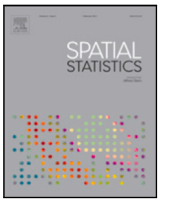
Spatial Statistics 61 (2024) 100824

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



A nonparametric penalized likelihood approach to density estimation of space–time point patterns☆

Blerta Begu [a], Simone Panzeri [b], Eleonora Arnone [c], Michelle Carey [a], Laura M. Sangalli [b,*]

[a] School of Mathematics & Statistics, University College Dublin, Belfield, Dublin 4, Ireland
[b] MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci, 32, Milano, 20133, Italy
[c] Dipartimento di Management, Università degli Studi di Torino, Corso Unione Sovietica, 218 bis, Torino, 10134, Italy

ARTICLE INFO

Keywords:
Space–time density estimation
Intensity estimation of inhomogeneous space–time Poisson point processes
Partial differential equation regularization
Finite element method

ABSTRACT

In this work, we consider space–time point processes and study their continuous space–time evolution. We propose an innovative nonparametric methodology to estimate the unknown space–time density of the point pattern, or, equivalently, to estimate the intensity of an inhomogeneous space–time Poisson point process. The presented approach combines maximum likelihood estimation with roughness penalties, based on differential operators, defined over the spatial and temporal domains of interest. We first establish some important theoretical properties of the considered estimator, including its consistency. We then develop an efficient and flexible estimation procedure that leverages advanced numerical and computation techniques. Thanks to a discretization based on finite elements in space and B-splines in time, the proposed method can effectively capture complex multi-modal and strongly anisotropic spatio-temporal point patterns; moreover, these point patterns may be observed over planar or curved domains with non-trivial geometries, due to geographic constraints, such as coastal regions with complicated shorelines, or curved regions with complex orography. In addition to providing estimates, the method's functionalities also include the introduction of appropriate uncertainty quantification tools. We thoroughly validate the proposed method, by means of simulation studies and applications to real-world data. The obtained results highlight significant advantages over state-of-the-art competing approaches.

1. Introduction

Geo-referenced event data, recorded continuously over time, are attracting increasing interest in many scientific fields, including epidemiology, seismology, social sciences, and many others. Examples of typical real applications include disease surveillance, occurrence analysis of extreme natural events, crime hot-spot detection, and roadway traffic crash mapping. Spatio-temporal event data are also frequently encountered in environmental sciences, for various purposes, including studying the presence of rare plant species, tracking the migration phenomena of endangered animals, and monitoring the occurrence of respiratory diseases due to pollution or global warming. In engineering and additive manufacturing, spatio-temporal event data are, for instance, observed in multi-target tracking problems and in the real-time monitoring of defects in selective laser melting processes for 3D printing.

☆ Blerta Begu and Simone Panzeri contributed equally to this work. Supplementary material providing the code and data necessary to reproduce the results presented in the simulation and application studies is available online.

* Corresponding author.
E-mail address: laura.sangalli@polimi.it (L.M. Sangalli).
URL: https://sangalli.faculty.polimi.it (L.M. Sangalli).

https://doi.org/10.1016/j.spasta.2024.100824
Received 1 November 2023; Received in revised form 5 March 2024; Accepted 19 March 2024
Available online 18 April 2024

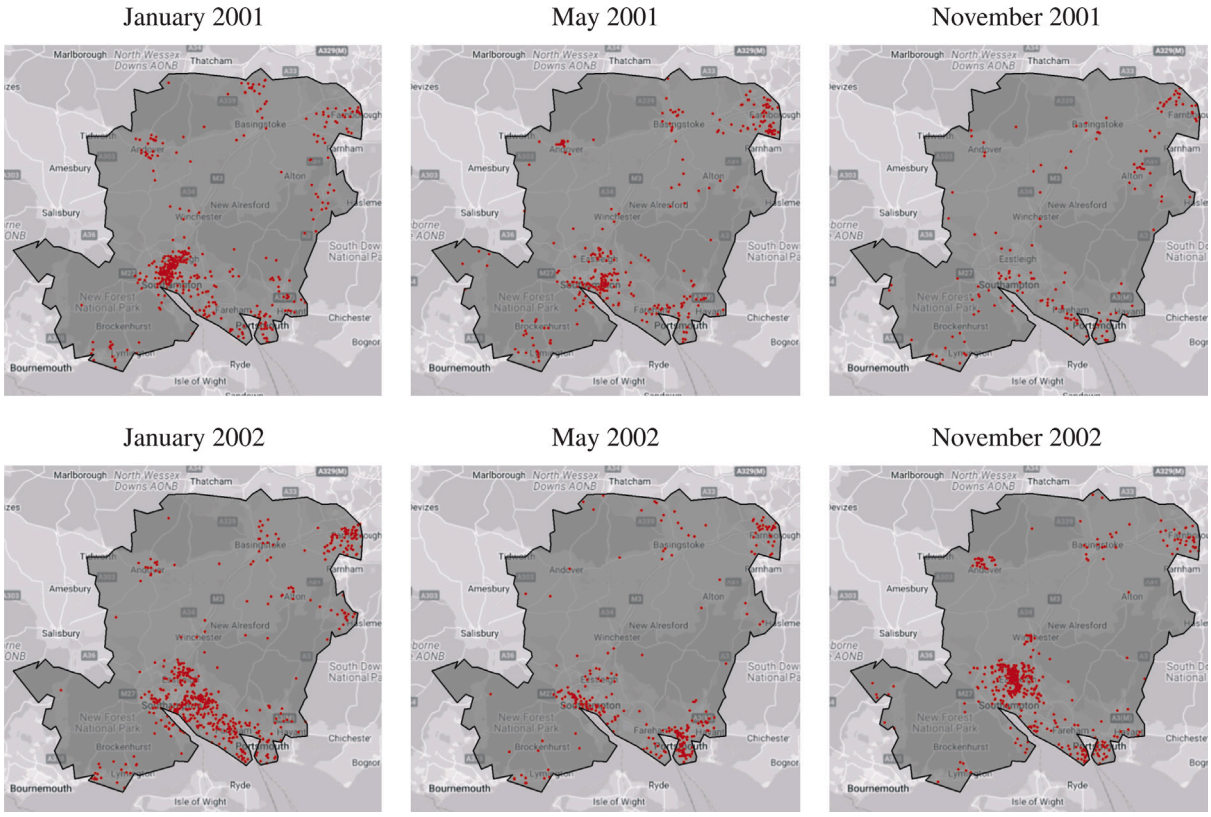


Fig. 1. Locations of gastrointestinal infections in Hampshire county, UK, in some months of 2001 and 2002. Although we here plot events on a monthly basis, the data are observed continuously over time (and space). The phenomenon is expressed differently across the irregular spatial domain, resulting for instance in different behaviors along the western and eastern shores of Southampton Water estuary. The background map was obtained using the `RgoogleMaps` package (Loecher, 2023).

Fig. 1 displays a benchmark application in epidemiology, concerning the evolution of non-specific gastrointestinal infections over time. Points correspond to the locations of cases reported to the National Health Service (NHS) Direct in Hampshire county, UK, in the period 2001–2003. Data are publicly available and collected during a collaborative surveillance project in the South of England; see Diggle et al. (2009) and Kaimi and Diggle (2011). The irregular shape of the spatial domain influences the disease spread, resulting for instance in different behaviors along the western and eastern shores of Southampton Water estuary. Fig. 2 illustrates instead an application in seismology. We here consider the locations of earthquakes, with magnitude greater than 4.75, that took place worldwide during 2020–2022. The dataset is sourced from the earthquake catalog of the United States Geological Survey (USGS). Since earthquakes result from the sudden release of energy, caused by the rupture of geological faults in the Earth's crust, points are mainly distributed along these elongated lines. In these varied frameworks, one crucial research interest lies in determining the underlying spatio-temporal evolution of the phenomenon. This can be rephrased in terms of estimating the unknown spatio-temporal probability density, or equivalently estimating the intensity of an inhomogeneous space–time point process, that generates the data.

Spatial point pattern analysis has been the object of an extensive and detailed literature, with a wealth of different models proposed for the study of point distribution over space. The reader may refer, for example, to Fishman and Snyder (1976), Daley and Vere-Jones (2003, 2008), Streit and Streit (2010), Baddeley et al. (2015), Møller and Waagepetersen (2007, 2017), and references therein. Much interest has been devoted to the investigation of methods for spatio-temporal point patterns (see, e.g., Diggle, 2013; González et al., 2016). Spatio-Temporal Kernel Density Estimation (STKDE) has been proposed by Zhang et al. (2011) and implemented in the R packages `stam` (Zhang, 2010) and `sparr` (Davies and Marshall, 2023). STKDE is utilized in various contexts, such as crime hot-spot prediction (Hu et al., 2018), epidemiology (Elson et al., 2020, 2021), and wildlife roadkill hot-spot detection for mitigation measures (Medinas et al., 2021). Intensity estimation in spatio-temporal Log Gaussian-Cox Processes (LGCP) has been described in Diggle et al. (2013), Shirota and Gelfand (2017), and implemented in the `lgcp` R package (Taylor et al., 2023). A very rich literature concerns spatio-temporal Bayesian hierarchical models based on Integrated Nested Laplace Approximations (INLA, Rue et al., 2009) and stochastic Partial Differential Equations (sPDE, Lindgren et al., 2011), reviewed in Blangiardo and Cameletti (2015), Krainski et al. (2018), Gómez-Rubio (2020), and implemented in the packages `R-INLA` (https://www.r-inla.org/) and `inlabru` (Lindgren and Bachl, 2023; Bachl et al., 2019). To mention just some of them, the reader may refer to Yuan et al.

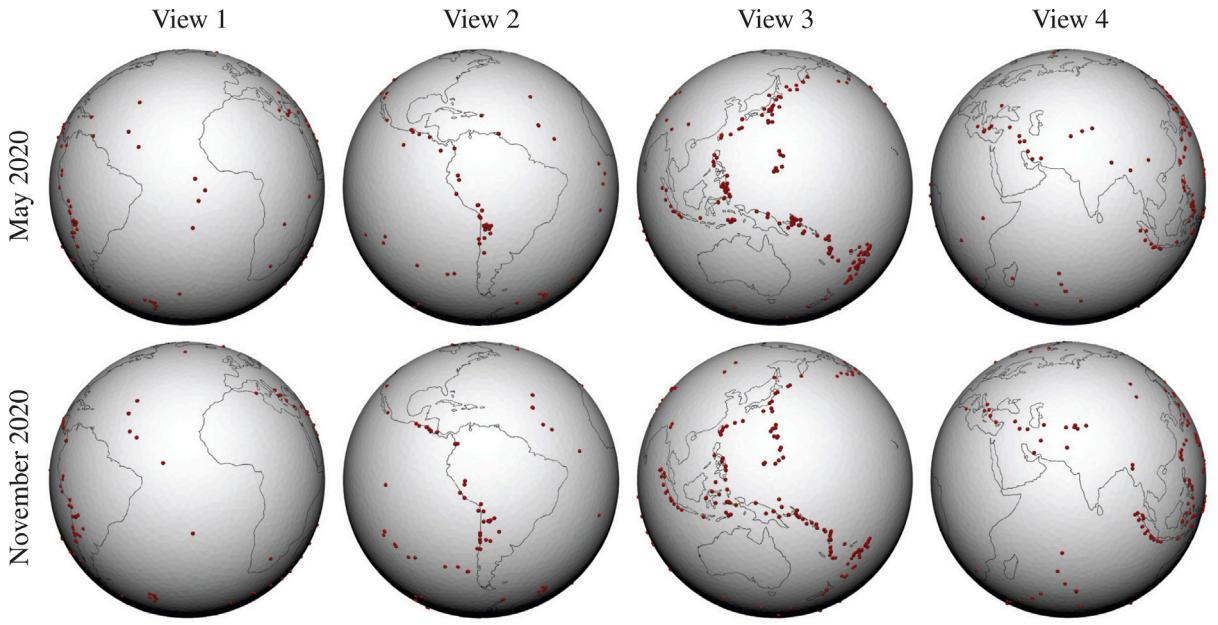


Fig. 2. Locations of earthquakes with a magnitude greater than 4.75 that occurred worldwide in two different months of 2021. Although we here plot events on a monthly basis, the data are observed continuously over time (and space). The pattern is highly localized and skewed along the faults in the Earth's crust, where earthquakes originate by the action of plate tectonic forces.

(2017), Python et al. (2019), Serra et al. (2014), Rodríguez de Rivera et al. (2018), Ugarte et al. (2014) and Valente and Laurini (2020).

Many methods have been proposed to address other research questions, concerning space–time point processes. Studies on spatio-temporal LGCPs have focused on advanced inference and the estimation of second-order parameters, using various approaches, including the joint minimum contrast fitting method and likelihood-oriented techniques (see, e.g., D'Angelo et al., 2023; Rodrigues and Diggle, 2012; Bayisa et al., 2020; Waagepetersen et al., 2016). Alternative research endeavors introduce novel classes of models for spatio-temporal point processes. Li and Guan (2014) proposed generalized linear mixed models for latent spatio-temporal point processes, incorporating functional principal component analysis and an empirical Bayes approach. Chen et al. (2021) combines continuous-time neural ordinary differential equations, to jointly capture spatial and temporal dependencies along with their propagation effects. Other works develop estimators of the spatio-temporal inhomogeneous K–function, and pair correlation function, to describe second-order characteristics of point processes (see, e.g., Gabriel, 2014; Hering et al., 2009).

Due to the inherent complexity of space–time settings, some of these state-of-the-art methodologies may come with some limitations, either for theoretical or computational reasons. Some methods, for instance, are defined to handle event data continuously over space and time, but their implementation requires discrete time instances. Most techniques, though theoretically defined for generic multivariate spatial domains, have implementations that work on regular planar spatial domains or spherical domains. Notable exceptions in this regards are for instance offered by methods implemented in `R-INLA`, and by the STKDE implementation in `sparr` (Davies and Marshall, 2023), that can handle space–time data recorded across irregularly shaped planar regions (Bertin et al., 2023). Other techniques make assumptions that allow for the simplification of the complexity characterizing the spatio-temporal phenomenon under study. One of these is the first-order separability hypothesis for spatio-temporal point processes, which requires that the density or intensity estimate can be factorized in spatial and temporal components. This is quite a strong assumption in the considered framework, that should be adequately investigated on a case-by-case basis, depending on the specific data at hand, by means of statistical tests, such as those proposed by Ghorbani et al. (2021).

In this work, we present an innovative methodology for Spatio-Temporal Density (and intensity) Estimation, with Partial Differential Equation Regularization, referred to as STDE-PDE. We provide an effective tool for estimating the unknown probability density on the spatio-temporal domain of interest, starting from an independent and identically distributed sample. This estimation problem is equivalent to intensity estimation of an inhomogeneous space–time Poisson point process. STDE-PDE combines a nonparametric maximum likelihood approach, with a regularization based on differential operators, in space and time, with the aim of balancing adherence to data and smoothness of the estimate. The method handles event data continuously over space and time, and does not make any assumption about the shape of the underlying density or intensity, nor about the separability in space and time. The estimation procedure relies on advanced optimization procedures and numerical analysis techniques, based on finite elements in space and B-splines in time. These choices ensure efficiency and flexibility, making the proposed STDE-PDE capable of capturing complex multi-modal and strongly anisotropic spatio-temporal point patterns, such as those characterizing the seismology application in Fig. 2. Moreover, these point patterns may be observed over planar or curved domains endowed by

non-trivial geometries, due to geographic constraints; this includes coastal regions with complicated shorelines, as in the case of the epidemiology data in Fig. 1, or curved spatial regions with complex orography.

For the underlying mathematical structure, STDE-PDE falls within the research field of nonparametric density estimation methods with roughness penalties. This field has so far been explored in space-only settings, by Silverman (1982), Gu and Qiu (1993), Gu (1993), Gu and Wang (2003) and Gu et al. (2013) for regular spatial domains, by Ferraccioli et al. (2021) for irregular planar domains, and by Arnone et al. (2022) for Riemannian manifold domains. In particular, we here extend the works of Ferraccioli et al. (2021) and Arnone et al. (2022), by incorporating the time dimension. This extension involves the utilization of roughness penalties in both space and time, similar to those commonly employed in semiparametric regression settings (Augustin et al., 2013; Marra et al., 2012; Bernardi et al., 2017; Arnone et al., 2023c). In contrast to Ferraccioli et al. (2021) and Arnone et al. (2022), who did not devise uncertainty quantification tools, we explicitly focus on addressing this aspect. In particular, following a Bayesian analogy, we develop approximate confidence bands for the spatio-temporal density or intensity estimate, based on the proposed estimator. With necessary simplifications, the proposed inference approach can also be applied in the space-only settings considered by Ferraccioli et al. (2021) and Arnone et al. (2022). Finally, STDE-PDE is a new addition to the class of statistical methods with PDE regularization originally inspired by Ramsay (2000, 2002), Sangalli et al. (2013), and reviewed in Sangalli (2021). The implementation of the proposed model further extends the functionalities of the R/C++ `fdaPDE` library (Arnone et al., 2023a), providing ready-to-use tools to deal with space–time point processes.

The remainder of the paper is organized as follows. Section 2 introduces the proposed statistical model. Section 3 delves into the theoretical properties, including the well-posedness of the estimation problem, and the consistency of the associated estimator. Section 4 addresses some necessary numerical aspects, focusing on the discretization procedure and briefly commenting on some optimization aspects. Section 5 introduces pointwise confidence intervals for the space–time density estimate, based on a Bayesian approach. Section 6 presents two simulation studies, comparing STDE-PDE with other state-of-the-art methodologies. Section 7 showcases the use of STDE-PDE in the two applications to real-world data, from epidemiology and seismology, illustrated in Figs. 1 and 2. Finally, Section 8 draws concluding remarks and outlines potential directions for future research.

