as γ'' increased.

- The use of large bandwidths with the median heuristic approach (spatial bandwidths between 0.3 and 0.4, and temporal bandwidths between 0.19 and 0.22) effectively mitigated the impact of small-scale clustering patterns in STLGCP intensity estimation. This approach significantly improved test performance, with DHSIC_H (across all γ'') and the S-test (for small values of γ'') producing empirical significance levels closer to the nominal 5%.
- The block permutation method slightly reduced the liberality of some tests compared to pure permutation. Particularly, the DHSIC_D and χ^2-tests exhibited empirical significance levels closer to the nominal level, even for $\gamma'' = 1$, under block permutation compared to pure permutation.
- The DHSIC_H and χ^2-tests employing block permutations consistently outperformed the other tests, achieving the nominal level 5% across all γ'' values.

Regarding the power of the tests (Fig. 4, bottom row), we observed:

- The power of all tests, except for DHSIC_D, increased along the non-separability parameter γ'.
- For $\gamma' \neq 0.5$, the power of the tests DHSIC_H and χ^2 slightly increased by increasing the number of blocks.
- The power of the DHSIC and S-based tests exhibited sensitivity to bandwidth variations, with larger bandwidths enhancing performance. Specifically, DHSIC_H outperformed DHSIC_D, and S_H surpassed S_D.
- When inspecting the powers, it should be recalled that the DHSIC and S statistics-based tests that used pure permutations, and the asymptotic χ^2-test had too high empirical significance levels for large γ''. The same was observed also for DHSIC_D and S_D both with block and pure permutations.
- The DHSIC_H and S_H tests with block permutations were consistently the most powerful, and their empirical significance levels were also close to the nominal levels, or slightly liberal (S_H for $\gamma'' = 1.0$).

4.4 Poisson processes

We also investigated the empirical significance level and power of the tests based on the three test statistics (DHSIC, S, χ^2), employing two permutation methods (block and pure), and considering two bandwidth selection methods under the Poisson assumptions. We considered the inhomogeneous spatio-temporal Poisson process models outlined in Sect. 4.4.1. Section 4.4.2 presents the results.

4.4.1 Models

We considered an inhomogenous spatio-temporal Poisson process with the intensity function $\rho(u,t) = (\nu - \gamma)\xi(u)\psi(t) + \gamma\phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(u,t)$, where the parameter ν controls the expected number of points in the pattern, the parameter $\gamma \in [0, \nu/2]$ controls the degree of separability, and $\xi(u)$ and $\psi(t)$ are spatial and temporal baseline densities, respectively (see Ghorbani et al. (2021), for more details). For $\gamma = 0$, the model is separable. The second part of the model, with $\phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(u,t)$, introduces non-separability in the model when $\gamma \neq 0$. The density $\phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(u,t)$ is a three dimensional normal

distribution with mean $\mu = (0.3, 0.3, 0.2)$ and a diagonal covariance matrix $\Sigma = \text{diag}(0.05, 0.05, 0.05)$, creating a "burst" of points around μ. For the baseline densities, two cases were considered:

(M1) Overall constant density with $\xi(u) = 1$ and $\psi(t) = 1$,
(M2) $\psi(t) = 1$ and the spatial density $\xi(u)$ is a bivariate normal distribution $\phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(u)$ with a mean vector $\boldsymbol{\mu} = (0.5, 0.5)$ and a diagonal covariance matrix $\boldsymbol{\Sigma}$ with variances 0.2 in both directions.

We considered different values of $\gamma = 25, 35, 50, 75$, and we determined ν by fixing the expected number of points in the pattern to be around 600. The model was designed to replicate real-world scenarios, such as those found in environmental epidemiology, where incidents typically occur randomly within the population. While there may be an overall trend over time, occasional sudden spikes in incidents can occur, particularly around contaminated sources. The choice of 600 points was made to mimic the number of cases observed in the FMD data discussed in Sect. 5.1.

4.4.2 Results

The empirical significance levels of the tests for models (M1) and (M2) are presented in Fig. 5 and powers in Fig. 6. Detailed numerical results can be found in Tables S14–S16 in the Supplementary materials. We can observe the following:

- The empirical significance levels were satisfactory for all tests except for S_{D}, which exhibited clear liberality, especially for $J = 5$ and $J = 6$.
- The use of small bandwidths of the Diggle method increased the power of the DHSIC test. The same was true for the S-statistic-based test, whether employing pure permutation, or block permutation with $J = 7$. (The S_D test with block permutation and $J = 5, 6$ was too liberal to be considered.)
- At the empirical significance level 0.05, the DHSIC tests exhibited slightly higher power for all γ values in both models compared to their S-statistics counterparts, except for $\gamma = 75$ with $J = 5$, where DHSIC_{H} was negligibly lower. Both tests performed better than the χ^2 test.
- As expected, employing the block permutation strategy led to a decrease in power for all tests. The drop was, however, rather small with the DHSIC statistic with Diggle bandwidths (D).
- As anticipated, an increase in the non-separability parameter γ increased the power of the tests.