2. Space–time density and intensity estimation

Let $\mathcal{D} \subset \mathbb{R}^d$, where $d = 2, 3$, and $[0, T] \subset \mathbb{R}^+$, be the spatial and temporal domains of interest. Depending on the specific application, $\mathcal{D}$ can be a compact (but possibly non-convex) subset of $\mathbb{R}^2$, or a two-dimensional Riemannian manifold, embedded in $\mathbb{R}^3$; see, e.g., Jost (2008). Other possible forms of $\mathcal{D}$ will be discussed in the concluding section. Define the spatio-temporal domain $\mathcal{D}_T := \mathcal{D} \times [0, T]$. Let the pairs $\{(\mathbf{p}_i, t_i)\}_{i=1}^n \in \mathcal{D}_T$ be n independent and identically distributed observations, drawn from an unknown distribution with density function $f : \mathcal{D}_T \to \mathbb{R}^+$, to be estimated. In order to estimate f, we adopt a nonparametric likelihood approach, which does not assume the form of f to be known *a priori*, and makes use of roughness penalties to ensure regularity of the estimate. This approach extends to space–time settings the framework considered for density estimation over spatial domains in Silverman (1982), Goodd and Gaskins (1971), Tapia and Thompson (1978), Gu and Qiu (1993), Ferraccioli et al. (2021) and Arnone et al. (2022). We parameterize the estimation problem in terms of the log-density $g = \log f$, to naturally satisfy the non-negativity constraint on f. Specifically, we propose to estimate f by finding the log-density function $g = \log f$ that minimizes the following penalized negative log-likelihood:

$$L(g) := -\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{p}_i, t_i) + \int_{\mathcal{D}_T} e^{g(\mathbf{p},t)}\, d\mathbf{p}\, dt + R(g; \lambda), \tag{1}$$

where $R(g; \lambda)$ is an appropriate roughness penalty, which depends on the smoothing parameters λ, as detailed below. Note that the first term in (1) is the negative log-likelihood. The second term instead guarantees that f has unitary integral over $\mathcal{D}_T$, thus ensuring that f is a density with respect to the Lebesgue measure over $\mathcal{D}_T$; see the detailed explanation in Appendix A. Finally, the regularization term $R(g; \lambda)$ in (1) prevents degenerate solutions, such as the sums of Dirac delta functions centered at the observations, which would lead to unbounded likelihoods. We here consider a flexible regularization $R(g; \lambda)$, comprising two terms that enable the control of the regularity of the estimate in both space and time,

$$R(g; \lambda) := \lambda_{\mathcal{D}} \int_{\mathcal{D}_T} \left(\Delta_{\mathcal{D}} g(\mathbf{p}, t)\right)^2 d\mathbf{p}\, dt + \lambda_T \int_{\mathcal{D}_T} \left(\frac{\partial^2 g(\mathbf{p}, t)}{\partial t^2}\right)^2 d\mathbf{p}\, dt, \tag{2}$$

where $\Delta_{\mathcal{D}}$ is the Laplace or Laplace–Beltrami operator, and $\lambda = \left[\lambda_{\mathcal{D}}, \lambda_T\right]^\top$ are smoothing parameters, with $\lambda_{\mathcal{D}}, \lambda_T > 0$. The Laplace and Laplace–Beltrami operators, respectively for planar and curved spatial domains $\mathcal{D}$, measure the local curvature of g over $\mathcal{D}$; similarly, the second derivative in time measures the local curvature of g over the temporal domain $[0, T]$. The presence of the two smoothing parameters, $\lambda_{\mathcal{D}}$ and λ_T, provides a means to control the regularity in space and time. These smoothing parameters balance the trade-off between the log-likelihood, responsible for the model's adaptation to the data, and the regularization term, which seeks reduced variation in g. The penalty in (2) is also common in regression contexts (see, e.g., Marra et al., 2012; Augustin et al., 2013), and permits the derivation of the large sample properties of the estimators, through suitable generalization of the arguments used in more classical spline settings by Silverman (1982).

Remark 1. The above model for density estimation can equivalently be reformulated in terms of intensity estimation of an inhomogeneous Poisson space–time point process on $\mathcal{D}_T$, as detailed in Appendix B (see also the simulation setting in Appendix F).

3. Theoretical properties

In this section, we address the main theoretical properties related the proposed method. In Section 3.1, we define the appropriate functional embedding for the considered estimation problem, and we address its well-posedness. In Section 3.2, we focus on the consistency of the estimator.

3.1. Functional embedding and well-posedness of the estimation problem

Let $L^2(\mathcal{D})$ denote the space of square integrable functions over the spatial domain $\mathcal{D}$, and consider the Sobolev space $H^k(\mathcal{D}) := \{u \in L^2(\mathcal{D}) : \ D^\alpha u \in L^2(\mathcal{D}) \ \forall |\alpha| \leq k\}$, equipped with the standard norm $\|u\|^2_{H^k(\mathcal{D})} := \sum_{|\alpha|\leq k} \|D^\alpha u\|^2_{L^2(\mathcal{D})}$, with D^α denoting the weak derivative of order α (see, e.g., Brezis, 2011). Then, consider the time-dependent Sobolev space defined as follows:

$$L^2(0,T;H^k(\mathcal{D})) := \Bigg\{ u(t,\boldsymbol{p}) : \ u(t,\cdot) \in H^k(\mathcal{D}) \text{ for almost every } t \in [0,T] \\ \text{and } \int_{\mathcal{D}} |u(t,\boldsymbol{p})|^2 \, d\boldsymbol{p} < \infty, \ \int_0^T \|u(t,\boldsymbol{p})\|^2_{H^k(\mathcal{D})} \, dt < \infty \Bigg\}.$$

When endowed with the norm $\|u\|^2_{L^2(0,T;H^k(\mathcal{D}))} := \int_0^T \|u(t,\boldsymbol{p})\|^2_{H^k(\mathcal{D})} \, dt$, the space $L^2(0,T;H^k(\mathcal{D}))$ is a Banach space (see, e.g., Adams and Fournier, 2003, p. 89). Now, take the space

$$V^{2,k}(0,T;\mathcal{D}) := \Big\{ u \in L^2(0,T;H^k(\mathcal{D})) : \ u_t, u_{tt} \in L^2(0,T;L^2(\mathcal{D})) \Big\}, \tag{3}$$

where u_t and u_{tt} denote the first- and second-order derivatives of u with respect to time. Finally, define V as the space of functions in $V^{2,k}(0,T;\mathcal{D})$ with $k=2$, subject to homogeneous Neumann conditions at the boundary of the spatial domain and a null velocity condition at $t=0$ and at $t=T$, that is:

$$V := \Big\{ u \in V^{2,2}(0,T;\mathcal{D}) : \ \big(\nabla_{\mathcal{D}} u\big)(t) \cdot \boldsymbol{n} = 0 \text{ for } t \in [0,T] \ \text{ and } \ u_t(0,\boldsymbol{p}) = u_t(T,\boldsymbol{p}) = 0 \text{ for } \mathbf{p} \in \mathcal{D} \Big\}, \tag{4}$$

where $\boldsymbol{n}$ denotes the outward orthonormal vector to the boundary $\partial\mathcal{D}$ of the spatial domain $\mathcal{D}$. In this space, we consider the associated norm:

$$\|u\|^2_V := \|u\|^2_{L^2(0,T;H^2(\mathcal{D}))} + \|u_t\|^2_{L^2(0,T;L^2(\mathcal{D}))} + \|u_{tt}\|^2_{L^2(0,T;L^2(\mathcal{D}))}. \tag{5}$$

The presented framework holds also for bounded and closed two-dimensional Riemannian manifolds $\mathcal{D}$. In particular, closed Riemannian manifolds are compact manifolds without boundaries, such as for instance the surface of a sphere. More specifically, when $\mathcal{D}$ is a closed manifold, then the functional space in (4) reduces to $V = \big\{u \in V^{2,2}(0,T;\mathcal{D}) : \ u_t(0,\boldsymbol{p}) = 0 \text{ for } \mathbf{p} \in \mathcal{D}\big\}$, where spatial boundary conditions are not needed.

The following theorem states that the density estimator introduced in Section 2 is well-posed in V.

Theorem 1. *The penalized log-likelihood $L(g)$ defined in (1) has a unique minimizer $\hat{g}$ in V.*

Proof. The proof is deferred to Appendix C. □

3.2. Asymptotic properties of the estimator

We now investigate the infill asymptotic properties of the proposed estimator, as the number of observations n tends to infinity. Let $g_0 = \log f_0$ be the true log-density. Denote by $D_{KL}(g_1,g_2)$ the Kullback–Leibler divergence between two log-densities $g_1 = \log f_1$ and $g_2 = \log f_2$ (see, e.g., Csiszár, 1975). Moreover, define the symmetrized Kullback–Leibler distance as

$$\begin{aligned} D_{sKL}(g_1,g_2) &:= D_{KL}(g_1,g_2) + D_{KL}(g_2,g_1) = \int_{\mathcal{D}_T} e^{g_1} \log\left(\frac{e^{g_1}}{e^{g_2}}\right) d\mathbf{p}\, dt + \int_{\mathcal{D}_T} e^{g_2} \log\left(\frac{e^{g_2}}{e^{g_1}}\right) d\mathbf{p}\, dt \\ &= \int_{\mathcal{D}_T} (g_1 - g_2) e^{g_1} \, d\mathbf{p}\, dt + \int_{\mathcal{D}_T} (g_2 - g_1) e^{g_2} \, d\mathbf{p}\, dt = \mathbb{E}_{g_1}[g_1 - g_2] + \mathbb{E}_{g_2}[g_2 - g_1], \end{aligned} \tag{6}$$

where we omit the integration variables $\mathbf{p}$ and t, to lighten the notation, and $\mathbb{E}_g[h] := \int_{\mathcal{D}_T} h e^g \, d\mathbf{p}\, dt$ represents the mean of $h(X)$ when the random variable X has log-density g.

To help the derivation of the result, we introduce the linear approximation $\hat{g}_*$ of $\hat{g}$, that is calculated by minimizing the expression

$$L_*(g) := -\frac{1}{n}\sum_{i=1}^n g(\mathbf{p}_i,t_i) + 1 + \mathbb{E}_{g_0}[g] + \mathrm{Var}_{g_0}(g - g_0) + \lambda_{\mathcal{D}} \int_{\mathcal{D}_T} (\Delta_{\mathcal{D}} g(\mathbf{p},t))^2 \, d\mathbf{p}\, dt + \lambda_T \int_{\mathcal{D}_T} \left(\frac{\partial^2 g(\mathbf{p},t)}{\partial t^2}\right)^2 d\mathbf{p}\, dt.$$

In essence, $L_*(g)$ is the quadratic approximation of $L(g)$ at g_0, meaning that the Taylor expansion of L and L_* coincide up to the second order when evaluated at g_0. The introduction of L_* and $\hat{g}_*$ is prompted by the observation that $D_{sKL}(\hat{g},g_0)$ can be divided into two components, $D_{sKL}(\hat{g},\hat{g}_*)$ and $D_{sKL}(\hat{g}_*,g_0)$, whose asymptotic behaviors are simpler to investigate.

We now make some standard assumptions on the smoothness of g_0, to simplify the derivation of the asymptotic results.

Assumption 1. The true log-density g_0 is at least twice differentiable over the spatio-temporal domain $\mathcal{D}_T$. Moreover, g_0 is bounded above and below, and it is such that $R(g_0; \lambda) < \infty$.

Assumption 2. For g in a convex set $B_0 \subset \mathcal{D}_T$ around g_0 containing $\hat{g}$ and $\hat{g}_*$, there exists a positive constant c such that $c\mathrm{Var}_{g_0} \leq \mathrm{Var}_g$ uniformly with respect to g.

Assumption 1 is a standard requirement for the consistency of density estimators (Silverman, 1982). It guarantees the equivalence between the weighted $L^2(0,T;L^2(\mathcal{D}))$–norm with the density function $f_0 = e^{g_0}$, that is $\int_{\mathcal{D}_T} h^2 f_0(\mathbf{p},t)\,d\mathbf{p}\,dt$, and the conventional $L^2(0,T;L^2(\mathcal{D}))$–norm. Assumption 2 is a generalization to the spatio-temporal setting of Assumption A.3 in Gu (1993). This condition is satisfied whenever the members of B_0 are bounded from above and below. Essentially, Assumption 2 requires Assumption 1 to be satisfied by functions that are near g_0, and in particular by $\hat{g}$ and $\hat{g}_*$.

For the sake of simplicity, in the following consistency result, we consider the case where $\lambda_D = \lambda_T = \lambda$. However, in general, the result holds for $\lambda_T = O(\lambda_D)$.

Theorem 2. *Under Assumptions 1 and 2, as $\lambda \to 0$ and $n\lambda^{1/2} \to \infty$, the estimator $\hat{g} \in V$ minimizing $L(g)$ in (1) is consistent with respect to the symmetrized Kullback–Leibler distance. Moreover,*

$$D_{sKL}(\hat{g}, g_0) = O(n^{-1}\lambda^{-1/2} + \lambda).$$

Proof. The proof is deferred to Appendix D. □

Remark 2. Theorem 2 states the consistency of the proposed STDE-PDE estimator in the symmetrized Kullback–Leibler distance, defined in (6). It should be pointed out that this distance controls other commonly used distances for density functions, such as the total variation and the Hellinger distances (see, e.g., Pollard, 2002; Silverman, 1982). This entails that the proposed estimator is also consistent with respect to the norms just mentioned.

4. Estimation procedure

In this section, we focus on the numerical strategies employed to effectively address the problem outlined in Section 2. Section 4.1 faces the aspects related to the discretization procedure, aiming to approximate the quantities involved in the estimation functional (1). Section 4.2 deals with the minimization of the estimation functional, in its discretized form, encompassing topics such as optimization algorithms and initialization procedures. Section 4.3 discusses the choice of the smoothing parameters.

4.1. Discretization of the objective functional

Our goal is to determine the minimizer of the functional $L(g)$ in (1) within the space of spatio-temporal density functions V in (4). However, this infinite-dimensional problem lacks a closed-form analytical solution. Hence, we must seek a numerical approximation. To achieve this, we proceed with a suitable discretization procedure, based on a finite element method in space and B-splines in time.

We discretize the spatial domain $\mathcal{D}$ through Delaunay triangulations (see, e.g., Hjelle and Dæhlen, 2006). Fig. 3 illustrates the triangulations for the spatial domains in the two applications under consideration. Such triangulation can for instance be created starting from the boundary $\partial\mathcal{D}$, and can be refined or coarsened according to criteria concerning maximum allowed triangle area, minimum allowed triangle angle, distribution of the data points and/or geometry of the domain. The library `fdaPDE` (Arnone et al., 2023a) includes various functionalities to create, refine or coarsen the mesh. To effectively represent point patterns with highly localized features, while employing a limited number of mesh nodes (and thus a reduced computational cost), it is advantageous to build adaptive data-driven spatial meshes. In this regard, for instance, for the application to earthquakes on the globe, we rely on the coarsening strategy presented by Dassi et al. (2015). Specifically, we start from a fine and regular mesh, and iteratively simplify it, taking into account both the surface's geometry and the data distribution, to obtain a mesh that is finer where more events are observed, while respecting the shape of the domain. As described in Dassi et al. (2015), such strategy also improves the inferential properties of the resulting estimators.

Let $\mathcal{T}_\mathcal{D}$ denote the triangulation and $V^r_{\mathcal{T}_\mathcal{D}}$ the space of piecewise polynomial functions of degree r over $\mathcal{T}_\mathcal{D}$. Formally:

$$V^r_{\mathcal{T}_\mathcal{D}} := \left\{ h_{\mathcal{T}_\mathcal{D}} \in C(\mathcal{T}_\mathcal{D}) : h_{\mathcal{T}_\mathcal{D}}\big|_\tau \in \mathbb{P}^r \ \forall \tau \in \mathcal{T}_\mathcal{D} \right\},$$

where $\mathbb{P}^r$ is the space of polynomials of degree r. Let $\{\boldsymbol{\xi}_l\}_{l=1}^K$ be the nodes of the mesh, which coincide with the vertices of the triangles in $\mathcal{T}_\mathcal{D}$ in the linear case. Additionally, let $\{\psi_k\}_{k=1}^K$ denote the finite element basis functions of $V^r_{\mathcal{T}_\mathcal{D}}$: these functions $\psi_k(\mathbf{p}) \in V^r_{\mathcal{T}_\mathcal{D}}$ are piecewise polynomials of degree r, and satisfy $\psi_k(\boldsymbol{\xi}_l) = \delta_{kl}$ for $k,l = 1,\dots,K$ (see, e.g., Quarteroni, 2009, for an introduction to finite elements). The library `fdaPDE` (Arnone et al., 2023a) currently supports linear and quadratic finite elements ($r = 1, 2$), over planar, surface as well as volumetric domains.

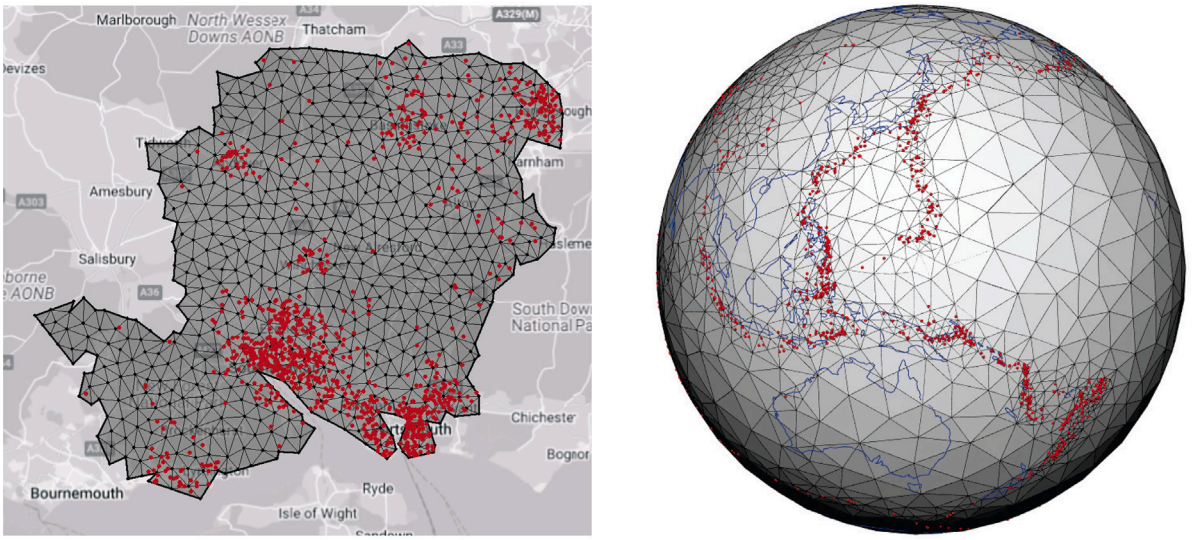


Fig. 3. Left: regular spatial mesh with 778 nodes of the Southampton region, Hampshire county, UK. Right: data-driven adaptive spatial mesh with 2000 nodes over the Earth's surface. Red points are 1000 locations randomly sampled among the data available in the original datasets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

For the temporal dimension, we adopt cubic B-spline functions $\{\varphi_m\}_{m=1}^{M}$, defined over $[0,T]$; see De Boor (1978) for a theoretical treatment. With the considered spatial and functional bases, we can now represent the spatio-temporal log-density function $g(\mathbf{p},t)$ by the basis expansion

$$g(\mathbf{p},t) = \sum_{k=1}^{K}\sum_{m=1}^{M} c_{km}\psi_k(\mathbf{p})\varphi_m(t), \tag{7}$$

where $\{c_{km} : k = 1,\dots,K \text{ and } m = 1,\dots,M\}$ are the expansion coefficients to be stored in $\mathbf{c} := [c]_{km}$. As a result, the original infinite-dimensional problem of estimating the function $g \in V^r_{\mathcal{T}_D}$ is now transformed into a finite-dimensional estimation problem of estimating the vector of coefficients $\mathbf{c}$. To this end, we need a discretization of the objective functional $L(g)$ in (1). We introduce the vector of K spatial basis functions $\boldsymbol{\psi} := \begin{bmatrix}\psi_1 & \dots & \psi_K\end{bmatrix}^\top$, the vector of M temporal basis functions $\boldsymbol{\varphi} := \begin{bmatrix}\varphi_1 & \dots & \varphi_M\end{bmatrix}^\top$, and the matrices $\Psi := [\Psi]_{ik} = \psi_k(\boldsymbol{p}_i) \in \mathbb{R}^{n\times K}$ and $\Phi := [\Phi]_{im} = \varphi_m(t_i) \in \mathbb{R}^{n\times M}$ that store the evaluations of the K spatial basis functions at the n spatial points $\{\mathbf{p}_i\}_{i=1}^n$, and the evaluations of the M temporal basis functions at the n temporal instances $\{t_i\}_{i=1}^n$. Moreover, we introduce the mass, stiffness, and penalty matrices, in space and time, as follows:

$$R_0 := \int_{\mathcal{T}_D} \boldsymbol{\psi}\boldsymbol{\psi}^\top \in \mathbb{R}^{K\times K}, \qquad K_0 := \int_0^T \boldsymbol{\varphi}\boldsymbol{\varphi}^\top \in \mathbb{R}^{M\times M}, \qquad R_1 := \int_{\mathcal{T}_D} (\nabla_{\mathcal{T}_D}\boldsymbol{\psi})(\nabla_{\mathcal{T}_D}\boldsymbol{\psi})^\top \in \mathbb{R}^{K\times K}$$

$$P_D := R_1^\top R_0^{-1} R_1 \in \mathbb{R}^{K\times K}, \qquad P_T := \int_0^T \boldsymbol{\varphi}_{tt}\boldsymbol{\varphi}_{tt}^\top \in \mathbb{R}^{M\times M}, \qquad P := 2\lambda_D(K_0 \otimes P_D) + 2\lambda_T(P_T \otimes R_0) \in \mathbb{R}^{KM\times KM}$$

where $[(\nabla_{\mathcal{T}_D}\boldsymbol{\psi})(\nabla_{\mathcal{T}_D}\boldsymbol{\psi})^\top]_{ij} = \nabla_{\mathcal{T}_D}\psi_j \cdot \nabla_{\mathcal{T}_D}\psi_i$ and $\boldsymbol{\varphi}_{tt} := \begin{bmatrix}\partial^2\varphi_1/\partial t^2 & \dots & \partial^2\varphi_M/\partial t^2\end{bmatrix}^\top$ is the vector of second-order derivatives of the temporal bases with respect to time.

We are now ready to discretize the functional $L(g)$.

Approximation of the log-likelihood. Let $\Upsilon \in \mathbb{R}^{n\times KM}$ be the matrix whose ith row in $\mathbb{R}^{1\times KM}$ is defined as the Kronecker product between the ith rows of Φ and Ψ. Denoting by $\mathbf{1}$ the unitary vector of length n, we can discretize the first term of $L(g)$ as:

$$-\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{p}_i,t_i) \approx -\frac{1}{n}\mathbf{1}^\top\Upsilon\mathbf{c}.$$

Approximation of the integral of the density function. We employ a standard Gaussian quadrature rule, with $q = 6$ quadrature nodes and associated weights $\mathbf{w}_D \in \mathbb{R}^q$ in space, and $r = 5$ quadrature nodes and associated weights $\mathbf{w}_T \in \mathbb{R}^r$ in time (Quarteroni, 2009). Therefore, given a right-open interval $\iota \subset [0,T]$, with endpoints coinciding with two consecutive knots of the chosen temporal discretization, we have:

$$\int_{D_T} e^{g(\mathbf{p},t)}\, d\mathbf{p}\, dt \approx \sum_{\iota\subset[0,T]}\sum_{\tau\in\mathcal{T}_D} \kappa(\mathbf{w}_T \otimes \mathbf{w}_D)^\top \exp\{(\Phi_\iota \otimes \Psi_\tau)\mathbf{c}\},$$

where: $\Psi_\tau \in \mathbb{R}^{q\times K}$ is the matrix containing the evaluations of the K spatial basis functions at the q spatial quadrature nodes in the triangle $\tau \in \mathcal{T}_D$; $\Phi_\iota \in \mathbb{R}^{r\times M}$ is the matrix having as entries the evaluations of the M temporal basis functions at the r temporal quadrature nodes; κ is the determinant of the transformation matrix mapping the reference triangle $\{(0,0),(0,1),(1,0)\}$ onto the triangle $\tau \in \mathcal{T}_D$ and the reference interval $[-1,1]$ onto the interval $\iota \subset [0,T]$.

Approximation of the penalization terms. In analogy with the works by Bernardi et al. (2017) and Arnone et al. (2023c), in spatio-temporal regression contexts, the discretization of the penalization term in space and time can be written as

$$\lambda_D \int_{D_T} \left(\Delta_D g(\mathbf{p},t)\right)^2 d\mathbf{p}\,dt + \lambda_T \int_{D_T} \left(\frac{\partial^2 g(\mathbf{p},t)}{\partial t^2}\right)^2 d\mathbf{p}\,dt \approx \lambda_D \mathbf{c}^\top (K_0 \otimes P_D)\mathbf{c} + \lambda_T \mathbf{c}^\top (P_T \otimes R_0)\mathbf{c}.$$

To summarize, the minimization of the objective functional $L(g)$ over the density functions $g \in V$ is numerically addressed as the minimization over the coefficients $\mathbf{c} \in \mathbb{R}^{KM}$ of the discretized objective functional $\tilde{L}(\mathbf{c})$, defined as:

$$\tilde{L}(\mathbf{c}) := -\mathbf{1}^\top Y \mathbf{c} + n \sum_{\iota\subset[0,T]} \sum_{\tau\in\mathcal{T}_D} \kappa(\mathbf{w}_T \otimes \mathbf{w}_D)^\top \exp\{(\Phi_\iota \otimes \Psi_\tau)\,\mathbf{c}\} + \lambda_D \mathbf{c}^\top (K_0 \otimes P_D)\mathbf{c} + \lambda_T \mathbf{c}^\top (P_T \otimes R_0)\,\mathbf{c}. \tag{8}$$

4.2. Minimization of the discretized estimation functional

We now cover the main aspects connected to the minimization of the functional $\tilde{L}(\mathbf{c})$ in (8). In Lemma 1 in Appendix C, we prove that the non-linear functional $L(g)$ in (1) is strictly convex. This entails that the associated minimization problem

$$\hat{\mathbf{c}} := \arg\min_{\mathbf{c}\in\mathbb{R}^{KM}} \tilde{L}(\mathbf{c}) \tag{9}$$

can be solved using standard iterative optimization algorithms (see, e.g. Nocedal and Wright, 1999; Lange, 2013). The `fdaPDE` library implements several optimization algorithms, with the default option being the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method, described in Liu and Nocedal (1989) and Nocedal (1980), with 5 correction vectors and two-loop recursion. Convergence is guaranteed regardless of the initialization. However, having a good initial guess $\mathbf{c}^0$ can significantly reduce the overall computational cost. To this aim, the library implements a data-driven initialization procedure, that extends to spatio-temporal settings the approach proposed by Ferraccioli et al. (2021), based on the spatial discretization of a heat diffusion process. This alternative shows significant advantages when the empirical distribution significantly deviates from the uniform distribution over D_T, achieving computational savings of up to 70% for highly non-uniform densities.

4.3. Selection of the smoothing parameters

In order to select appropriate values for the smoothing parameters λ_D and λ_T, we perform K–fold cross-validation based on the L^2–norm of the distance between the estimated density $\hat{f}$ and the true density f, that is $\int_{D_T}(\hat{f}(\mathbf{p},t) - f(\mathbf{p},t))^2\,d\mathbf{p}\,dt$. This leads to the following cross-validation error on the kth validation set, for $k = 1,\dots,K$,

$$\text{CVerr}_k(\lambda_D,\lambda_T) := \int_{D_T} \left(\hat{f}^{-[k]}_{\lambda_D,\lambda_T}(\mathbf{p},t)\right)^2 d\mathbf{p}\,dt - \frac{2}{n_k}\sum_{i=1}^{n_k} \hat{f}^{-[k]}_{\lambda_D,\lambda_T}(\mathbf{p}_i^{[k]}, t_i^{[k]}), \tag{10}$$

where: $\hat{f}^{-[k]}_{\lambda_D,\lambda_T}$ is the density estimate calculated using the training data (contained in all the K folds except the kth one), $(\mathbf{p}_i^{[k]}, t_i^{[k]})$ is the ith observation in the validation set (kth fold), and n_k is the cardinality of this last set. The quantity in (10) is obtained neglecting the term that does not depend on the pair (λ_D,λ_T) (see Ferraccioli et al., 2021 for more details), and can be efficiently calculated by exploiting the finite element formulation introduced in Section 4.1. At the beginning of the cross-validation procedure, data are randomly distributed in the folds, which are designed in such a way that they do not exclusively contain observations recorded at specific time instances.

The K resulting errors in (10) are then averaged to produce the final cross-validation error $\text{CVerr}(\lambda_D,\lambda_T)$. The procedure is repeated for each pair (λ_D,λ_T), exploring all the possible values provided in the input. Among those, the optimal proposals in space and time are selected by minimizing CVerr.

5. Confidence bands

In this section, we leverage a Bayesian analogy to derive pointwise confidence intervals for spatio-temporal densities, based on the proposed estimator. It should be noted that previous works on density estimation in space-only domains, such as those by Ferraccioli et al. (2021) and Arnone et al. (2022), did not develop tools for quantifying uncertainty. With the necessary simplifications, the inference approach introduced here can also be used to compute confidence intervals for densities over spatial domains, based on the estimators in Ferraccioli et al. (2021) and Arnone et al. (2022). The confidence bands we propose rely on the approach discussed in Leonard (1982) to address univariate problems related to a penalized likelihood functional, and subsequently considered by O'Sullivan (1988), who presented numerical results for univariate log-densities. These confidence bands significantly enhance the proposed method by providing valuable insights into the uncertainty associated with the estimated density.

Setting

$$\tilde{L}_0(\mathbf{c}) := -\frac{1}{2}\mathbf{c}^\top R\mathbf{c} \qquad \tilde{L}_1(\mathbf{c}) := \frac{1}{n}\mathbf{1}^\top Y\mathbf{c} - \sum_{\iota\subset T}\sum_{\tau\in\mathcal{T}_D} \kappa(\mathbf{w}_T \otimes \mathbf{w}_D)^\top \exp\{(\Phi_\iota \otimes \Psi_\tau)\,\mathbf{c}\},$$

we can rewrite the discretized functional in (8) as follows:

$$-\frac{\tilde{L}(\mathbf{c})}{n} = \tilde{L}_1(\mathbf{c}) + \tilde{L}_0(\mathbf{c}).$$

We can then reinterpret the problem of estimating $\mathbf{c}$ that minimizes functional (8) as the Bayesian problem of finding the posterior distribution of $\mathbf{c}$ under the log-likelihood $\tilde{L}_1(\mathbf{c})$, with the prior distribution $\mathbf{c} \sim \mathcal{N}(\mathbf{0}, R^{-1})$. Continuing with this Bayesian analogy, the posterior distribution for $\mathbf{c}$ is proportional to the following term:

$$\exp\left(\tilde{L}_1(\mathbf{c}) - \frac{1}{2}\mathbf{c}^\top R\mathbf{c}\right).$$

Moreover, the estimated coefficients $\hat{\mathbf{c}}$ is the posterior mode and provides an approximation of the distribution's mean. The posterior covariance can be approximated by the dispersion matrix D, that is,

$$D^{-1} = -\frac{\partial^2 log\ \pi(\hat{\mathbf{c}}|\ \{(\mathbf{p}_i, t_i)\}_{i=1}^n)}{\partial(\hat{\mathbf{c}}\,\hat{\mathbf{c}}^\top)} = -\frac{\partial^2 \tilde{L}_1(\hat{\mathbf{c}})}{\partial(\hat{\mathbf{c}}\,\hat{\mathbf{c}}^\top)} + R = \hat{H} + R,$$

where $\hat{H}$ is the Hessian of the log-likelihood, defined as

$$\hat{H} = \sum_{\iota \subset T} \sum_{\tau \in \mathcal{T}_D} \kappa\,(\boldsymbol{\Phi}_\iota \otimes \boldsymbol{\Psi}_\tau)^\top \left[(\mathbf{w}_T \otimes \mathbf{w}_D)^\top \exp\{(\boldsymbol{\Phi}_\iota \otimes \boldsymbol{\Psi}_\tau)\hat{\mathbf{c}}\}\right](\boldsymbol{\Phi}_\iota \otimes \boldsymbol{\Psi}_\tau).$$

This leads to a Gaussian approximation of the posterior distribution: $\mathbf{c}\,|\,\{(\mathbf{p}_i, t_i)\}_{i=1}^n \sim \mathcal{N}(\hat{\mathbf{c}}, D)$. Following this analogy, the approximated posterior distribution of the log-density function g is $\mathcal{N}((\Psi\otimes\Phi)^\top \hat{\mathbf{c}}, (\Psi\otimes\Phi)^\top (2D/n)\,(\Psi\otimes\Phi))$. Specifically, the pointwise approximated Bayesian posterior variance for the estimated log-density is given by:

$$\hat{\sigma}^2(\mathbf{p}, t) = \frac{2}{n}(\boldsymbol{\psi}(\mathbf{p}) \otimes \boldsymbol{\varphi}(t))^\top\, D\,(\boldsymbol{\psi}(\mathbf{p}) \otimes \boldsymbol{\varphi}(t)), \tag{11}$$

from which we can build pointwise $(1-\alpha)\%$ credible bands, that we might in turn reinterpret as confidence intervals.

It is worth noting that, as often discussed in nonparametric regression settings (see, e.g., Genovese and Wasserman, 2008; Hall and Horowitz, 2013, and related references), such approaches may lead to overly conservative confidence bands. Various strategies have been explored in the literature to address this issue. In our setting, we solve this problem by scaling the smoothing parameters λ_D and λ_T, which are used to compute $\hat{\sigma}^2(\mathbf{p}, t)$ in (11), by an appropriate scaling factor, determined through a standard bootstrapping procedure.

6. Simulation studies

In this section, we present some simulation studies that illustrate the performances of the proposed method. In Section 6.1, we introduce state-of-the-art competing methods. In Section 6.2, we define three common measures of error for spatio-temporal density functions, to quantify the methods' performance. In Section 6.3, we show a simulation over a simple planar domain. In Sections 6.4 and 6.5, we present simulation studies over a spherical domain. This very specific form of curved domain is considered here to enable comparisons with at least one competing method. Indeed, to the best of our knowledge, there currently is a lack of software implementing possible competing methods over general two-dimensional Riemannian manifolds. In Section 6.6 we outline some addition simulation studies reported in the appendix; in particular, Appendix H shows a simulation on a non-spherical two-dimensional manifold. It should be pointed out that some competing methods are designed for intensity estimation; for these methods, the estimates are appropriately normalized over $\mathcal{D}_T$. As detailed in Appendix B, STDE-PDE can also be used to estimate the intensity of inhomogeneous space–time Poisson processes. A simulation study in an intensity estimation setting is reported in Appendix F.

6.1. Competing methods

We consider the following state-of-the-art competitors:

INLA: the model for intensity estimation of spatio-temporal point processes developed in Chapter 8.3 of Krainski et al. (2018), and implemented in the R package `R-INLA` (www.r-inla.org). The model is based on the sPDE approach described in Lindgren et al. (2011), with the penalized complexity priors derived in Fuglstad et al. (2019) for the range and the marginal standard deviation.

LGCP: the model for intensity estimation of spatio-temporal log-Gaussian Cox processes, as presented by Brix and Diggle (2001) and Diggle et al. (2005), and implemented in the R package `LGCP` (Taylor et al., 2023). We here employ a bivariate Gaussian kernel in space and lowess smoother in time.

STKDE: the model for fixed-bandwidth kernel density estimation from spatio-temporal point patterns, as presented in Fernando and Hazelton (2014) and Davies et al. (2018), and implemented in the R package `sparr` (Davies and Marshall, 2023). For the simulations over curved surfaces, we rely on the STKDE version for data aggregated over time, described by Zhang et al. (2011) and implemented in the R package `stam` (Zhang, 2010).

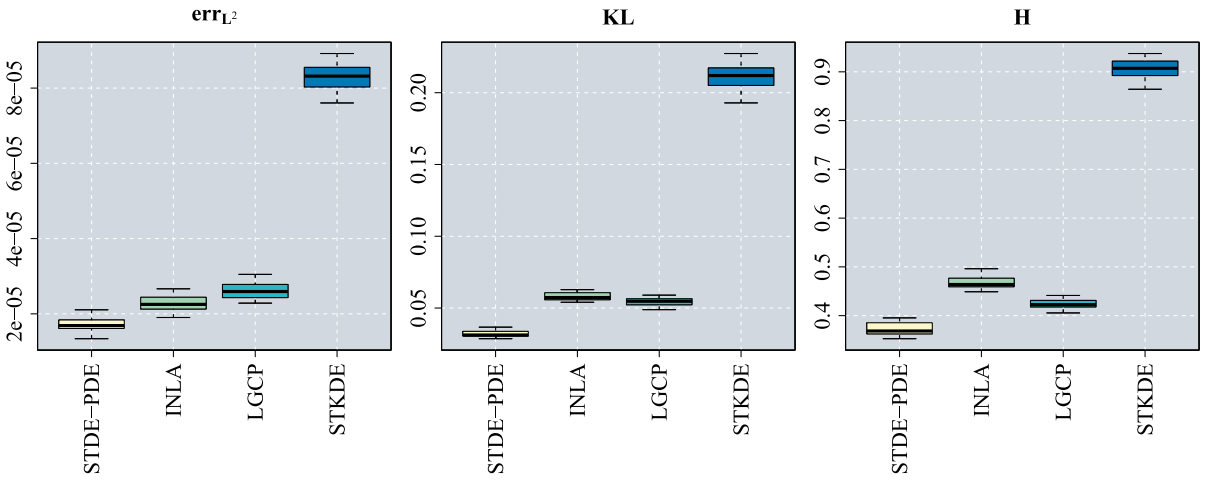


Fig. 4. Simulation 1. Boxplots of L^2–norm of the error (left), Kullback–Leibler divergence (center), and Hellinger distance (right) of the estimates provided by proposed STDE-PDE and the competing methods introduced in Section 6.1, namely INLA, LGCP, and STKDE, over 30 simulation repetitions.