5 Case studies

As real-world case studies, we investigated the first-order separability of the spatial and temporal components in two datasets: the UK's 2001 epidemic foot-and-mouth disease (FMD) dataset in Cumbria (Sect. 5.1) and the varicella in Valencia (Sect. 5.2). To assess the separability, we employed the DHSIC separability test and Algorithm 3.1. Additionally, we compared the obtained results with the results of the global ERL

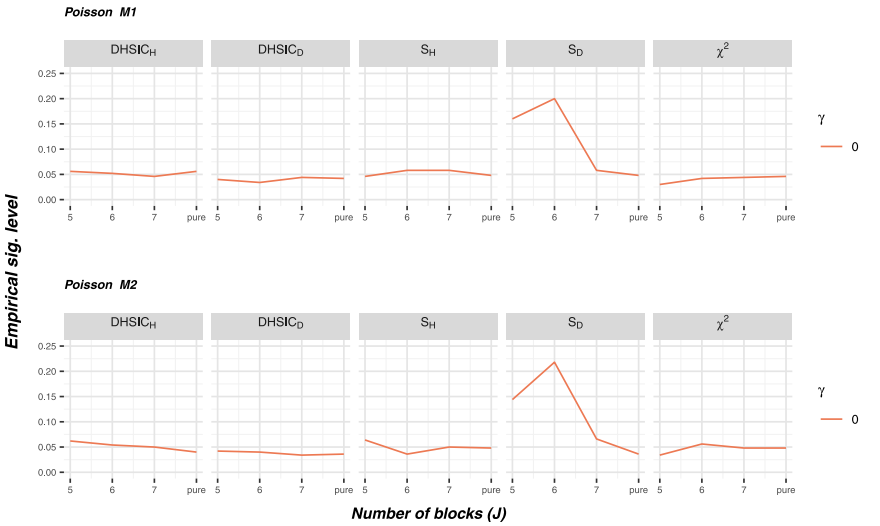

Fig. 5 Empirical significance levels for the Poisson process model across DHSIC, S-based global envelope, and χ^2 for the two inhomogeneous Poisson process models M1 and M2 (see Sect. 4.4.1 for details). All tests were performed using block permutation with different numbers of blocks $J = 5, 6, 7$, and also with 1999 pure permutation, except for the χ^2-test, where 'pure' corresponds to the asymptotic test. The indices H and D for DHSIC and S statistics refer to the median heuristic and the Diggle bandwidth selection methods, respectively. Lines represent parameter γ as specified in the legend. The χ^2-test was computed based on a $3 \times 2 \times 2$ grid, and all S-statistics were based on $25 \times 25 \times 20$ grids. The significance level was set to 0.05

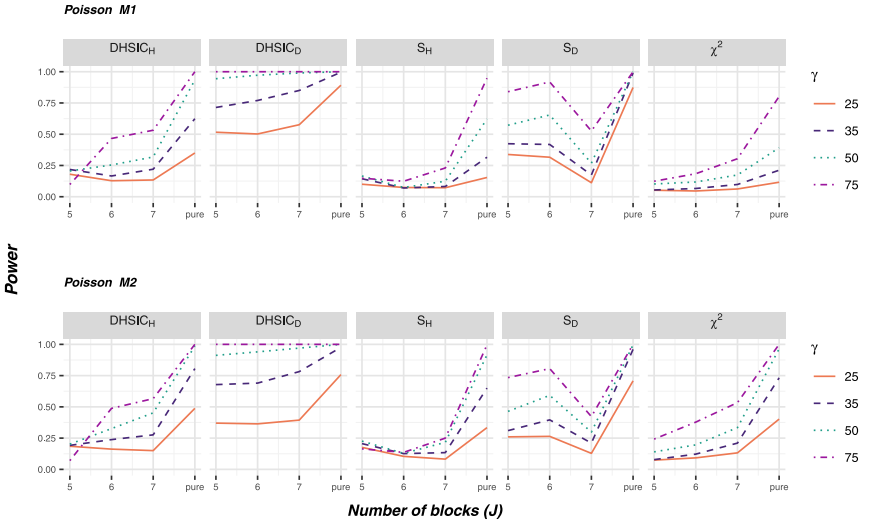

Fig. 6 Rejection rates of the separability tests as in Fig. 5 but with the non-separable intensity functions with parameter $\gamma > 0$ shown in legend. The simulations of each model had on average about 600 points

envelope test based on the S-statistic and pure permutations of times, as outlined in Ghorbani et al. (2021). To further enrich our comparative analysis, we also applied the S- and χ^2-statistics within Algorithm 3.1, incorporating the block permutation strategy. To differentiate between block and pure permutations, the notation bp is consistently added as a subscript to all tests employing block permutation.

5.1 UK 2001 epidemic foot-and-mouth disease data

Foot-and-mouth disease is a severe and highly contagious viral disease of livestock, with a significant economic impact. The disease affects cattle, swine, sheep, goats, and other ruminant cloven-hoofed animals. The data we analyze here come from the R package stpp (Gabriel et al. 2018) and were previously analyzed in Keeling et al. (2001), Diggle (2006, 2007), Gabriel et al. (2018), Møller and Ghorbani (2012), Ghorbani (2013), and Ghorbani et al. (2021). For more information on the data, see Diggle (2013).

The data we examined in this study pertain to Cumbria, a county located in northwest England, with an area of 5556.298km^2. The data were collected over a 200-day period starting February 1, 2001, resulting in a time interval of $T = [0, 200]$. Figure 7 displays the spatial point pattern of 648 infected farms in the irregular region W defined by Cumbria (upper left panel), marked based on the infection date, with smaller points representing older events. The lower panel of the figure illustrates the daily number of infected farms using bars. Further, the upper right and lower panels show, respectively, the estimated spatial intensity $\hat{\rho}_{\text{space}}$ using the kernel method with a bandwidth $b = 1.83$ km and the estimated temporal intensity $\hat{\rho}_{\text{time}}$ using the kernel method with a bandwidth of 3.59 days. The choice of bandwidths here was made by minimizing the MSE criterion (see Diggle (1985)) for the spatial intensity estimate and using the bw.SJ$(\cdot)$ function of the R package stats, which implements the methods of Sheather and Jones (1991) for bandwidth selection based on pilot estimation of derivatives, for the temporal intensity estimate. We used the Gaussian kernel (7) in both cases. Further, the spatio-temporal pattern of all FMD cases is shown in Fig. 8. The x- and y-axes show the spatial coordinates in kilometers, and the z-axis shows the temporal component in days.