6.2. Measures of error

To compare the various competing methods, we consider the following measures of error:

L^2–norm of the error (err_{L^2}):

$$\text{err}_{L^2}(\hat{f}, f) := \|\hat{f} - f\|^2_{L^2(0,T;L^2(\mathcal{D}))} = \int_{D_T} \left(\hat{f}(\mathbf{p}, t) - f(\mathbf{p}, t)\right)^2 d\mathbf{p}\, dt,$$

where the integral over D_T is numerically approximated by summing over a fine evaluation grid in $[0, T]$ and using a regular lattice that covers D.

Kullback–Leibler divergence (KL):

$$\text{KL}(\hat{f}, f) := \int_{D_T} \hat{f}(\mathbf{p}, t) \log\left(\frac{\hat{f}(\mathbf{p}, t)}{f(\mathbf{p}, t)}\right) d\mathbf{p}\, dt,$$

computed resorting to the functionalities provided by the R package `philentropy` (Drost and Nowosad, 2024).

Hellinger distance (H):

$$\text{H}(\hat{f}, f) := \left(\frac{1}{2} \int_{D_T} \left(\sqrt{\hat{f}(\mathbf{p}, t)} - \sqrt{f(\mathbf{p}, t)}\right)^2 d\mathbf{p}\, dt\right)^{1/2}.$$

6.3. Simulation 1: a mixture of time-dependent Gaussian distributions on a squared region

In Simulation 1, we consider point data scattered over a squared spatial domain and recorded over the time interval from 0 to 1. The true density is defined as a mixture of 4 time-dependent Gaussian distributions, whose means and variances evolve over time as reported in Appendix E. The second row in Fig. 5 displays the true density at some time instances, showing its multimodal pattern, with modes having different directions and intensities of anisotropy. We generate $n = 5000$ independent observations from this density, adopting the acceptance-rejection technique (Casella et al., 2004; Neal, 2003). The sampling process is repeated 30 times. The first row of Fig. 5 shows the data obtained in the first simulation repetition.

In this simulation, we consider all the competing methods introduced in Section 6.1. For discretization over space for the proposed STDE-PDE and for INLA, we employ regular spatial meshes of the squared domain, having respectively 793 nodes for STDE-PDE and 789 nodes for INLA, generated using the respective software. For LGCP and STKDE we use a regular 32×32 spatial grid. For discretization in time, we use 7 evenly spaced knots for STDE-PDE and INLA, and 9 evenly spaced time instances for LGCP and STKDE, from 0 to 1, extremes excluded. Smoothing parameters for STDE-PDE are selected via 10–fold cross-validation, as presented in Section 4.3.

Fig. 5 qualitatively compares the mean estimates obtained by the various methods over the 30 simulation repetitions. The figure highlights that STDE-PDE accurately captures the spatial structure of the true density and its variation in time. Fig. 4 gives the boxplots of the measures of error introduced in Section 6.2, over the 30 simulation repetitions, computed on a fine grid of 1024 locations in space and 9 instances in time. The proposed method outperforms all others according to all measures of error.

Fig. 6 presents inference results, comparing STDE-PDE with INLA, the only competitor equipped with built-in inference capabilities. We compute the coverage probabilities of the 95% confidence intervals for the estimates over the 30 simulation

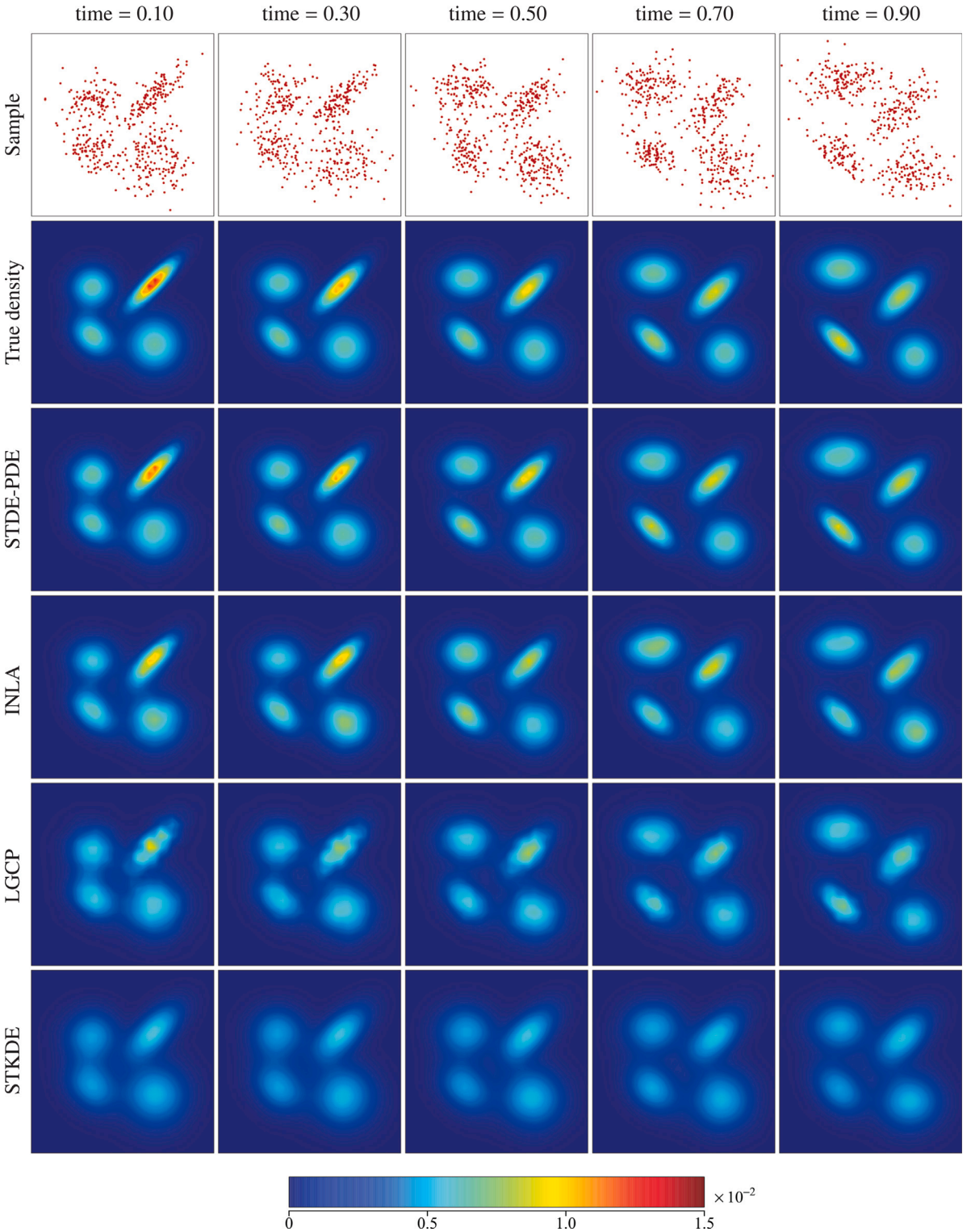


Fig. 5. Simulation 1. View of the sample (first row), true density (second row), and mean estimates of STDE-PDE (third row), INLA (fourth row), LGCP (fifth row), and STKDE (sixth row), over 30 simulation repetitions, at different time instances. All the plots have the same color scale, with bluish colors corresponding to low-density values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

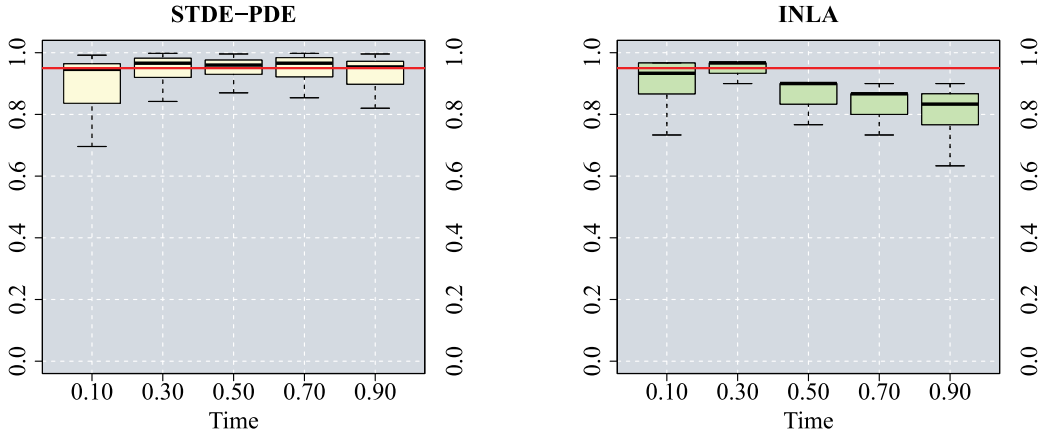


Fig. 6. Simulation 1. Boxplots of empirical coverage probabilities of 95% confidence intervals over 30 simulation repetitions: STDE-PDE (left) and INLA (right).

repetitions. For STDE-PDE we follow the procedure outlined in Section 5. We evaluate the results at 5 different time instances and on a regular 32×32 spatial grid, focusing on areas characterized by non-negligible density values, specifically where the true density exceeds 10^{-3}. We obtain an overall median empirical coverage of 0.962 (with an IQR of 0.048) for STDE-PDE, and of 0.881 (with an IQR of 0.094) for INLA. Fig. 6 illustrates the empirical coverage for the different time instances. For the proposed method, the median empirical coverage is close the nominal value, 0.95, for any time considered instant. Furthermore, the interquartile range of the empirical coverage remains narrow and more stable across time, with respect to the competing method, avoiding both under- and over-conservative behaviors.

6.4. Simulation 2: a mixture of time-dependent Kent distributions on a spherical domain

In Simulation 2 we examine a space–time point pattern over the surface of the unit sphere recorded from 0 to 1 in time. We generate 10 000 independent observations from a mixture of 5 time-dependent Kent distributions (Kent, 1982; Mardia and Jupp, 2000), with parameters defined in Appendix E. Correct sampling on a sphere, and more generally on a two dimensional manifold, is not trivial; we refer the reader to Arnone et al. (2022) for details on how this can be appropriately performed.

We compare the proposed method to STKDE for data aggregated over time (Zhang et al., 2011), considering 9 evenly spaced time instances from 0 to 1, extremes excluded. We could not find available implementations of other techniques acting on non-planar spatial domains. For the proposed STDE-PDE we use a regular spatial mesh with 606 nodes over the surface of the unit sphere, and 9 evenly spaced knots in time from 0 to 1, extremes excluded; the smoothing parameters are selected by 10–fold cross-validation.

Fig. 7 qualitatively compares the mean estimates obtained by the various methods over the 30 simulation repetitions. The figure highlights the remarkably good fit returned by STDE-PDE, which can effectively capture each component of the mixture distribution, as it shifts around the sphere. The good performances of the proposed method are confirmed by Fig. 8, that gives the boxplots, over the 30 simulation repetitions, of the measures of error introduced in Section 6.2, computed on a grid of 5016 locations in space and 9 instances in time.

6.5. Simulation 3: a highly localized and skewed density on a spherical domain

In Simulation 3, we consider another space–time point pattern over the unit sphere, that is highly localized along elongated regions, mimicking the behavior of earthquake data, that cluster around Earth's fault lines. We sample 10 000 observations from a mixture of 25 time-dependent Kent distributions, whose parameters are provided in Appendix E; the resulting density is highly anisotropic, localized, and skewed along elongated ridges. Mean estimates and measures of error, over 30 simulation repetitions, are provided in Figs. 9 and 10. STKDE appears to oversmooth the signal, while the proposed STDE-PDE proves to be particularly well suited to this complex data setting, capturing the strongly localized and skewed modes of the signal.

6.6. Additional simulation studies

In addition to Simulations 1, 2, and 3, presented in Sections 6.3 6.4 and 6.5, we have conducted several other tests, reported in the appendix. Appendix F presents intensity estimation of an inhomogeneous Poisson process, considering the same point pattern in Simulation 1. We here compare the proposed method to INLA and LGCP, obtaining the same ordering among the techniques obtained for Simulation 1. Appendix G shows an additional simulation on a planar domain, considering a true density that combines high-frequency and low-frequency components, that significantly overlap over time. Appendix H considers a space–time point pattern on a non-spherical two-dimensional manifold. We here compare the proposed STDE-PDE with a separable method for density estimation based on k–nearest neighbors, which we implemented solely for comparison purposes, since none of the main available competing methods is currently implemented to handle such data. STDE-PDE outperforms such competing method, that relies on the first-order separability assumption, even though the data are generated from a density that satisfies such assumption.

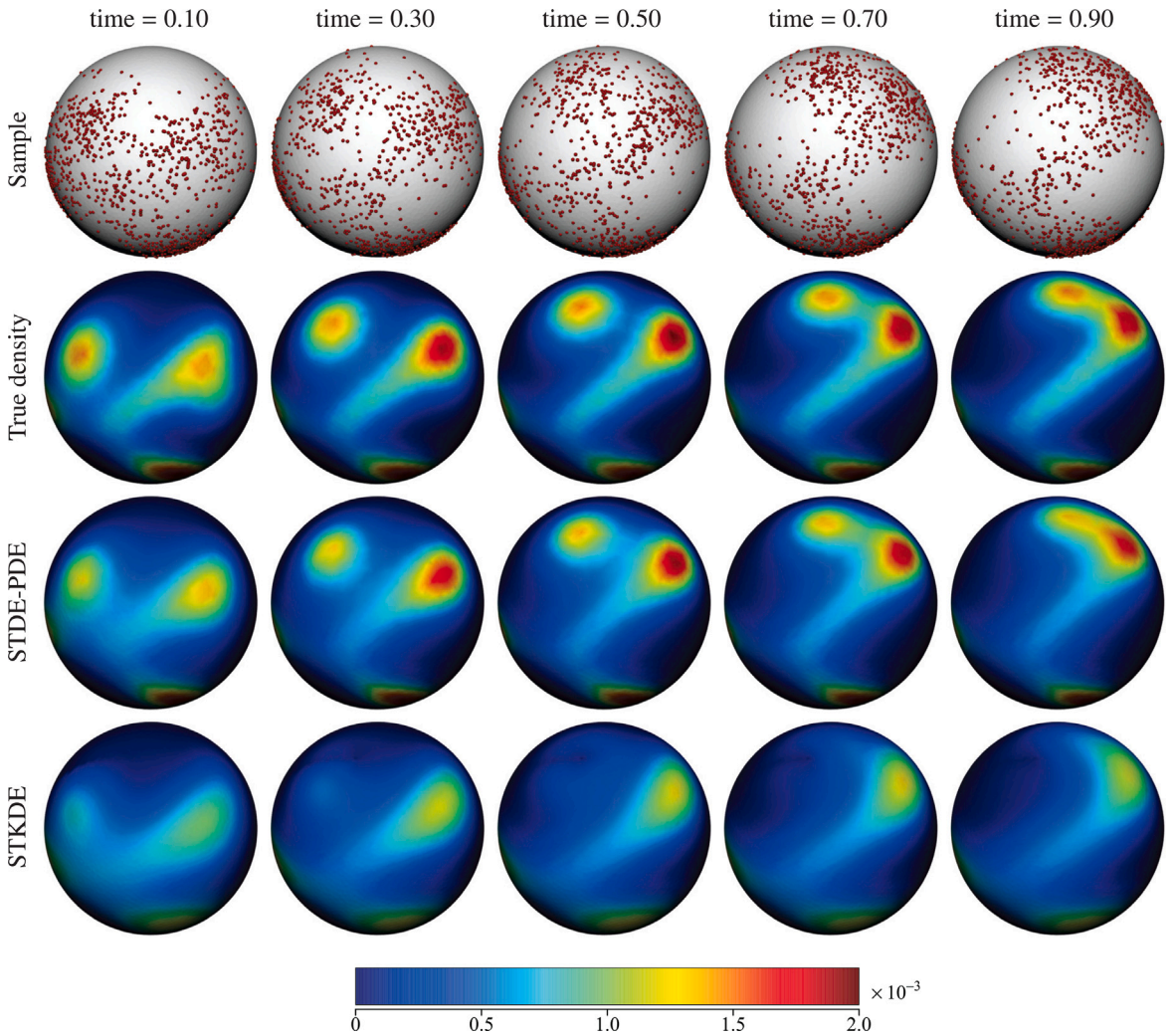


Fig. 7. Simulation 2. Front view of the sample (first row), true density (second row), and mean estimates of STDE-PDE (third row), STKDE (fourth row), over 30 simulation repetitions, at different time instances. All the plots have the same color scale, with bluish colors corresponding to low density values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

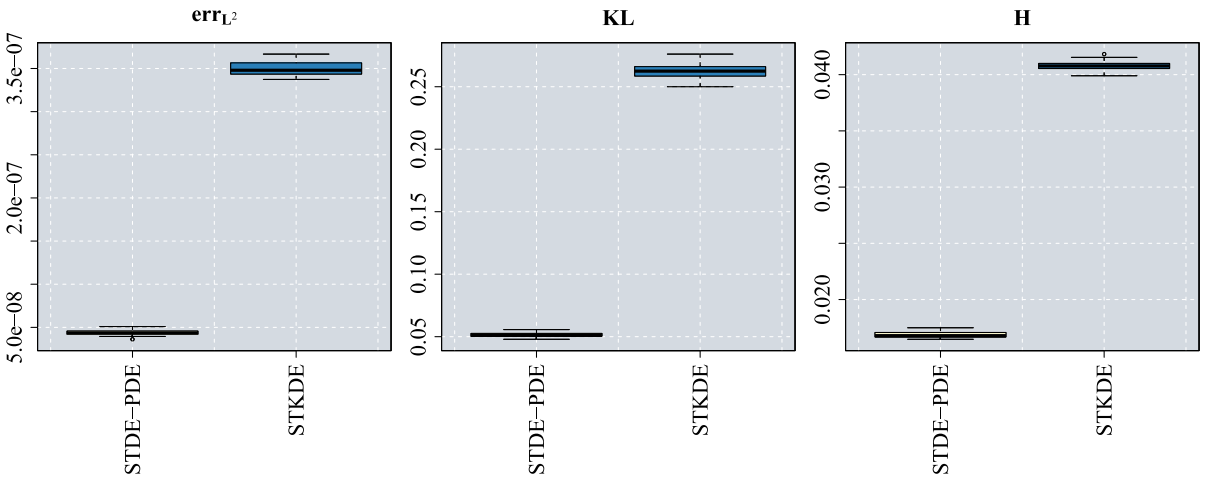


Fig. 8. Simulation 2. Boxplots of L^2–norm of the error (left), Kullback–Leibler divergence (center) and Hellinger distance (right) of the estimates provided by the STDE-PDE and the STKDE methods, over 30 simulation repetitions.