The primary objective in analyzing the FMD dataset is to explore any potential relationship between the location of farms and the times t_i at which particular farms reported FMD cases (Keeling et al. 2001; Diggle 2013). Here, for comparison purposes, we used various tests to assess whether the spatial and temporal components for FMD data are first-order separable. Notably, Ghorbani et al. (2021) applied the global ERL envelope test (Myllymäki et al. 2017) with the S-function, which strongly rejected the null hypothesis of separability ($p$value $= 4.10^{-4}$), assuming an inhomogeneous Poisson process for the FMD data. Here, we used the DHSIC test together with block permutation which avoids making model assumptions such as Poissonity about the data.

Based on the simulation studies conducted in Sect. 4.3, the power of the DHSIC$_{\text{bp.H}}$ test increased with the number of blocks J. Thus, assuming that the FMD dataset exhibits small-scale clustering and given that the number of points is rather large

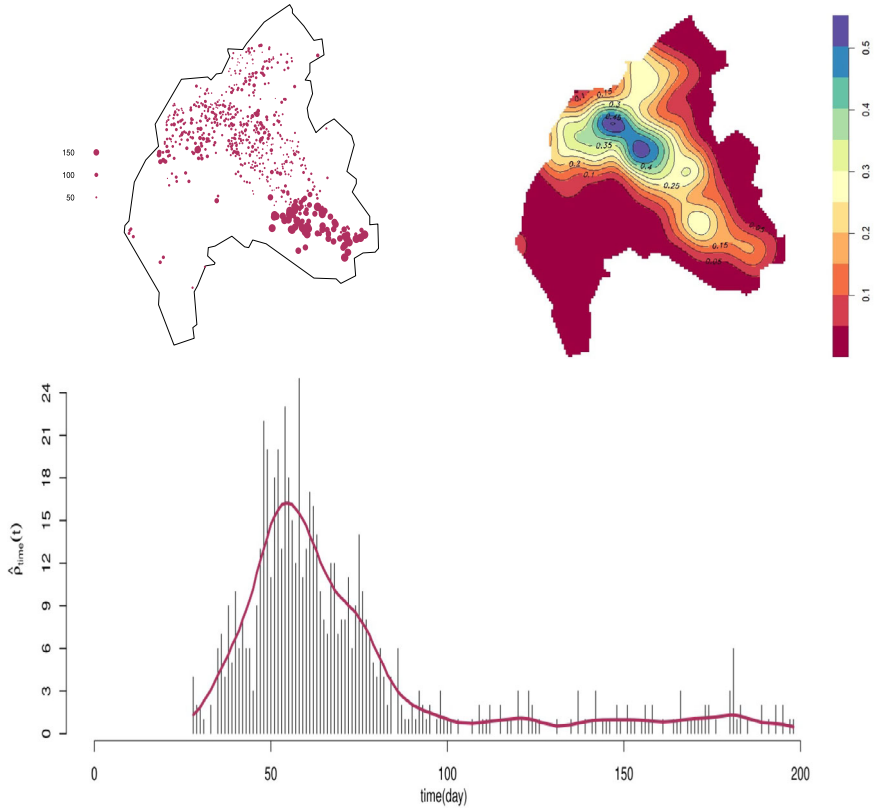

Fig. 7 Spatial point pattern of infected farms (top left panel) (the smaller points correspond to older events), $\hat{\rho}_{\text{space}}(r)$ (top right panel), and $\hat{\rho}_{\text{time}}(t)$ together with the daily number of infected farms (bottom panel) for the FDM data

(> 600), at least $J = 5$ can be recommended. However, since smaller values of J are preferable to avoid disturbing the interactions of the process, we did not consider $J > 8$. For comparison purposes, and according to the simulation studies in Sect. 4.3, we also performed the χ^2_{bp} test, which was performed based on $3 \times 3 \times 3$ cells, and the $S_{\text{bp.H}}$ test, which was evaluated on a $50 \times 50 \times 10$ grid. The $\text{DHSIC}_{\text{bp.H}}$ and $S_{\text{bp.H}}$ tests strongly rejected the separability hypothesis regardless of the number of blocks (J) (Table 1). The p values of the S-based global envelope test were similar to those of the $\text{DHSIC}_{\text{bp.H}}$ test, while the χ^2_{bp}-test failed to reject the separability hypothesis for $J \leq 6$. Thus, the $\text{DHSIC}_{\text{bp.H}}$ and $S_{\text{bp.H}}$ tests rejected the separability hypothesis similarly as Ghorbani et al. (2021) (permutation test with S and global ERL envelope test: p value $= 4 \cdot 10^{-4}$), but without restricting to the Poisson process case.

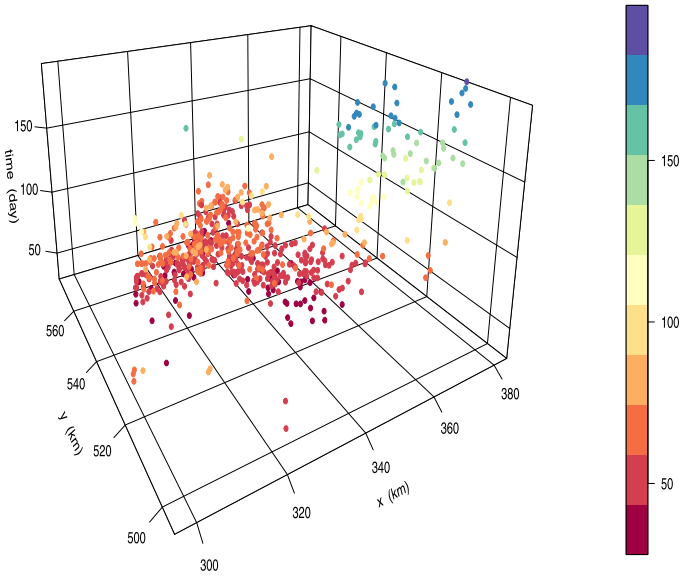

Fig. 8 Spatio-temporal point pattern of infected farms for the FDM cases in Cumberia, England. The color scale corresponds to infection time

Table 1 The p values of the $\text{DHSIC}_{\text{bp.H}}$, and $S_{\text{bp.H}}$, and χ^2_{bp} tests for the FMD data. The median heuristic spatial and temporal bandwidths, used in the $\text{DHSIC}_{\text{bp.H}}$ and $S_{\text{bp.H}}$ tests, were 18.44 km and 13.43 days, respectively

Tests	$J = 5$	$J = 6$	$J = 7$	$J = 8$
$\text{DHSIC}_{\text{bp.H}}$	0.0081	0.0072	0.0005	0.0005
$S_{\text{bp.H}}$	0.0082	0.0132	0.0071	0.0022
χ^2_{bp}	0.0751	0.1251	0.0175	0.0160

5.2 Varicella in Valencia

This section analyzes the varicella-zoster virus (also known as chickenpox) cases registered in Valencia, Spain, in 2013. Valencia is the third most populated municipality in Spain, with 791,413 inhabitants in the administrative center (districts) and an area of approximately 134 km^2 (Wikipedia of Valencia). The study area is represented by districts 1 to 16. The remaining districts are very sparsely populated and located far from the urban center. In 2013, 921 cases of varicella were recorded in the study area over a 52-week period (Department of Health, 2013). The dataset analyzed here was previously studied by Iftimi et al. (2015) to identify spatio-temporal clusters of chickenpox in Valencia and by Iftimi et al. (2017b) to model spatial locations of varicella using hybrids of Gibbs point processes (Baddeley et al. 2013). Additionally, Iftimi et al. (2017a) discovered spatio-temporal interactions in varicella by developing a spatial multi-scale area-interaction model to the spatio-temporal framework.

The spatial coordinates of chickenpox cases were initially given in latitude and longitude. To facilitate the analysis, these coordinates were converted from the lati-

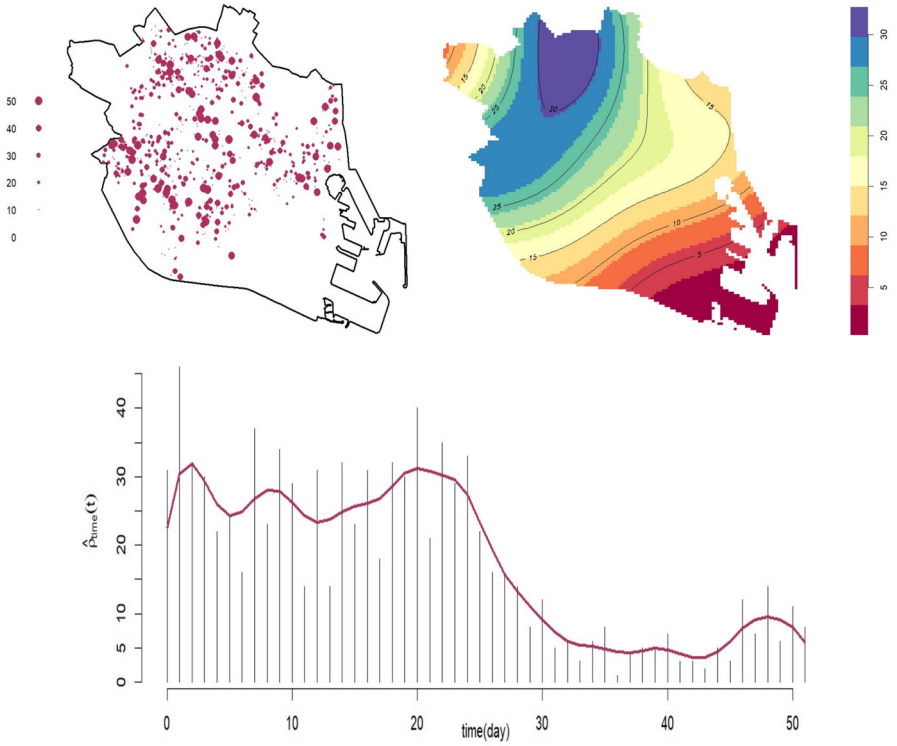

Fig. 9 Spatial point pattern of infected cases (top left panel) (the smaller points correspond to older events), $\hat{\rho}_{\text{space}}(r)$ (top right panel), and $\hat{\rho}_{\text{time}}(t)$ together with the daily number of infected cases (bottom panel) for the varicella data in Valencia

tude/longitude scale to the Universal Transverse Mercator (UTM) scale and expressed in meters. Figure 9 (left) shows the Valencia boundaries along with a spatial projection of the spatio-temporal point pattern given in Fig. 10. Smaller points represent older events. For the analysis purpose, the spatial coordinates were rescaled to kilometers, i.e., the study region was then enclosed within the box $[0, 9\text{km}] \times [0, 9\text{km}]$.