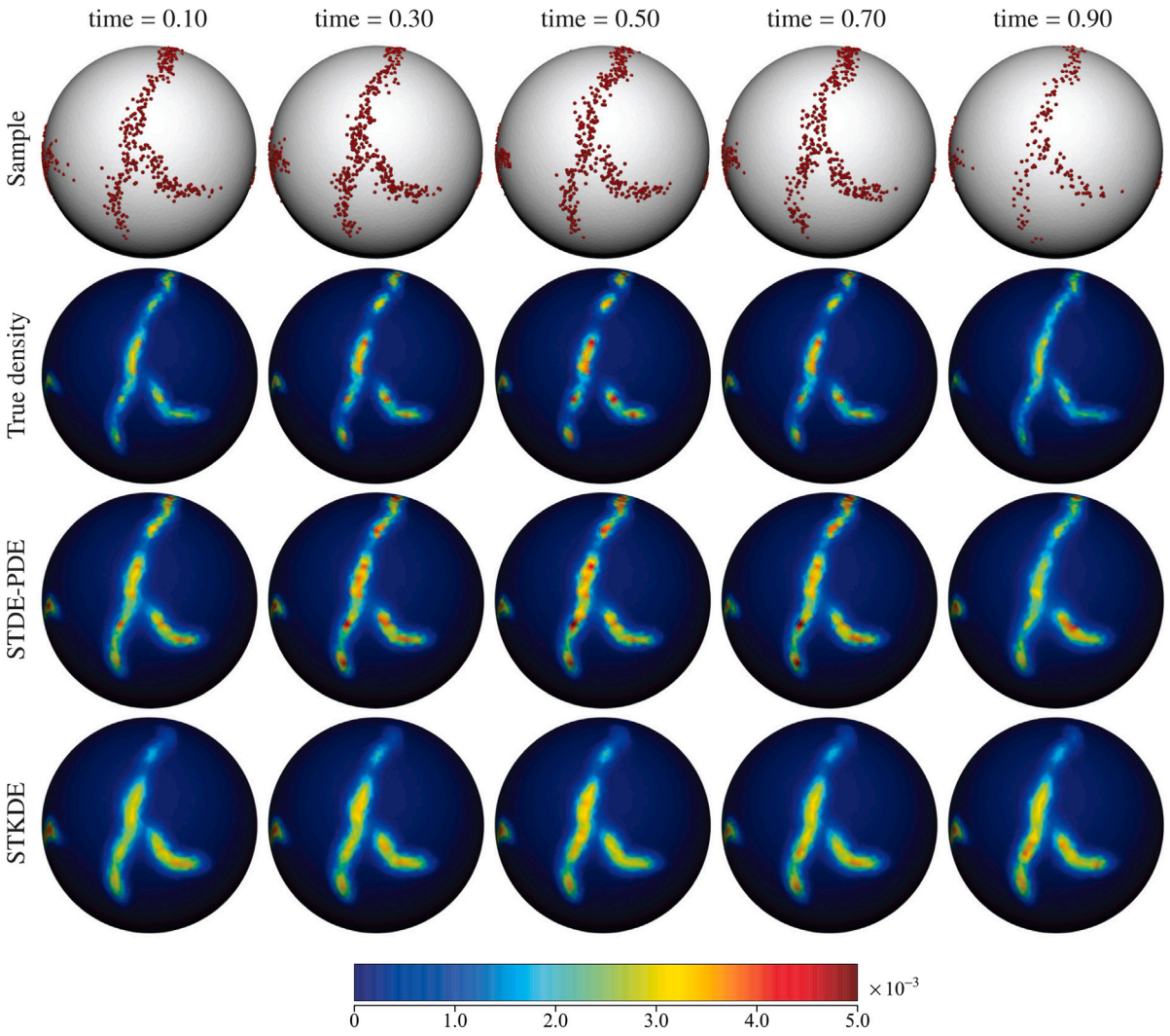


Fig. 9. Simulation 3. Front view of the sample (first row), true density (second row), and mean estimates of STDE-PDE (third row), STKDE (fourth row), over 30 simulation repetitions, at different time instances. All the plots have the same color scale, with bluish colors corresponding to low density values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

7. Applications to epidemiological and seismological studies

In this section, we show two illustrative real data examples, concerning gastrointestinal infections (Section 7.1) and earthquakes with large magnitude (Section 7.2).

7.1. Gastrointestinal infections

We analyze cases of gastrointestinal symptoms reported to the National Health Service (NHS) Direct in Hampshire county, United Kingdom, during the period 2001 to 2003. The dataset, provided by Taylor et al. (2013), comprises 10 069 observations. Some temporal snapshots of the observed data are illustrated in Fig. 1

In this application, we examine all the competing methods introduced in Section 6.1 and already considered in the simulation study presented in Section 6.3. We maintain the settings and options adopted for competing methods in Simulation 1. For estimation over the Southampton region of interest, we employ the regular spatial mesh with 778 nodes displayed in the left panel of Fig. 3 for STDE-PDE, a mesh with 776 nodes produced with `R-INLA` for INLA, and a regular 32×32 grid for LGCP and STKDE. For estimation in time, we use 7 knots. Smoothing parameters for STDE-PDE are selected via 10–fold cross-validation.

Fig. 11 shows qualitative results. The proposed STDE-PDE returns an estimate that captures the strongly localized peaks of new cases; the estimate highlights different behaviors of the phenomenon along the western and eastern shores of Southampton Water

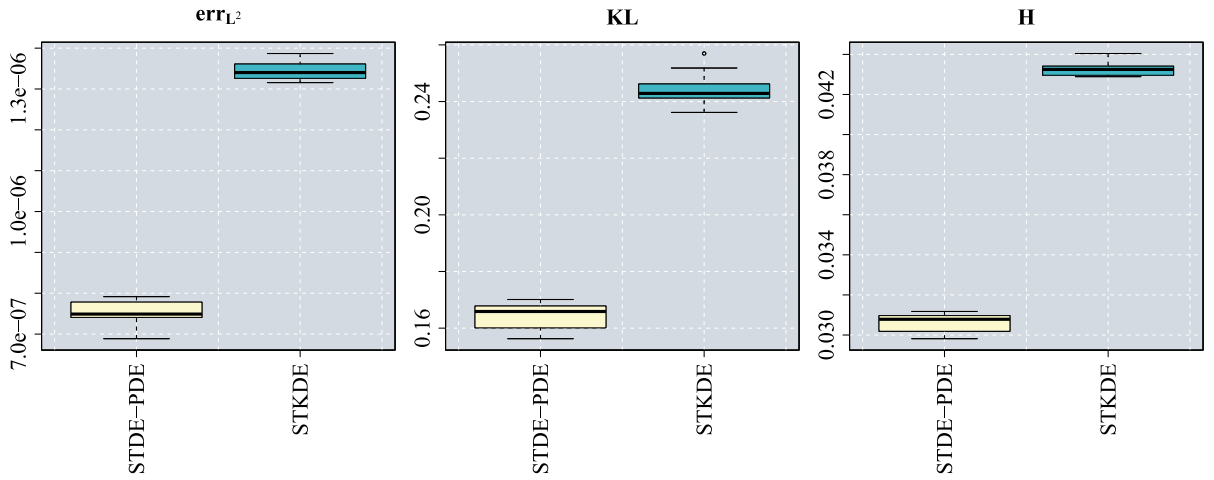


Fig. 10. Simulation 3. Boxplots of L^2–norm of the error (left), Kullback–Leibler divergence (center), and Hellinger distance (right) of the estimates provided by the STDE-PDE and STKDE methods, over 30 simulation repetitions.

estuary. In Fig. 12 (left panel) we also provide quantitative comparisons among the estimates provided by the different methods, in terms of the 10–fold cross-validation error in (10), highlighting the comparative advantage of the proposed method.

7.2. Earthquakes of large magnitude

We now analyze the locations of earthquakes with magnitude greater than 4.75 worldwide from 2020 to 2022. The dataset contains 8080 observations and is taken from the USGS earthquakes catalog (https://earthquake.usgs.gov/earthquakes/search/). In this setting, we compute the estimates with STDE-PDE over a regular discretization in time, with 7 knots, and the adaptive data-driven spatial mesh shown in the right panel of Fig. 3, with 2000 nodes. As indicated in Section 4.1, this mesh naturally permits a more accurate estimation along the faults, at a lower computational cost with respect to a uniformly fine mesh. For the considered competing methods, we follow the same choices detailed for the simulation study presented in Section 6.4. For STKDE, we aggregate locations on a three-monthly basis.

Fig. 2 shows some temporal snapshots of the observed data and Fig. 13 the estimate at the corresponding times. The density estimate produced by STDE-PDE correctly displays elongated ridges of high values in correspondence of the faults in the Earth's crust. Fig. 12 (right panel) displays the boxplots of 10–fold cross-validation errors, offering quantitative confirmation of the accuracy of the estimates obtained through the proposed method.

8. Discussion and future research directions

The proposed STDE-PDE offers an interesting addition to the literature on space–time density and intensity estimation. Simulation studies and real-world applications highlight the robust performance of STDE-PDE in comparison to state-of-the-art methodologies, showing a strong ability to capture complex multi-modal spatio-temporal densities. The proposed method is shown to accurately work on two-dimensional spatial domains with non-trivial shapes, including general two-dimensional Riemannian manifolds. In future studies, we will show the ability of the method to model time-varying point patterns in complicated three-dimensional spatial domains, leveraging the techniques recently developed by Arnone et al. (2023b) for data observed in non-convex volumes.

STDE-PDE can be extended in various directions. One interesting research line concerns replicated spatio-temporal point process, addressed by Gervini (2022) by means of log-linear models for latent intensity functions. Another fascinating direction is related to the modeling of space-varying temporal densities, as those considered by Menafoglio et al. (2014, 2016) and Scimone et al. (2022). These densities differ from the spatio-temporal densities considered in this work, as they maintain unitary integration at each spatial location. Complementary research could instead focus on time-varying spatial density estimation problems.

Of particular significance is the expansion of the proposed method into semiparametric approaches, particularly in scenarios where supplementary information is available, either through covariates or specific insights into the nature of the phenomenon being studied. For instance, the considered point process could be endowed with pairwise interactions, as in Gibbs processes. This situation could be accommodated by penalized composite-likelihood approaches, such as those considered by Illian and Hendrichsen (2010), Guan et al. (2015) and Daniel et al. (2018). Additionally, in the proposed modeling framework, knowledge about the physics governing the phenomenon under investigation could also be included through physics-informed penalties, such as those discussed in Arnone et al. (2019) in space–time regression settings.

An interesting theoretical investigation goes towards the exploration of a potential conceptual analogy between the nonparametric penalized likelihood approach, followed in this work, and stochastic methods such as classical or sPDE-based log Gaussian

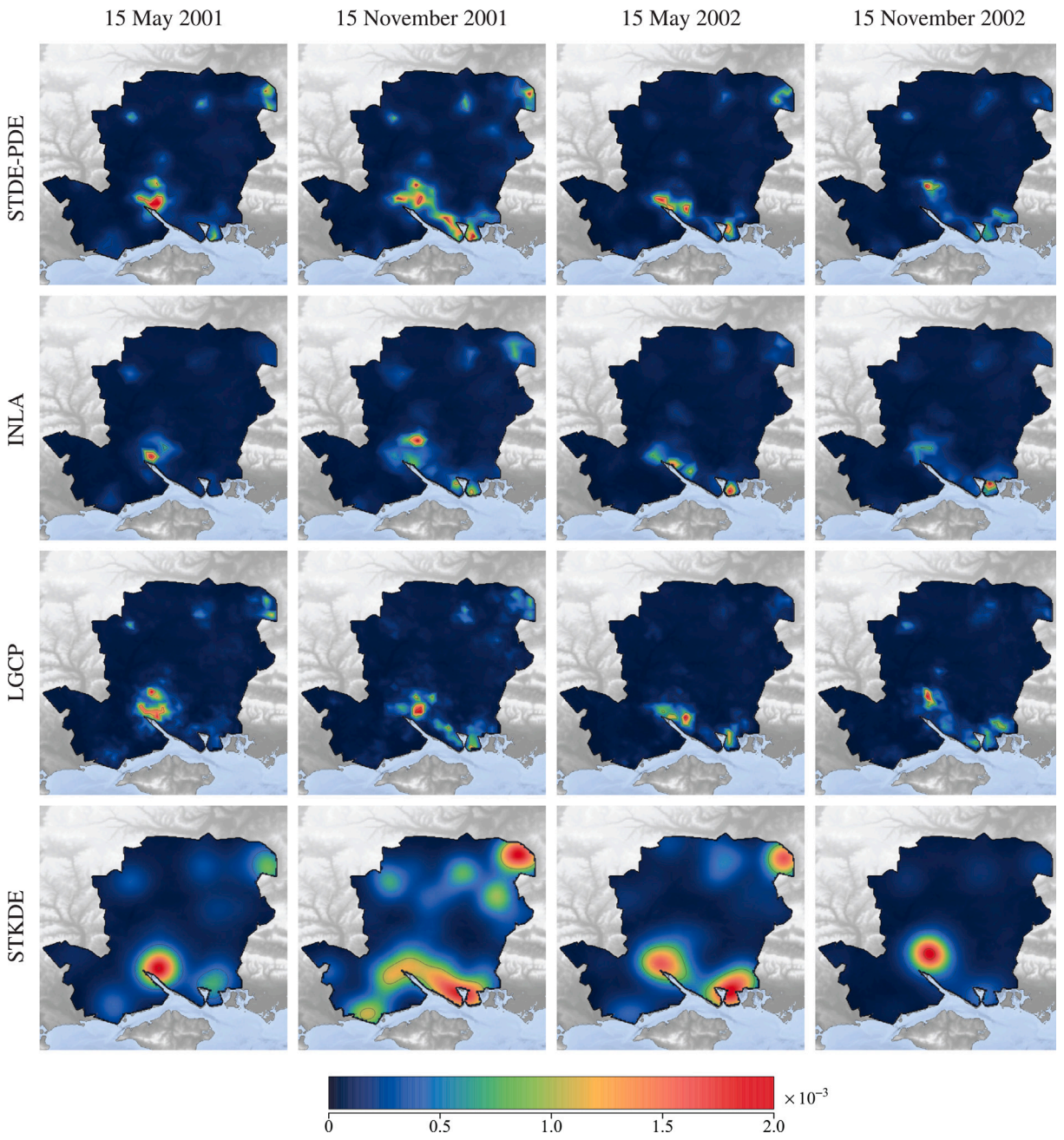


Fig. 11. Application to gastrointestinal infections. View of the density estimates for STDE-PDE (first row), INLA (second row), LGCP (third row), and STKDE (fourth row), at different time instances. All the plots have the same color scale, with bluish colors corresponding to low density values of gastrointestinal infections in Hampshire county, UK. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Cox processes. In the geostatistical setting, the connection between penalized spline smoothing and kriging has for instance been investigated by Kimeldorf and Wahba (1970), Furrer and Nychka (2007) and Laslett (1994). To the best of our knowledge, instead, in the point process framework, such potential analogy has not yet been investigated in the literature, not even in the simpler one-dimensional and two-dimensional settings considered by Silverman (1982) and Gu and Qiu (1993).

Finally, a challenging development goes towards generalizations of STDE-PDE to model marked point processes. In the application to earthquake data presented in Section 7.2, this extension would enable the inclusion of the magnitude in the analysis. In addition, modeling marked point processes could for instance allow the monitoring of spatio-temporal extreme value trends, in relation to meteorological and socio-economic factors, as done in the study on air pollution in Spain conducted by Wang et al. (2023).

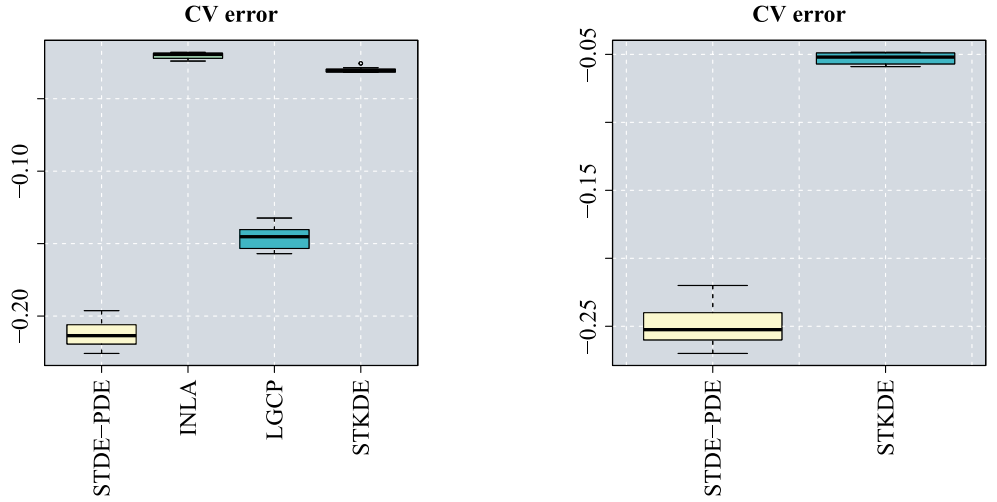


Fig. 12. Left: Application to gastrointestinal infections; boxplots of 10–fold cross-validation errors for STDE-PDE and the competing methods introduced in Section 6.1, namely INLA, LGCP, and STKDE. Right: Application to earthquakes of large magnitude; boxplots of 10–fold cross-validation errors for STDE-PDE and STKDE.