The time component of the process takes integer values from 0 to 51 because the data were collected over 52 weeks. Therefore, we set the spatio-temporal study area to $([0, 9] \times [0, 9]) \times [0, 52](\text{km}^2 \times \text{week})$. Figure 10 shows the spatio-temporal pattern of all varicella cases. The x- and y-axes show the spatial coordinates in kilometers, and the z-axis shows the temporal component in weeks.

Similar to the analysis performed on the FMD dataset, our primary focus in the analysis of varicella data revolves around uncovering significant associations or dependencies between the geographical locations of infected cases and the reported timing, which enriches our understanding of the underlying dynamics in the dataset. To explore this relationship thoroughly, we employed the DHSIC separability test and Algorithm 3.1 (block permutations). Additionally, we compared the obtained results with the global ERL envelope test based on the S-statistic, either pure (S-test) or block permutations ($S_{\text{bp.H}}$-test), as suggested by the simulation studies. Namely, based on our

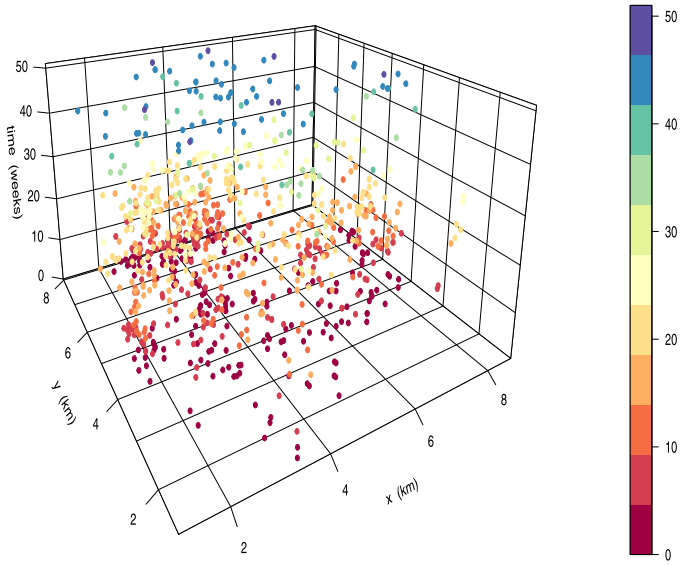

Fig. 10 Spatio-temporal pattern of varicella cases in Valencia, Spain. The color scale corresponds to infection time

Table 2 The p values of the DHSIC$_{\text{bp.H}}$, and $S_{\text{bp.H}}$, and χ^2_{bp} tests for the varicella data in Valencia. The median heuristic spatial and temporal bandwidths, used in the DHSIC$_{\text{bp.H}}$ and $S_{\text{bp.H}}$ tests, were 2.12 km and 8.5 weeks, respectively

blocks	5	6	7	8
DHSIC$_{bp.H}$	0.383	0.282	0.477	0.266
$S_{bp.H}$	0.651	0.430	0.972	0.453
χ^2_{bp}	0.375	0.443	0.416	0.321

visual assessment, the varicella data exhibits a regular pattern; thus, we considered the STDPPs as our reference case from the simulation study. To further enrich our comparative analysis, we also applied the S_{bp} and χ^2_{bp} tests, with a $3 \times 3 \times 3$ cell configuration. For the S-statistic, we estimated the spatial intensity, $\hat{\rho}_{\text{space}}$, with a bandwidth of $b_s = 2.12$ km and the temporal intensity, $\hat{\rho}_{\text{time}}$, with a bandwidth of $b_t = 8.5$ weeks and evaluated the S-statistic on a $50 \times 50 \times 10$ grid. These bandwidths were selected using the median heuristic approach explained in Sect. 3. Presumably, the choice $J = 6$ would also be a reasonable one here. However, for comparison purposes, we again performed all block permutation-based tests with $J = 5, 6, 7$, and 8. The resulting p values from almost all tests indicate substantial evidence supporting the null hypothesis of first-order separability (see Table 2).

We further tested the separability hypothesis for the varicella data using the global ERL envelope test based on the S-statistic (8) and *pure permutations*. The resulting $p$value $= 0.007$ supports rejecting the null hypothesis of separability. This underscores the importance of an appropriate simulation method in hypothesis testing.

6 Software

The R codes used in this work are available upon request from the corresponding author (mohammad.ghorbani@ltu.se).

7 Discussion and conclusion

We introduced a novel separability test for spatio-temporal point process models and evaluated its performance across various scenarios, including STDPPs, LGCPs, and inhomogeneous Poisson processes. For regular STDPPs, the new DHSIC test with block permutations and large bandwidths determined using the median heuristic approach ('H', spatial bandwidths between 0.3 and 0.39, and temporal bandwidths between 0.13 and 0.23 in our examples) performed the best in our simulation experiments. Its empirical significance levels were fine or a bit conservative. It had higher power than the tests based on previously suggested S- and χ^2-statistics. On the other hand, the DHSIC using small Diggle's bandwidths ('D', spatial bandwidths between 0.03 and 0.06, and temporal bandwidths between 0.05 and 0.07) previously suggested for the S-statistic was extremely conservative and had low power. For the LGCP models, the empirical significance levels of the DHSIC_H-based test were close to the nominal level, while the DHSIC_D as well as S- and χ^2-statistic-based tests were unacceptably liberal for some LGCPs (with large γ'' values). The tests based on the S-statistic and bandwidths from the median heuristic approach also performed well. The DHSIC_H-based test had comparable power with these S_H tests. Thus, the test with DHSIC_H and block permutations can be recommended for regular STDPPs and clustered LGCPs.

We also compared the DHSIC statistic to the S and χ^2 tests proposed by Ghorbani et al. (2021) in the Poisson process case. At the empirical significance level 0.05, the DHSIC tests exhibited slightly higher power for all γ values compared to their S-statistics counterparts. The DHSIC statistic using small bandwidths with the Diggle method demonstrated satisfactory empirical level with high power. It is worthwhile noting that the new proposed test is computationally more efficient. For example, for FMD data, the S-test took 700 s, whereas the DHSIC test only took 30 s, making the DHSIC test approximately 23 times faster.

We also included the χ^2-statistic (9) with the block permutation approach into our simulation studies for STDPPs and LGCPs. However, slightly surprisingly, the power of this test was lower than the power of the asymptotic χ^2-test not only for Poisson processes but also for STDPPs where both tests had appropriate significance levels. For LGCPs, the asymptotic test was liberal.

It is worth noting that there is no general rule to choose number of blocks J, and it somewhat depends on the range of dependence in the data. If there are long range dependencies, then one should choose large blocks, not to break them. However, these two points can be served as a guide: First, J should not be too large, so as not to completely destroy the structure of the point pattern; second, it should not be too small because the permutations may then not represent the null hypothesis. In our experiments for STDPP, when the number of points in a pattern was about 100, the choice $J = 5, 6$ or 7 performed well. We also originally experimented with the

choices $J = 4$ and $J = 3$, which allow to work at the significance levels 0.042 and 0.17. The empirical significance levels were close to the nominal levels with these values of J as well, but with $J = 3$ the significance level is obviously too large for a typical situation and with $J = 4$ the power of the test tended to be clearly lower than with $J = 5$ or 6 (results omitted here).

We used the aforementioned tests to investigate the first-order separability of FMD data and varicella data in Valencia. The results of these tests provided convincing evidence for the rejection of first-order separability for FMD data. Conversely, there was no evidence against the first-order separability hypotheses for the varicella data in Valencia.

It is worth to note that in this study, we calculated the DHSIC using the kernel embedding of the joint and marginal distributions. However, an alternative approach is to employ the correlation distance approximation, which offers an efficient and scalable method by computing the correlation distance between the kernel matrices associated with the joint and marginal distributions (Edelmann and Goeman 2022; Shen and Vogelstein 2021). The correlation distance serves as a similarity measure that quantifies the linear relationship between two sets of variables.

Supplementary materials

The supplementary materials include the following key items:

1. Illustration of the block permutation with an example (S1): This section explains the block permutation approach using an example. It demonstrates how data are partitioned and permuted while preserving the second-order structure of the spatio-temporal point process.
2. Realizations of STDPP models (S2): This section discusses separable and non-separable intensity functions and three spectral density models. It includes illustrative examples of realizations from these models to clarify the concept of separability and highlight differences between separable and non-separable processes. These visualizations aid in understanding the characteristics of the STDPP models used in the simulation study.
3. Tables of test results (S3): This section presents detailed numerical results on the performance of the proposed separability tests, which are summarized in Figs. 1-6. The tables include empirical significance levels and test power under various scenarios discussed in the main document.

A Simulation of inhomogeneous spatio-temporal determinantal point processes

Suppose that X is a SOIRS and isotropic STDPP with kernel function

$$C((u_1, t_1), (u_2, t_2)) = \sqrt{\rho(u_1, t_1)} R_0(u/\alpha_s, t/\alpha_t) \sqrt{\rho(u_2, t_2)}, \tag{14}$$

where ρ plays the role of the intensity function of the process, $\alpha_s > 0$ and $\alpha_t > 0$ are the spatial and temporal correlation parameters, respectively, and $R_0(\cdot)$ is the correlation function correspondent to C. The algorithms of Lavancier et al. (2015) can be used to simulate stationary DPPs. Then, the inhomogeneous STDPP with (14) can be simulated by applying independent thinning to an appropriate stationary STDPP, or another SOIRS STDPP. In general, suppose that Y is a STDPP with kernel C_y and p is a probability function that takes values in $[0, 1]$. If we denote Y_p as a point process obtained by retaining all $(u, t) \in Y$ independently and with probability $p(u, t)$, then according to Proposition A.2 in Lavancier et al. (2015), Y_p is a STDPP with kernel function

$$C_p((u_1, t_1), (u_2, t_2)) = \sqrt{p(u_1, t_1)} C_y((u_1, t_1), (u_2, t_2)) \sqrt{p(u_2, t_2)}. \tag{15}$$

Now, assume that $X_{\max}$ is a stationary version of X with kernel $C_{\max}(u, t) = \rho_{\max} R_0(u/\alpha_s, t/\alpha_t)$ such that $\rho_{\max}$ is the maximum of $\rho(u, t)$ on the given observation window $W \times T$. Let X_{thin} be the point process obtained by independent thinning of $X_{\max}$ with retention probability $p(u, t) = \rho(u, t)/\rho_{\max}$. Then applying (15), X_{thin} is a STDPP with the kernel function

$$\begin{aligned} C_{\text{thin}}((u_1, t_1), (u_2, t_2)) &= \sqrt{p(u_1, t_1)} C_{\max}(u, t) \sqrt{p(u_2, t_2)} \\ &= \sqrt{\rho(u_1, t_1)} R_0(u/\alpha_s, t/\alpha_t) \sqrt{\rho(u_2, t_2)}. \end{aligned}$$

Thus, the thinned process has the kernel and intensity of the original process X. Therefore, the process X can be simulated by first generating a realization of $X_{\max}$ and then performing independent thinning with retention probability $p(u, t) = \rho(u, t)/\rho_{\max}$.