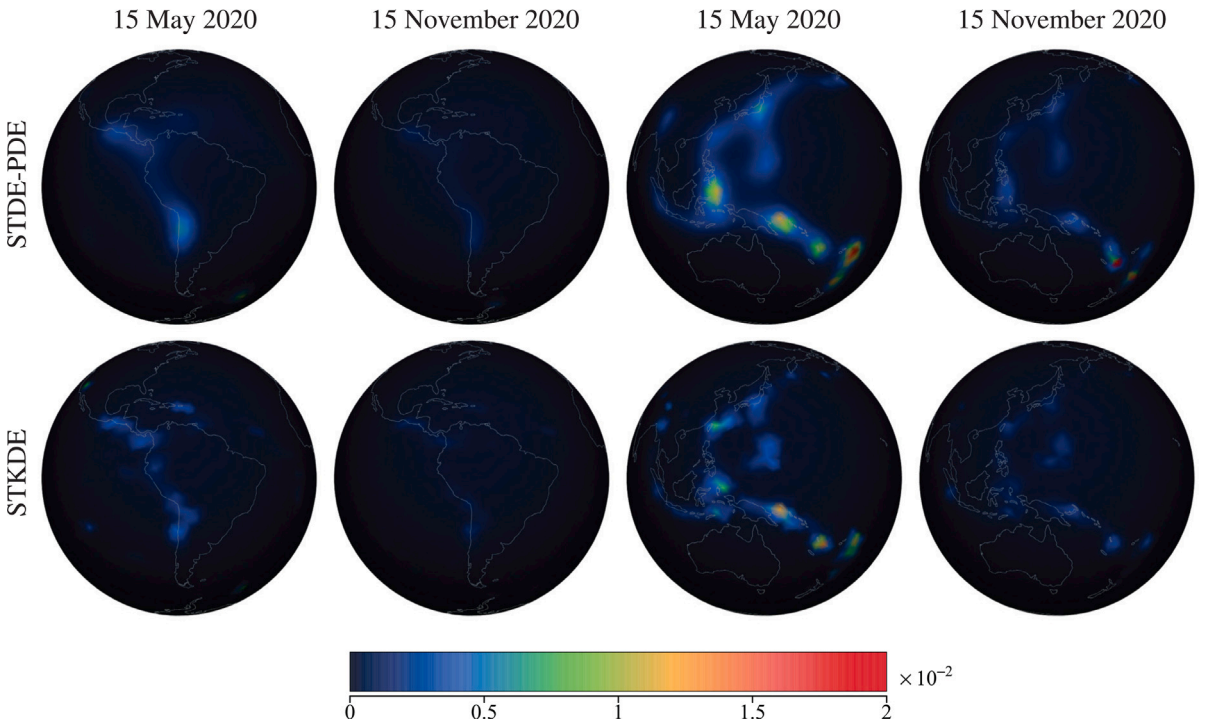


Fig. 13. Application to earthquakes of large magnitude. View of the density estimates over the Australian, Eurasian and Pacific plates, for STDE-PDE (first row), STKDE (second row), at different time instances. All the plots have the same color scale, with red colors corresponding to high density of earthquakes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Acknowledgments

We are grateful to the Editor and Reviewers for their constructive feedback, which has significantly enhanced the quality of the manuscript. We would like to thank Aldo Clemente for discussions concerning the proof of Lemma 2 in Appendix C. B. Begu was funded by NexSys SFI/21/SPP/3756, Ireland. L.M. Sangalli acknowledges funding by the European Union NextGenerationEU programme, research project CoEnv — Complex Environmental Data and Modeling (PRIN2022 - CUP 2022E3RY23). S. Panzeri and L.M. Sangalli acknowledge the research project Dipartimento di Eccellenza 2023–2027, Dipartimento di Matematica, Politecnico di Milano, Italy, funded by the Ministero dell'Università e della Ricerca.

Appendix A. Derivation of the unconstrained minimization problem

In this section, we demonstrate that the second term in the functional $L(g)$ in (1) ensures that $\int_{\mathcal{D}_T} e^{\hat{g}(\mathbf{p},t)}\,d\mathbf{p}\,dt = 1$, and thus that $\hat{f} = e^{\hat{g}}$ is a proper density over $\mathcal{D}_T$. Indeed, in order to estimate the log-density g, by using a penalized likelihood approach, one would minimize the objective functional

$$L_c(g) := -\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{p}_i, t_i) + R(g;\lambda), \tag{A.1}$$

under the constraint that $\int_{\mathcal{D}_T} e^{\hat{g}(\mathbf{p},t)}\,d\mathbf{p}\,dt = 1$. Equivalently, one would minimize the functional $L_c(g)$ in (A.1) over the space $V_c := \left\{ g \in V : \int_{\mathcal{D}_T} e^{g(\mathbf{p},t)}\,d\mathbf{p}\,dt = 1 \right\}$. Note that V_c is the restriction of the space V, defined in Eq. (4) in Section 3.1, to functions such that $\int_{\mathcal{D}_T} e^{\hat{g}(\mathbf{p},t)}\,d\mathbf{p}\,dt = 1$. In other words, V_c is the restriction of the space V to log-densities. The following theorem states the equivalence of the constrained minimization problem, introduced above, with the unconstrained minimization introduced in Section 2.

Theorem 3. *The minimization of $L_c(g)$ in (A.1) in the constrained space V_c is equivalent to the minimization of $L(g)$ in (1) in the unconstrained space V.*

Proof. We start by observing that when $g \in V_c$, $L(g) = L_c(g) + 1$; thus the functional L and L_c differ by a constant on V_c. This entails that in the space V_c the minimization of $L(g)$ is equivalent to the minimization of $L_c(g)$. We now show that for each $g \in V$ there exist a $g^* \in V_c$ such that $L(g^*) \le L(g)$. Given $g \in V$, define $g^* = g - \log \int_{\mathcal{D}_T} e^{g(\mathbf{p},t)}\,d\mathbf{p}\,dt$. By construction, $g^* \in V$ and e^{g^*} satisfies the unitary integration constraint, therefore $g^* \in V_c$. Moreover, since g and g^* differ only by a constant, the differential operators in space and time applied to g and g^* coincide, implying that $R(g;\lambda) = R(g^*;\lambda)$. Let $h := \int_{\mathcal{D}_T} e^{g(\mathbf{p},t)}\,d\mathbf{p}\,dt$, we derive the following:

$$\begin{aligned}
L(g^*) &= -\frac{1}{n}\sum_{i=1}^{n} g^*(\mathbf{p}_i, t_i) + 1 + R(g^*;\lambda) \\
&= -\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{p}_i, t_i) + \log \int_{\mathcal{D}_T} e^{g(\mathbf{p},t)}\,d\mathbf{p}\,dt + 1 + R(g;\lambda) \\
&= L(g) - \int_{\mathcal{D}_T} e^{g} + \log \int_{\mathcal{D}_T} e^{g(\mathbf{p},t)}\,d\mathbf{p}\,dt + 1 \\
&= L(g) - h + \log h + 1.
\end{aligned}$$

Since the inequality $h - \log h \ge 1$ holds for all $h > 0$, then $L(g^*) \le L(g)$ with equality only if $h = 1$. This implies that $\hat{g}$ minimizes $L(g)$ if and only if $\int_{\mathcal{D}_T} e^{\hat{g}(\mathbf{p},t)}\,d\mathbf{p}\,dt = 1$, hence if and only if $\hat{g}$ minimizes $L_c(g)$ in V_c. □

Remark 3. The proof of the theorem extends the arguments proposed by Silverman (1982) to the case of interest. The result critically depends on the choice of the penalizing term $R(g;\lambda)$, which must be expressible as a bilinear form with differential operators depending solely on partial derivatives. In our case, after defining the bilinear form as

$$b(g_1, g_2) := \int_{\mathcal{D}_T} \Delta_{\mathcal{D}}(g_1)\Delta_{\mathcal{D}}(g_2)\,d\mathbf{p}\,dt + c\int_{\mathcal{D}_T} \frac{\partial^2 g_1}{\partial t^2}\frac{\partial^2 g_2}{\partial t^2}\,d\mathbf{p}\,dt, \qquad \text{where } c := \frac{\lambda_T}{\lambda_D}, \tag{A.2}$$

then, the regularization term can be written as $R(g;\lambda) = \lambda_D\, b(g,g)$.

Appendix B. Intensity estimation of an inhomogeneous Poisson process

In this section, we briefly discuss the connection between the spatio-temporal density estimation problem, introduced in Section 2, and the problem of intensity estimation of an inhomogeneous space–time Poisson point process. Firstly, we derive the log-likelihood of such a point process with inhomogeneous intensity function $\gamma : \mathcal{D}_T \to \mathbb{R}^+$. According to González et al. (2016), the underlying distribution can be factorized into two terms. The first term refers to the Poisson univariate random variable N, with mean $\Gamma(\mathcal{D}_T) := \int_{\mathcal{D}_T} \gamma(\mathbf{p},t)\,d\mathbf{p}\,dt$, that characterizes the number of events observed over $\mathcal{D}_T$. The second term regards the set of mutually independent and identically distributed spatio-temporal observations drawn from the random variable X, whose occurrence density is $p_X(\mathbf{p}_i, t_i) := \gamma(\mathbf{p}_i, t_i)/\Gamma(\mathcal{D}_T)$. Following Daley and Vere-Jones (2008) and Diggle (2013), the likelihood can be defined as the probability of getting n points $\{(\mathbf{p}_i, t_i)\}_{i=1}^n$ in $\mathcal{D}_T$, multiplied by the joint conditional density for these observations, that is,

$$p_N(n)p_{X|N}\left((\mathbf{p}_1,t_1),\dots,(\mathbf{p}_n,t_n)|n\right) = \underbrace{\frac{\Gamma(\mathcal{D}_T)^n}{n!}e^{-\Gamma(\mathcal{D}_T)}}_{p_N(n)}\, n! \prod_{i=1}^{n}\underbrace{\frac{\gamma(\mathbf{p}_i,t_i)}{\Gamma(\mathcal{D}_T)}}_{p_X(\mathbf{p}_i,t_i)} = e^{-\Gamma(\mathcal{D}_T)}\prod_{i=1}^{n}\gamma(\mathbf{p}_i,t_i), \tag{B.1}$$

where $p_{X|N}\left((\mathbf{p}_1,t_1),\dots,(\mathbf{p}_n,t_n)|n\right)$ is obtained similarly to Streit and Streit (2010) and Daley and Vere-Jones (2003) in the simpler spatial framework. Hence, starting from the likelihood in (B.1) and setting $\phi = \log\gamma$, we end up with the penalized negative

log-likelihood:

$$P(\phi) := -\sum_{i=1}^{n} \phi(\mathbf{p}_i, t_i) + \int_{D_T} e^{\phi(\mathbf{p},t)} + R(\phi; \tilde{\lambda}), \tag{B.2}$$

The following proposition states the equivalence between the density and the intensity estimation problem.

Proposition 1. *The estimate of the intensity $\hat{\gamma} = e^{\hat{\phi}}$ can be obtained by minimizing $P(\phi)$ in V or equivalently as $\hat{\gamma} = n\hat{f}$, where $\hat{f} = e^{\hat{g}}$ is the density estimated by minimizing the objective functional $L(g)$ in V, with $\lambda = \tilde{\lambda}/n$ and $L(g)$ defined as in (1) in Section 2.*

Proof. Let us assume that the intensity function takes the form $\gamma = Cf$, where C is a positive real-valued constant, and $f : \mathcal{D}_T \to \mathbb{R}^+$ is a density function over $\mathcal{D}_T$, for which $\int_{D_T} f(\mathbf{p}, t)\, d\mathbf{p}\, dt = 1$. Hence, $\log(\gamma) = \log(C) + \log(f) = \log(C) + g$, and the functional in (B.2) can be expressed in terms of C and g as follows:

$$P(C, g) = -n \log C - \sum_{i=1}^{n} g(\mathbf{p}_i, t_i) + C + R(g, \tilde{\lambda}).$$

We recall that the regularizing term $R(\cdot\,; \tilde{\lambda})$ solely depends on derivatives of the argument, implying that it annihilates constant terms. Since C is a constant, the minimization of (B.2) can be carried out in two steps. Firstly, by differentiating $P(C, g)$ with respect to C and solving the resulting equation for which the derivative equals zero, we get $\hat{C} = n$. Following this initial step, neglecting constant terms and diving all by n, the functional simplifies to

$$P(n, g) = -\frac{1}{n} \sum_{i=1}^{n} g(\mathbf{p}_i, t_i) + R\left(g; \frac{\lambda}{n}\right). \tag{B.3}$$

We are only left to differentiate $P(n, g)$ with respect to g, but this is exactly the density estimation problem we tackled in Appendix A. In accordance with the reasoning presented for the density estimation procedure, we can address the unitary integration constraint of f over $\mathcal{D}_T$ by incorporating the integral term directly into the functional in (B.3). Then, setting $\lambda = \lambda/n$ leads to the same functional in (1). In light of these considerations, a spatio-temporal intensity estimate can be obtained either by directly minimizing the functional $P(\phi)$ in (B.2), or equivalently, by multiplying the density estimate, attained by minimizing $L(g)$, by the number of observations n. □

Appendix C. Proof of Theorem 1

In this section, we address the uniqueness and the existence of a minimizer for the functional $L(g)$ defined in (1) in the functional space V in (4). The proof of Theorem 1 is assured by Theorem 4.1 from Gu and Qiu (1993), which was formulated within a general framework, provided that all the necessary hypothesis for its validity are met. The following two lemmas ensure these hypothesis.

Lemma 1. *The functional*

$$J(g) := -\frac{1}{n} \sum_{i=1}^{n} g(\mathbf{p}_i, t_i) + \int_{D_T} e^{g(\mathbf{p},t)}\, d\mathbf{p}\, dt$$

is continuous and strictly convex in V. Consequently, if a minimizer for $L(g)$ exists, it is unique.

Proof. J is continuous, being the sum of a linear term and a combination of continuous operators. Let now $g_1, g_2 \in V$, $\alpha \in [0, 1]$, and $g = \alpha g_1 + (1 - \alpha) g_2$. J is convex. Indeed:

$$\begin{aligned} J(g) &= \alpha \left\{ -\frac{1}{n} \sum_{i=1}^{n} g_1(\mathbf{p}_i, t_i) \right\} + (1-\alpha) \left\{ -\frac{1}{n} \sum_{i=1}^{n} g_2(\mathbf{p}_i, t_i) \right\} + \int_{D_T} e^{\alpha g_1(\mathbf{p},t)} e^{(1-\alpha) g_2(\mathbf{p},t)}\, d\mathbf{p}\, dt \\ &\leq \alpha \left\{ -\frac{1}{n} \sum_{i=1}^{n} g_1(\mathbf{p}_i, t_i) \right\} + (1-\alpha) \left\{ -\frac{1}{n} \sum_{i=1}^{n} g_2(\mathbf{p}_i, t_i) \right\} + \int_0^T \left\{ \int_D e^{g_1(\mathbf{p},t)}\, d\mathbf{p} \right\}^{\alpha} \left\{ \int_D e^{g_2(\mathbf{p},t)}\, d\mathbf{p} \right\}^{1-\alpha} dt \\ &\leq \alpha J(g_1) + (1-\alpha) J(g_2) \quad \forall \alpha \in [0, 1] \quad \forall g_1, g_2 \in V, \end{aligned}$$

where we first applied Hölder's and then Young's inequalities to the integral term. Strict convexity is a consequence of the following observations: Hölder's inequality becomes an equality if and only if e^{g_1} and e^{g_2} are linearly dependent, that is $\exists\, c_1, c_2 \in \mathbb{R}$ s.t. $g_1 = g_2 + \log(c_2/c_1)$; Young's inequality becomes an equality if and only if $\int_D e^{g_1(\mathbf{p},t)}\, d\mathbf{p} = \int_D e^{g_2(\mathbf{p},t)}\, d\mathbf{p}$, meaning that strict convexity is only achieved when $g_1 = g_2$. □

Let V_0 denote the null space of the differential operator with respect to the regularization term $R(g; \lambda)$ in (2). Formally, $V_0 := \{g \in V : \|g\|_b = 0\}$, where $\|\cdot\|_b$ is the norm induced by the inner product $b(g, g)$ in (A.2). Let $V_0^{\perp}$ be the orthogonal space of V_0 such that $V = V_0 \oplus V_0^{\perp}$.

Lemma 2. *V_0 is a finite-dimensional space. Moreover, $\|\cdot\|_b$ is a norm in the space $V_0^{\perp}$ and is equivalent to the V–norm defined in (5).*

Proof. Let $g_0 \in V_0$, meaning that g_0 is the solution of

$$\begin{cases} b(g,g) = 0 & \text{in } \mathcal{D}_T \\ \nabla_D\, g(\mathbf{p},t)\cdot \boldsymbol{n} = 0 & \text{on } \partial\mathcal{D} \text{ for } t \in [0,T) \\ g_t(\mathbf{p},0) = 0 & \text{for } \mathbf{p} \in \mathcal{D}. \end{cases}$$

This is possible if and only if:

$$\begin{cases} \Delta_D g(\mathbf{p},t) = 0 & \text{in } \mathcal{D} \text{ for } t \in [0,T] \\ \nabla_D\, g(\mathbf{p},t)\cdot \boldsymbol{n} = 0 & \text{on } \partial\mathcal{D} \text{ for } t \in [0,T] \end{cases} \wedge \begin{cases} g_{tt}(\mathbf{p},t) = 0 & \text{in } (0,T) \text{ for } \mathbf{p} \in \mathcal{D} \\ g_t(\mathbf{p},0) = g_t(\mathbf{p},T) = 0 & \text{for } \mathbf{p} \in \mathcal{D}. \end{cases}$$

The only admissible solution is the constant function $g_0 = c_0$, where $c_0 \in \mathbb{R}$, defined over $\mathcal{D}_T$. This proves that V_0 is of finite dimension.

Before moving to the second requirement, we observe that

$$\|g\|_b^2 = \int_0^T \|\Delta_D g\|^2_{L^2(D)}\, dt + c\int_0^T \|g_{tt}\|^2_{L^2(D)}\, dt = \|\Delta_D g\|^2_{L^2(0,T;L^2(D))} + c\|g_{tt}\|^2_{L^2(0,T;L^2(D))},$$

where c is the constant defined in (A.2). Moreover, the V–norm can also be defined as follows, using an equivalent norm of the standard $H^2(D)$–norm (Maz'ya, 2013):

$$\|g\|_V^2 = \|g\|^2_{L^2(0,T;L^2(D))} + \|\nabla_D g\|^2_{L^2(0,T;L^2(D))} + \|\Delta_D g\|^2_{L^2(0,T;L^2(D))} + \|g_t\|^2_{L^2(0,T;L^2(D))} + \|g_{tt}\|^2_{L^2(0,T;L^2(D))}.$$

For the equivalence in $V_0^\perp$ we need to show that there exist two real positive constants m and M such that:

$$m\|\tilde{g}\|_b^2 \le \|\tilde{g}\|_V^2 \le M\|\tilde{g}\|_b^2, \qquad \forall \tilde{g} \in V_0^\perp. \tag{C.1}$$

We focus first on the right-hand side inequality, $\|\tilde{g}\|_V^2 \le M\|\tilde{g}\|_b^2$. Applying the provided analogies and noting that $\tilde{g} \in V_0^\perp$ satisfies $\int_0^T \big(g_0, \tilde{g}(\mathbf{p},t)\big)_{L^2(D)}\, dt = 0 \;\; \forall g_0 \in V_0$, we obtain:

$$\|\tilde{g}\|^2_{L^2(0,T;L^2(D))} \le C_p^2 \|\nabla_{D_T}\tilde{g}\|^2_{L^2(0,T;L^2(D))} = C_p^2\|\nabla_D\tilde{g} + \tilde{g}_t\|^2_{L^2(0,T;L^2(D))} \le 2C_p^2\Big(\|\nabla_D\tilde{g}\|^2_{L^2(0,T;L^2(D))} + \|\tilde{g}_t\|^2_{L^2(0,T;L^2(D))}\Big),$$

where, in the first inequality we used classical Poincaré–Sobolev inequality, see Maz'ya (2008, p. 40), and C_p is the Poincaré constant. Then,

$$\begin{aligned} \|\tilde{g}\|_V^2 &\le (2C_p^2+1)\Big(\|\nabla_D\tilde{g}\|^2_{L^2(0,T;L^2(D))} + \|\tilde{g}_t\|^2_{L^2(0,T;L^2(D))}\Big) + \|\Delta_D\tilde{g}\|^2_{L^2(0,T;L^2(D))} + \|\tilde{g}_{tt}\|^2_{L^2(0,T;L^2(D))} \\ &\le \tilde{C}\Big(\|\Delta_D\tilde{g}\|^2_{L^2(0,T;L^2(D))} + \|\tilde{g}_{tt}\|^2_{L^2(0,T;L^2(D))}\Big) \le M\|\tilde{g}\|_b^2, \end{aligned}$$

where $M = \tilde{C}\max\{1, 1/c\}$, and the second inequality holds thanks to the homogeneous Neumann boundary conditions. The left-hand side inequality in (C.1) is straightforward and is satisfied for $m = \min\{1, 1/c\}$. □

Appendix D. Proof of Theorem 2

In this section, our focus turns to the consistency of the minimizer for $L(g)$ as outlined in (1), operating within the functional space V as characterized in (4). The proof of Theorem 2 follows from Theorem 5.3 in Gu and Qiu (1993), subject to the fulfillment of the essential hypotheses for its validity. To ensure the requirements we rely on the following two Lemmas.