B Simulation study: Explanation for the choice of parameters of the determinantal point processes

The range parameter values of the spectral densities were chosen for the models of the simulation study of Sect. 4.2 as follows: According to the existence of a STDPP for different spectral densities, the maximum intensity of the models is affected by the degree of interaction between the points. Distance parameters of each model were chosen to satisfy the corresponding existence condition of the a STDPP, i.e., $\varphi(\omega, \tau) \leq 1$ (for more details, see Section 3 in Vafaei et al. (2023)). Further, for given values of the temporal scale parameter and intensity, we chose maximal spatial scale parameters for the three spectral density models (i)-(iii):

(i) For the spectral density model (i) with the variance parameters $\sigma_s^2 = \sigma_t^2 = 1$ and considering that $\rho_{\max} = (2\pi\alpha_s^2\alpha_t\sigma_s^2\sigma_t^2)^{-1}$: the spatial range parameter, α_s, has the maximum value $(2\pi\rho\alpha_t)^{-1/2}$ for a given value of the intensity ρ and the temporal delay parameter α_t. Thus, for $\rho = 200$ and with fixed parameters $\alpha_t = 0.15$ and 0.05, the maximum of α_s is, respectively, 0.07 and 0.12 (choices listed in the legend of Fig. 1 (Top row)). As the parameters α_s and α_t increase, the value of

pair correlation function decreases which means that repulsion among the points increases (see Figure 1 in Vafaei et al. (2023)).

(ii) In the model (ii), with $\epsilon = 1$ and $\nu = 2$: the minimum value of the spatial scale parameter α_s for a given ρ and α_t is equal to $\{4\rho/(\pi^2\alpha_t)\}^{1/2}$. Therefore, for $\rho = 200$ and with fixed parameters $\alpha_t^{-1} = 0.05$ and 0.02.
Therefore, we obtained corresponding values, 0.1, 0.05, and 0.02 for the α_s^{-1} (See legend of Fig. 1 (Middle row)).

(iii) In the model (iii), with $\epsilon = 0$ and $\nu = 2$: the minimum value of spatial scale parameter α_s for the given ρ and α_t is equal to $\{(2\rho)/(\pi^2\alpha_t)\}^{1/2}$. Therefore, for $\rho = 200$ and $\alpha_t^{-1} = 0.05$ and 0.02 we obtained corresponding values for the α_s^{-1} (See legend of Fig. 1 (Bottom row)). For models (ii) and (iii), the value of pair correlation function decreases with increasing the α_s^{-1} and α_t^{-1} (see Figure 1 in Vafaei et al. (2023)), i.e., dispersion between the points increases, and therefore, we select the possible maximum values of the α_s^{-1}.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11749-025-00972-y.

Acknowledgements The authors express their gratitude to the anonymous referees and the associate editor for their constructive feedback and suggestions, which significantly improved the paper. Additionally, the authors are thankful to Adina Iftimi for sharing the Varicella data.

Funding Open access funding provided by Lulea University of Technology. MG was financially supported by the Kempe Foundations (grant number JCSMK22-0134) and the Olle Engkvists Foundation (grant number 224-0071), NV by the Kempe Foundations (grant number JCSMK22-0134), and MM by the Research Council of Finland (grant numbers 295100 and 327211).

References

Ahsanullah M (2017) *Characterizations of univariate continuous distributions*. Atlantis Press and the author(s)

Aronszajn N (1950) Theory of reproducing kernels. Transactions of the American Mathematical Society 68:337–404

Assunção R, Maia A (2007) A note on testing separability in spatial-temporal marked point processes. Biometrics 63:290–294

Baddeley A, Rubak E, Turner R (2015) *Spatial point patterns: Methodology and applications with R*. Chapman and Hall/CRC Press

Baddeley A, Turner R, Mateu J, Bevan A (2013) Hybrids of Gibbs point process models and their implementation. Journal of Statistical Software 55:1–43

Barnard GA (1963) Discussion of Professor Bartlettâ€™s paper. Journal of Royal Statistical Society Series B 25:294

Berlinet A, Thomas-Agnan C (2004) Reproducing kernel Hilbert spaces in probability and statistics. Springer, Boston

Borrajo M, González-Manteiga W, Martínez-Miranda M (2020) Bootstrapping kernel intensity estimation for inhomogeneous point processes with spatial covariates. Computational Statistics and Data Analysis 144:106875

Chwialkowski K, & Gretton A (2014) A kernel independence test for random processes. E.P. Xing and T. Jebara (Eds.), *Proceedings of the 31st international conference on machine learning* (Vol. 32, pp. 1422–1430)

Diaz-Avalos C, Juan P, Mateu J (2013) Similarity measures of conditional intensity functions to test separability in multidimensional point processes. Stochastic Environmental Research and Risk Assessment 27:1193–1205

Diggle PJ (1985) A kernel method for smoothing point process data. Applied Statistics 34:138–147

Diggle PJ (2006) Spatio-temporal point processes, partial likelihood, foot and mouth disease. Statistical Methods in Medical Research 15:325–336

Diggle PJ (2007) Spatio-temporal point processes: methods and applications. B. Finkenstädt, L. Held, and V. Isham (Eds.), *Statistical Methods for Spatio-Temporal Systems* (p.1-45). Chapman and Hall/CRC, Boca Raton

Diggle PJ (2013) Statistical analysis of spatial and spatio-temporal point patterns. Chapman and Hall/CRC, Boca Raton

Edelmann D, Goeman J (2022) A regression perspective on generalized distance covariance and the Hilbert-Schmidt independence criterion. Statistical Science 37:562–579

Fuentes-Santos I, González-Manteiga W, Mateu J (2018) A first-order, ratio-based nonparametric separability test for spatio-temporal point processes. Environmetrics 29:1–18

Fukumizu K, Gretton A, Sun X, Schölkopf B (2007) Kernel measures of conditional dependence. J. Platt, D. Koller, Y. Singer, and S. Roweis (Eds.), *Advances in neural information processing systems* (Vol. 20). Curran Associates, Inc

Gabriel E, Diggle PJ (2009) Second-order analysis of inhomogeneous spatio-temporal point process data. Statistica Neerlandica 63:43–51

Gabriel E, Diggle PJ, Rowlingson B, Rodriguez-Cortes FJ (2018) *Stpp: Space-time point pattern simulation, visualisation and analysis.* (R package version 2.0-3)

Ghojogh B, Crowley M, Karray F, Ghodsi A (2023) Elements of dimensionality reduction and manifold learning. Springer, Cham, Switzerland

Ghorbani M (2013) Testing the weak stationarity of a spatio-temporal point process. Stochastic Environonmental Research and Risk Assessesment 27:517–524

Ghorbani M, Vafaei N, Dvořák J, Myllymäki M (2021) Testing the first-order separability hypothesis for spatio-temporal point patterns. Computational Statistics and Data Analysis 161:107245

Gonzalez JA, Hahn U, Mateu J (2019) Analysis of tornado reports through replicated spatio-temporal point patterns. Journal of Royal Statistical Society Series C 69:1–21

Gretton A, Bousquet O, Smola A, Schölkopf B (2005) Measuring statistical dependence with Hilbert-Schmidt norms. Springer-Verlag, In Algorithmic learning theory, pp 63–77

Gretton A, Fukumizu K, Teo CH, Song L, Schölkopf B, Smola AJ (2007) A kernel statistical test of independence. *In Advances in Neural Information Processing Systems (NIPS 20)*, 585-592

Guan Y (2008) On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. Journal of the American Statistical Association 103(483):1238–1247

Iftimi A, Martínez-Ruiz FE, Santiyán AM, Montes F (2015) Spatio-temporal cluster detection of chickenpox in valencia, spain in the period 2008–2012. Geospatial health 10(1):341

Iftimi A, Montes F, Mateu J, Ayyad C (2017) Measuring spatial inhomogeneity at different spatial scales using hybrids of gibbs point process models. Stochastic Environmental Research and Risk Assessment 31:1455–1469

Iftimi A, van Lieshout M, Montes F (2017a) A multi-scale area-interaction model for spatio-temporal point patterns. *spatial statistics*, *26*, 38-55

Keeling M, Woolhouse MEJ, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, Grenfell BT (2001) Dynamics of the 2001 UK foot-and-mouth epidemic dispersal in a heterogeneous landscape. Science 294:813–817

Lavancier F, Møller J, Rubak E (2015) Determinantal point process models and statistical inference. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77(4):853–877

Liu K, Ruan F (2024) On the limitation of kernel dependence maximization for feature selection. ArXiv 2406:06903

Liu Z, Peach LR, Laumann F, Mengod SV, Barahona M (2023) Kernel-based joint independence tests for multivariate stationary and non-stationary time series. *Royal Society Open Science*, *10*

Møller J, Christensen HS, Cuevas-Pacheco F, Christoffersen AD (2021) Structued space-sphere point processes and K-functions. Methodology and Computing in Applied Probability 23:569–591

Møller J, Ghorbani M (2012) Aspects of second-order analysis of structured inhomogeneous spatio-temporal point processes. Statistica Neerlandica 66(4):472–491

Muandet K, Fukumizu K, Sriperumbudur B, Schölkopf B (2017) Kernel mean embedding of distributions: a review and beyond. Foundations and Trends in Machine Learning 10:1–144

Myllymäki M, & Mrkvička T (2023) *GET: Global envelopes in R.* (arXiv:1911.06583 [stat.ME])

Myllymäki M, Mrkvička T, Seijo H, Grabarnik P, Hahn U (2017) Global envelope tests for spatial processes. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79(2):381–404

Nguyen D, Eisenstein J (2017) Kernel independence test for geographical language variation. Computational Linguistics 43:567–592

Pfister N, Bühlmann P, Schölkopf B, Peters J (2017) Kernel-based tests for joint independence. Royal statistical society, Series B 80:5–31

R Core Team (2023). Vienna, Austria

Rustamov RM, Klosowski JT (2020) Kernel mean embedding based hypothesis tests for comparing spatial point patterns. Spatial Statistics 38:100459

Schneider M, Ertel W, Palm G (2016) Constant time expected similarity estimation for large-scale anomaly detection. G.A. Kaminka et al. (Eds.), *ECAI 2016 - 22nd european conference on artificial intelligence, 29 august-2 september 2016, the hague, the netherlands - including prestigious applications of artificial intelligence (PAIS 2016)* (Vol. 285, pp. 12–20)

Schoenberg FP (2004) Testing separability in spatial-temporal marked point processes. Biometrics 60:471–481

Sheather SJ, Jones MC (1991) A reliable data-based bandwidth selection method for kernel density estimation. Journal of Royal Statistical Society Series B 53:683–690

Shen C, Vogelstein J (2021) The exact equivalence of distance and kernel methods in hypothesis testing. AStA Advances in Statistical Analysis 105:385–403

Vafaei N, Ghorbani M, Ganji M, Myllymäki M (2023) Spatio-temporal determinantal point processes. ArXiv 2301:02353

Zhang X, Song L, Gretton A, Smola A (2008) Kernel measures of independence for non-iid data. In Advances in Neural Information Processing Systems 21:1937–1944