Lemma 3. *Let g_0 be the true log-density and $\hat{g}$ be the minimizer of $L(g)$ defined as in (1). Then*

$$\begin{aligned} D_{sKL}(g_0,\hat{g}) =& 2\lambda\left[\int_{D_T} \Delta_D\hat{g}\Delta_D(g_0-\hat{g})\, d\mathbf{p}\, dt + c\int_{D_T}\frac{\partial^2\hat{g}}{\partial t^2}\frac{\partial^2(g_0-\hat{g})}{\partial t^2}\, d\mathbf{p}\, dt\right] \\ &+ \frac{1}{n}\sum_{i=1}^n(\hat{g}-g_0)(X_i) - \mathbb{E}_{g_0}[\hat{g}-g_0]. \end{aligned}$$

Proof. Set

$$A_{g,h}(\alpha) := -\frac{1}{n}\sum_{i=1}^n (g+\alpha h)(X_i) + \int_{D_T} e^{g+\alpha h}\, d\mathbf{p}\, dt + \lambda\left[\int_{D_T}\big(\Delta_D(g+\alpha h)\big)^2\, d\mathbf{p}\, dt + c\int_{D_T}\left(\frac{\partial^2(g+\alpha h)}{\partial t^2}\right)^2 d\mathbf{p}\, dt\right].$$

By differentiating $A_{g,h}$ with respect to α and setting $\alpha = 0$, it results that

$$\frac{\partial A_{g,h}}{\partial\alpha}(0) = -\frac{1}{n}\sum_{i=1}^n h(X_i) + \mathbb{E}_g[h] + 2\lambda\left[\int_{D_T}\Delta_D g\Delta_D h\, d\mathbf{p}\, dt + c\int_{D_T}\frac{\partial^2 g}{\partial t^2}\frac{\partial^2 h}{\partial t^2}\, d\mathbf{p}\, dt\right].$$

Observe that the term above is null, since $\hat{g}$ is the minimizer of (1). Finally, setting $g = \hat{g}$ and $h = \hat{g} - g_0$ we obtain the desired result. □

Lemma 4. *Under Assumption 1, there exists an infinite set of functions $\{\phi_j(\mathbf{p})\}$ and $\{\psi_i(t)\}$ of space and time eigenfunctions such that, by setting $\{\omega_k(\mathbf{p},t)\} = \{\phi_j(\mathbf{p})\psi_i(t)\}$, then in $L^2(0,T;L^2(D))$ it follows that:*

$$Cov(\omega_k,\omega_l) = \delta_{kl} \quad \text{and} \quad \int_{D_T}\left(\Delta_D\omega_k\Delta_D\omega_l + c\,\frac{\partial^2\omega_k}{\partial t^2}\frac{\partial^2\omega_l}{\partial t^2}\,d\mathbf{p}\,dt\right) = \left(\eta_k^2 + c\,\xi_k^2\right)\delta_{kl},$$

where δ_{kl} denotes the Kronecker delta function, $0 \le \eta_k \to 0$, and $0 \le \xi_k \to 0$. Moreover, as $k \to \infty$ the following asymptotic property holds:

$$\eta_k^2 + c\,\xi_k^2 \sim k^2.$$

Proof. The proof exploits the peculiar formulation chosen for the regularization term $R(g;\lambda)$, which accounts separately for the regularity of the density function in spatial and temporal domains. This implies that we can consider the following two problems: the one of finding the eigenvalues and eigenfunctions of the operator Δ_D when time is fixed; and the one of finding the eigenvalues and eigenfunctions of the operator $\partial^2/\partial t^2$ when space is fixed, with homogeneous Neumann boundary conditions in space and time respectively. That is:

$$\begin{cases}\Delta_D\phi(\mathbf{p}) = \eta\phi(\mathbf{p}) & \text{in } D\\ \nabla_D\,\phi(\mathbf{p})\cdot\boldsymbol{n} = 0 & \text{on } \partial D\end{cases} \quad \text{and} \quad \begin{cases}\psi_{tt}(t) = \xi\psi(t) & \text{in } (0,T)\\ \psi_t(0) = \psi_t(T) = 0.\end{cases}$$

According to the related literature (Brezis, 2011; Chavel, 1984; Agmon, 2010), the eigenvalues η_k and ξ_k are infinitely many and $\eta_k, \xi_k \to \infty$ as k increases. In addition, if D_T is bounded and regular enough, then $\eta_k \sim \frac{|D|}{4\pi}k$ and $\xi_k \sim k^{1/2}$ as $k \to \infty$, leading the desired asymptotic behavior. The corresponding eigenfunctions can be orthonormalized in $L^2(D)$ and $L^2(T)$ respectively. Therefore:

$$\int_{D_T}\omega_k(\mathbf{p},t)\omega_l(\mathbf{p},t)\,d\mathbf{p}\,dt = \int_0^T\psi_k(t)\psi_l(t)\,dt\int_D\phi_k(\mathbf{p})\phi_l(\mathbf{p})\,d\mathbf{p} = \delta_{kl}$$

and

$$\int_{D_T}\left(\Delta_D\omega_k\Delta_D\omega_l + \tilde{\lambda}\frac{\partial^2\omega_k}{\partial t^2}\frac{\partial^2\omega_l}{\partial t^2}\right)d\mathbf{p}\,dt = \int_{D_T}\left[\left(\psi_k\Delta_D\phi_k\right)\left(\psi_l\Delta_D\phi_l\right) + c\left(\phi_k\frac{\partial^2\psi_k}{\partial t^2}\right)\left(\phi_l\frac{\partial^2\psi_l}{\partial t^2}\right)\right]d\mathbf{p}\,dt = \left(\eta_k^2 + c\,\xi_k^2\right)\delta_{kl}.$$

Finally, under Assumption 1, the standard $L^2(0,T;L^2(D))$ scalar product can be substituted by the scalar product induced by the true log-density g_0, that is $\int_{D_T}\omega_k\omega_l e^{g_0}\,d\mathbf{p}\,dt$, obtaining that $\mathrm{Cov}(\omega_k,\omega_l) = \delta_{kl}$. □

Appendix E. Data generation for simulations 1, 2 and 3

The true spatio-temporal density function used in Simulation 1 of Section 6.3 is a mixture of 4 time-dependent Gaussian distributions over the square $[-6,6]\times[-6,6]$, obtained by introducing variability over time to the density adopted in Ferraccioli et al. (2021). Mean and variances are given by

$$\mu_1 = \begin{bmatrix}-2\\ -1.5-0.5t\end{bmatrix} \qquad \Sigma_1 = \begin{bmatrix}0.8 & -0.2-0.4t\\ -0.2-0.4t & 0.8\end{bmatrix}$$

$$\mu_2 = \begin{bmatrix}2+t\\ -2-t\end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix}1.5-0.5t & 0\\ 0 & 1.5-0.5t\end{bmatrix}$$

$$\mu_3 = \begin{bmatrix}-2\\ 1.5+1.5t\end{bmatrix} \qquad \Sigma_3 = \begin{bmatrix}0.8+t & 0\\ 0 & 0.8\end{bmatrix}$$

$$\mu_4 = \begin{bmatrix}2\\ 2-t\end{bmatrix} \qquad \Sigma_4 = \begin{bmatrix}1 & 0.9-0.3t\\ 0.9-0.3t & 1\end{bmatrix}$$

for $t \in [0,1]$. We set mixing weights as $\pi(t) \propto \left[\eta_{avg}(\Sigma_1|t),\quad \eta_{avg}(\Sigma_2|t),\quad \eta_{avg}(\Sigma_3|t),\quad \eta_{avg}(\Sigma_4|t)\right]^\top$, where $\eta_{avg}(A)$ denotes the average of the eigenvalues of the matrix A. Thanks to these mixing weights, a larger number of points is sampled from the distributions presenting stronger anisotropy.

The true spatio-temporal density function used in Simulation 2 of Section 6.4 is a mixture of 5 time-dependent Kent distributions over the surface of the unit sphere, with scale parameters (concentration κ_i and ovalness β_i) given by

$\kappa_1 = 18$	$\kappa_2 = 15$	$\kappa_3 = 20$	$\kappa_4 = 20$	$\kappa_5 = 20$
$\beta_1 = 0$	$\beta_2 = 7$	$\beta_3 = 10$	$\beta_4 = 7$	$\beta_5 = 4$

and position parameters (mean directions $\gamma_{1,i}$, major axis $\gamma_{2,i}$ and minor axis $\gamma_{3,i}$) given by

$$\gamma_{1,1} = \begin{bmatrix}-0.5+t\\ -0.5+t\\ 0.8+t\end{bmatrix} \qquad \gamma_{2,1} = \begin{bmatrix}-0.7789\\ 0.6157\\ 0.1188\end{bmatrix} \qquad \gamma_{3,1} = \begin{bmatrix}-0.5696\\ -0.6154\\ -0.5449\end{bmatrix}$$

$$\gamma_{1,2} = \begin{bmatrix}-0.3\\ -0.3\\ 0.2\end{bmatrix} \qquad \gamma_{2,2} = \begin{bmatrix}-0.8651\\ 0.3803\\ -0.3270\end{bmatrix} \qquad \gamma_{3,2} = \begin{bmatrix}0.1483\\ -0.4289\\ -0.8911\end{bmatrix}$$

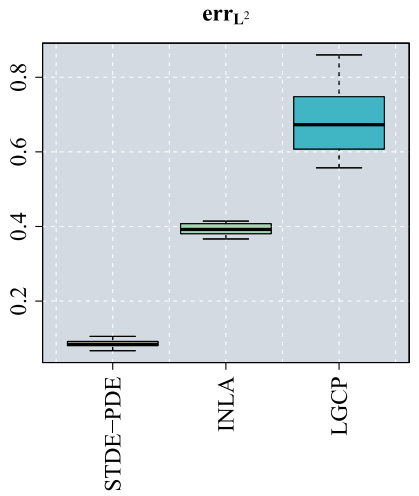


Fig. F.14. Simulation 4: intensity estimation of inhomogeneous Poisson Process. Boxplots of L^2–norm of the error of the estimates of the intensity provided by the proposed STDE-PDE and the considered competing method for intensity estimation, namely INLA and LGCP, over 30 simulation repetitions.

$$\gamma_{1,3} = \begin{bmatrix} 0.5 \\ -0.5 \\ 0.8 \end{bmatrix} \quad \gamma_{2,3} = \begin{bmatrix} -0.6665 \\ -0.7432 \\ -0.0584 \end{bmatrix} \quad \gamma_{3,3} = \begin{bmatrix} 0.5754 \\ -0.4629 \\ -0.6743 \end{bmatrix}$$

$$\gamma_{1,4} = \begin{bmatrix} 0.2 \\ -1 \\ 0 \end{bmatrix} \quad \gamma_{2,4} = \begin{bmatrix} 0.6365 \\ -0.0304 \\ -0.7707 \end{bmatrix} \quad \gamma_{3,4} = \begin{bmatrix} -0.7546 \\ -0.2314 \\ -0.6140 \end{bmatrix}$$

$$\gamma_{1,5} = \begin{bmatrix} 0.6 + t \\ -0.5 + t \\ 0.3 + t \end{bmatrix} \quad \gamma_{2,5} = \begin{bmatrix} 0.6365 \\ -0.0304 \\ -0.7707 \end{bmatrix} \quad \gamma_{3,5} = \begin{bmatrix} 0.7546 \\ -0.2314 \\ -0.6140 \end{bmatrix}$$

for $t \in [0, 1]$. The mean directions $\gamma_{1,1}, \dots, \gamma_{1,5}$ are then normalized to lay on the unit sphere (while major and minor axes are already normalized in the definitions above). Finally, we use equal mixing weights, thus sampling approximately the same number of points from each distribution in the mixture.

The true spatio-temporal density function used in Simulation 3 of Section 6.5 is a mixture of 25 Kent distributions over the surface of the unit sphere. All the components have the same time-dependent scale parameters: $\kappa = 400$ and $\beta = 125 + 100\,|t - 0.5|$. Position parameters, instead, are constant over time, and their values are here omitted due to space constraints.

Appendix F. Simulation 4: intensity estimation of inhomogeneous Poisson process

In this section, we report the results obtained in a simulated intensity estimation problem, analogous to the density estimation problem described in Section 6.3. We here consider the proposed STDE-PDE for intensity estimation, as described in Section 2 and detailed in Appendix B, and compare it to LGCP and INLA. All methods are run under the same specifications described in Section 6.3 for Simulation 1. Fig. F.14 displays the boxplots over 30 simulation repetitions of the err_{L^2}, computed as detailed in Section 6.2. The other measures of error defined in Section 6.2 are metrics in the space of probability distributions and therefore are not suitable for intensity functions. The ordering of methods remains unchanged, and we can draw the same considerations already discussed for Simulation 1 in the density estimation setting.

Appendix G. Simulation 5: additional simulation on planar domain

In this section, we show an additional simulation on a planar domain. The true spatio-temporal density function is a mixture of 4 time-dependent Gaussian distributions over the square $[-6, 6] \times [-6, 6]$, with the following set of parameters:

$$\mu_1 = \begin{bmatrix} -3.5 + 1.5t \\ -3.5 + 1.5t \end{bmatrix} \quad \Sigma_1 = 3.0 \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.8 \end{bmatrix}$$

$$\mu_2 = \begin{bmatrix} 1.7 + t \\ -2 - 1.4t \end{bmatrix} \quad \Sigma_2 = 3.0 \begin{bmatrix} 1.5 - 0.5t & -0.5t \\ -0.5t & 1.5 - 0.5t \end{bmatrix}$$

$$\mu_3 = \begin{bmatrix} -2 - t \\ 1.5 + 1.7t \end{bmatrix} \quad \Sigma_3 = 1.0 \begin{bmatrix} 1.6 & -1.2t \\ -1.2t & 1.6 \end{bmatrix}$$

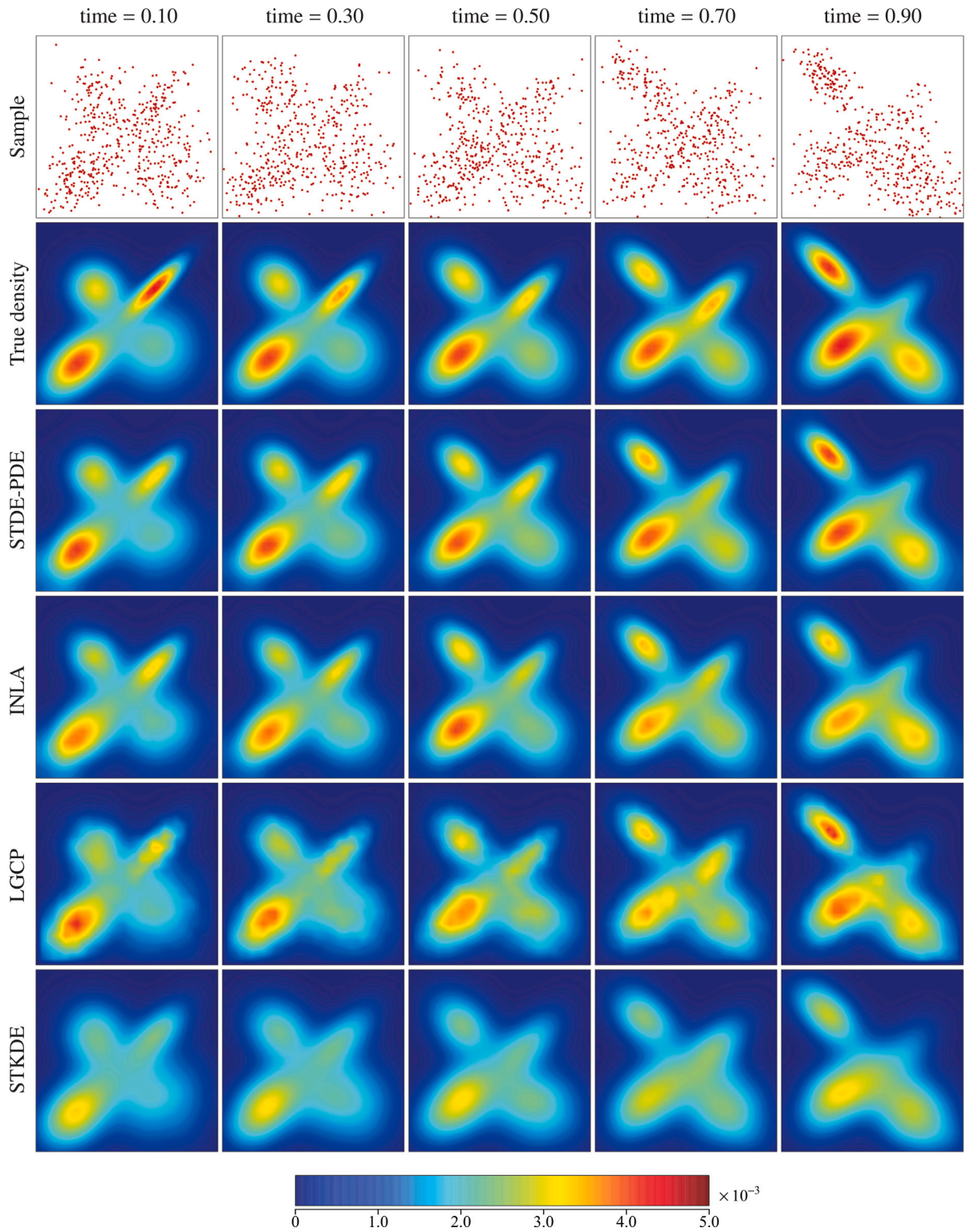


Fig. G.15. Simulation 5, additional simulation on planar domain. View of the sample (first row), true density (second row), and mean estimates of STDE-PDE (third row), INLA (fourth row), LGCP (fifth row), and STKDE (sixth row), over 30 simulation repetitions, at different time instances. All the plots have the same color scale, with bluish colors corresponding to low-density values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

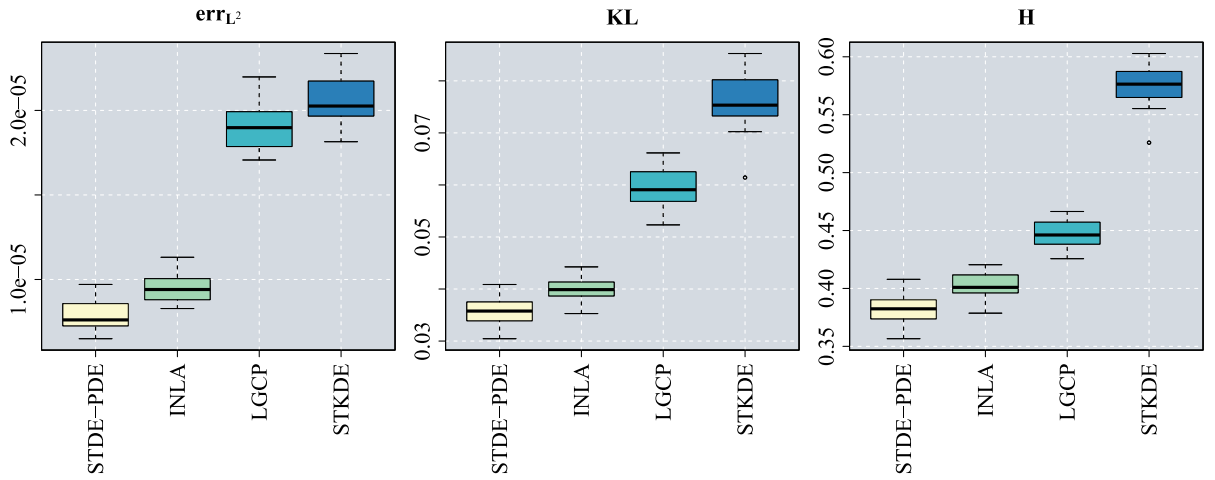


Fig. G.16. Simulation 5, additional simulation on planar domain. Boxplots of L^2–norm of the error (left), Kullback–Leibler divergence (center), and Hellinger distance (right) of the estimates provided by proposed STDE-PDE and the competing methods introduced in Section 6.1, namely INLA, LGCP, and STKDE, over 30 simulation repetitions.

$$\mu_4 = \begin{bmatrix} 2 \\ 2 - 1.5t \end{bmatrix} \qquad \Sigma_4 = 1.5 \begin{bmatrix} 1 & 0.9 - 0.3t \\ 0.9 - 0.3t & 1 \end{bmatrix}$$

for $t \in [0, 1]$. We set mixing weights as $\pi(t) = \left[0.7\,\eta\,(a), \quad 0.8\,\eta\,(a), \quad 0.4\,\eta\,(a), \quad 0.25\,\eta\,(b)\right]^\top$, where $a = 10t + 1$, $b = 10\,(0.5 + \text{sign}(0.5 - t)(t - 0.5)) + 2$, and $\eta(\cdot)$ denotes the Gamma density with shape parameter 7.5 and unit scale parameter. The top panels of Fig. G.15 show the true density: such density encompasses both high-frequency and low-frequency components, that may significantly overlap; moreover, the density has significant mass close to the boundary of the domain.

Fig. G.15 shows the estimated fields and Fig. G.16 reports the errors, showing that the ordering of the methods remains the same as for Simulation 1. For the two methods that consider the shape of the domain, STDE-PDE and INLA, we observe some minor edge-effects in the estimates. This is to be expected in this simulation setting, where the shape of the domain only influences the observability of the process, but not the process itself. This fact has a higher impact on the coverage of the confidence intervals. INLA turns out to be over-conservative, with a coverage equal to one (median coverage = 1.000, IQR = 0.000), while STDE-PDE appears under-conservative (median coverage = 0.933, IQR = 0.133).

Appendix H. Simulation 6: point pattern on a non-spherical two-dimensional manifold

In this section, we include an additional simulation on a general two-dimensional Riemannian manifold. Since we could not find available implementations of competing methods capable to handle space–time point patterns over this type of domains, we here consider as a benchmark method k-Nearest Neighbors Spatio-Temporal Density Estimation, referred to as KNNSTDE. This model separably combines the estimates obtained in space and time (Mack and Rosenblatt, 1979). We implement KNNSTDE using the `R` package `TDA` (Fasy et al., 2024), which relies on Euclidean distances in $\mathbb{R}^d$ (with $d = 2, 3$) rather than on geodetic distances. To facilitate this competing technique, we sample data according to a first-order separable density function, meaning that $f(\mathbf{p}, t) = f_D(\mathbf{p}) f_T(t)$, and we define the spatial component based on Euclidean distances. More specifically, we consider $f_D(\mathbf{p}) = \left(\sin(\mathbf{p}_x) + \cos(\mathbf{p}_y) + \mathbf{p}_z + 5.1590\right)^3$, with $\mathbf{p}_x$, $\mathbf{p}_y$ and $\mathbf{p}_z$ denoting the Cartesian coordinates of the location $\mathbf{p}$, whereas $f_T(t)$ is a Beta distribution over $[0, 1]$ with shape parameters $\alpha = \beta = 2$. Such components are then normalized to get a proper density on the considered D_T. Given the shape of $f_T(t)$, the sample size evolves symmetrically in $[0, 1]$, increasing until time $t = 0.50$ and then decreasing again. We sample 10 000 data from the true density, and repeat the sampling 30 times. Fig. H.17 shows the data sampled on the surface domain for the first simulation repetition.

We compute STDE-PDE estimates using a mesh of 964 nodes, and 6 time knots in $[0, 1]$. Smoothing parameters are selected by 10–fold cross-validation, as detailed in Section 4.3. Fig. H.18 compares the true density and the mean estimates, over 30 simulation repetitions, at $t = 0.10$. The presented method accurately estimates all the modes of the true density, also those falling close to the boundary. This is also true at the beginning of the temporal interval, at $t = 0.10$, even though STDE-PDE does not exploit the first-order separability assumption, differently from KNNSTDE, and solely a fairly limited number of observations is available, as appreciable from the left panel in Fig. H.17. Although this simulation has been engineered to facilitate the benchmark method KNNSTDE, the proposed STDE-PDE outperforms the results achieved by KNNSTDE, as also illustrated by the error boxplots in Fig. H.19.

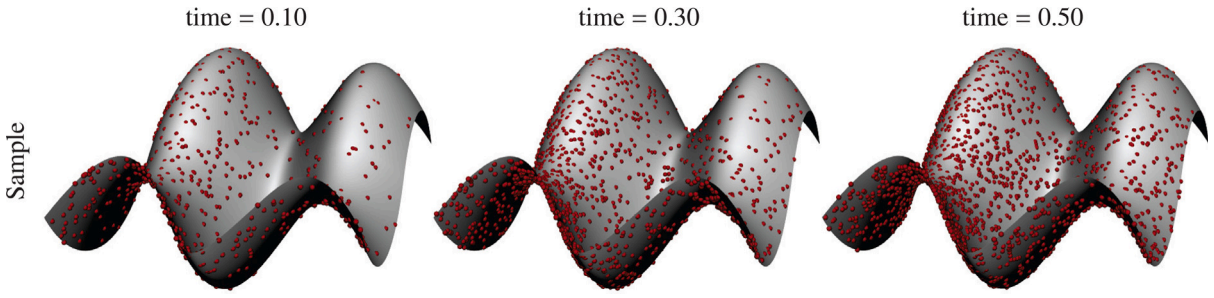


Fig. H.17. Simulation 6: point pattern on a non-spherical two-dimensional manifold. Front view of the sample at different time instances over the temporal domain $[0, 1]$.

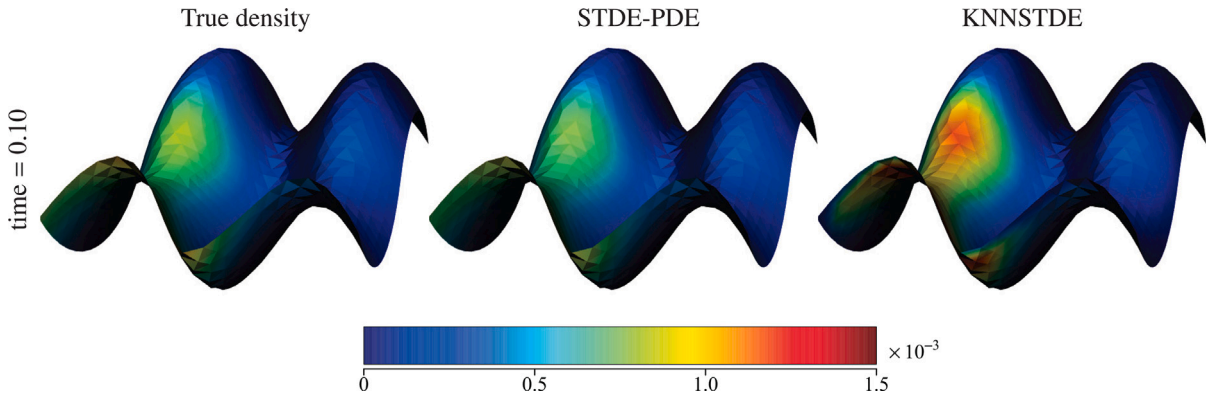


Fig. H.18. Simulation 6: point pattern on a non-spherical two-dimensional manifold. True density (left), and mean estimates over 30 simulation repetitions obtained by the proposed STDE-PDE (center) and KNNSTDE (right), at a time instant at the beginning of the temporal domain $[0, 1]$. All the plots have the same color scale, with bluish colors corresponding to low density values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

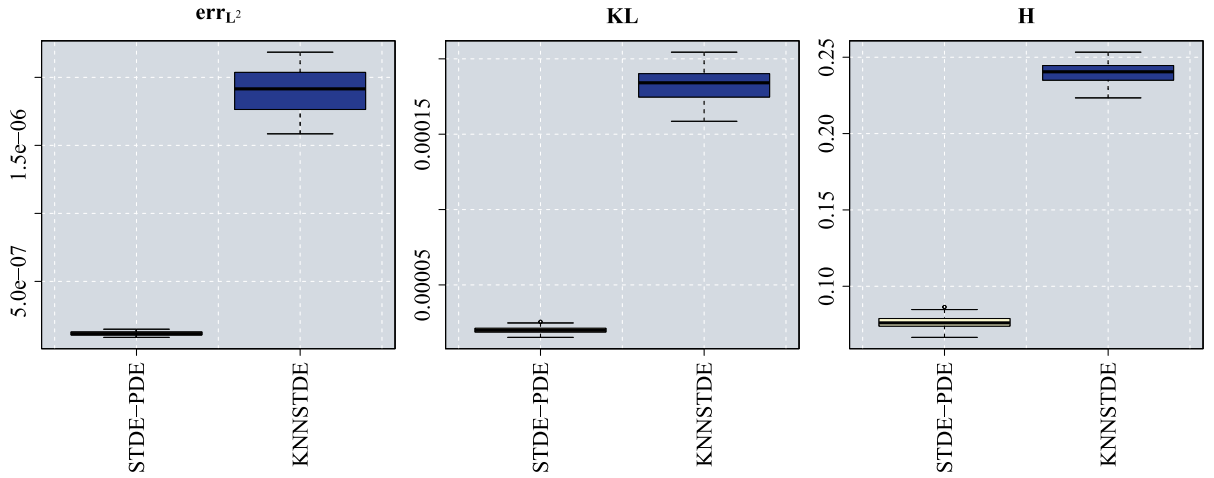


Fig. H.19. Simulation 6: point pattern on a non-spherical two-dimensional manifold. Boxplots of L^2–norm of the error (left), Kullback–Leibler divergence (center) and Hellinger distance (right), of the estimates provided by the proposed STDE-PDE, and the considered competing method KNNSTDE, over 30 simulation repetitions.

References

Adams, R.A., Fournier, J.J., 2003. Sobolev Spaces. Elsevier.
Agmon, S., 2010. Lectures on Elliptic Boundary Value Problems, vol. 369, American Mathematical Society.

Arnone, E., Azzimonti, L., Nobile, F., Sangalli, L.M., 2019. Modeling spatially dependent functional data via regression with differential regularization. J. Multivariate Anal. 170, 275–295.
Arnone, E., Clemente, A., Sangalli, L.M., Lila, E., Ramsay, J., Formaggia, L., 2023a. fdaPDE: physics-informed spatial and functional data analysis. R package version 1.1-17, https://CRAN.R-project.org/package=fdaPDE.
Arnone, E., Ferraccioli, F., Pigolotti, C., Sangalli, L.M., 2022. A roughness penalty approach to estimate densities over two-dimensional manifolds. Comput. Statist. Data Anal. 174.
Arnone, E., Negri, L., Panzica, F., Sangalli, L.M., 2023b. Analyzing data in complicated 3D domains: Smoothing, semiparametric regression, and functional principal component analysis. Biometrics.
Arnone, E., Sangalli, L.M., Vicini, A., 2023c. Smoothing spatio-temporal data with complex missing data patterns. Stat. Model. 23 (4), 327–356.
Augustin, N.H., Trenkel, V.M., Wood, S.N., Lorance, P., 2013. Space-time modelling of blue ling for fisheries stock management. Environmetrics 24 (2), 109–119.
Bachl, F.E., Lindgren, F., Borchers, D.L., Illian, J.B., 2019. inlabru: an R package for Bayesian spatial modelling from ecological survey data. Methods Ecol. Evol. 10 (6), 760–766.
Baddeley, A., Rubak, E., Turner, R., 2015. Spatial Point Patterns: Methodology and Applications with R. CRC Press.
Bayisa, F.L., Ådahl, M., Rydén, P., Cronie, O., 2020. Large-scale modelling and forecasting of ambulance calls in northern Sweden using spatio-temporal log-Gaussian Cox processes. Spatial Stat. 39.
Bernardi, M.S., Sangalli, L.M., Mazza, G., Ramsay, J.O., 2017. A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province. Stoch. Environ. Res. Risk Assess. 31, 23–38.
Bertin, K., Klutchnikoff, N., Ouimet, F., 2023. A new adaptive local polynomial density estimation procedure on complicated domains. arXiv preprint arXiv:2308.01156.
Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-Temporal Bayesian Models with R-INLA. John Wiley & Sons.
Brezis, H., 2011. Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer.
Brix, A., Diggle, P.J., 2001. Spatiotemporal prediction for log-Gaussian Cox processes. J. R. Stat. Soc. Ser. B Stat. Methodol. 63 (4), 823–841.
Casella, G., Robert, C.P., Wells, M.T., 2004. Generalized accept-reject sampling schemes. In: Lecture Notes-Monograph Series, JSTOR, pp. 342–347.
Chavel, I., 1984. Eigenvalues in Riemannian Geometry. Academic Press.
Chen, R.T., Amos, B., Nickel, M., 2021. Neural spatio-temporal point processes. In: International Conference on Learning Representations.
Csiszár, I., 1975. I-divergence geometry of probability distributions and minimization problems. Ann. Probab. 146–158.
Daley, D.J., Vere-Jones, D., 2003. An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods. Springer.
Daley, D.J., Vere-Jones, D., 2008. An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure. Springer.
D'Angelo, N., Adelfio, G., Mateu, J., 2023. Locally weighted minimum contrast estimation for spatio-temporal log-Gaussian Cox processes. Comput. Statist. Data Anal. 180.
Daniel, J., Horrocks, J., Umphrey, G.J., 2018. Penalized composite likelihoods for inhomogeneous Gibbs point process models. Comput. Statist. Data Anal. 124, 104–116.
Dassi, F., Ettinger, B., Perotto, S., Sangalli, L.M., 2015. A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain. Appl. Numer. Math. 90, 111–131.
Davies, T.M., Marshall, J.C., 2023. sparr: spatial and spatiotemporal relative risk. R package version 2.3-10, https://CRAN.R-project.org/package=sparr.
Davies, T.M., Marshall, J.C., Hazelton, M.L., 2018. Tutorial on kernel estimation of continuous spatial and spatiotemporal relative risk. Stat. Med. 37 (7), 1191–1221.
De Boor, C., 1978. A Practical Guide to Splines, vol. 27. Springer.
Diggle, P.J., 2013. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns. CRC Press.
Diggle, P.J., Knorr-Held, L., Rowlingson, B.S., Su, T.L., Hawtin, P., Bryant, T.N., 2009. On-line monitoring of public health surveillance data. In: Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance. Oxford University Press, pp. 233–266.
Diggle, P.J., Moraga Serrano, P.E., Rowlingson, B.S., Taylor, B., 2013. Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. Statist. Sci. 28 (4), 542–563.
Diggle, P., Rowlingson, B., Su, T.-l., 2005. Point process methodology for on-line spatio-temporal disease surveillance. Environmetrics 16 (5), 423–434.
Drost, H.G., Nowosad, J., 2024. philentropy: similarity and distance quantification between probability. R package version 0.8.0, https://CRAN.R-project.org/package=philentropy.
Elson, R., Davies, T.M., Jenkins, C., Vivancos, R., O'Brien, S.J., Lake, I.R., 2020. Application of kernel smoothing to estimate the spatio-temporal variation in risk of STEC O157 in England. Spatial Spatio-Temporal Epidemiol. 32.
Elson, R., Davies, T.M., Lake, I.R., Vivancos, R., Blomquist, P.B., Charlett, A., Dabrera, G., 2021. The spatio-temporal distribution of COVID-19 infection in England between january and june 2020. Epidemiol. Infect. 149.
Fasy, B.T., Kim, J., Lecci, F., Maria, C., Millman, D.L., Rouvreau, V., 2024. TDA: statistical tools for topological data analysis. R package version 1.9.1, https://CRAN.R-project.org/package=TDA.
Fernando, W.S., Hazelton, M.L., 2014. Generalizing the spatial relative risk function. Spatial Spatio-Temporal Epidemiol. 8, 1–10.
Ferraccioli, F., Arnone, E., Finos, L., Ramsay, J.O., Sangalli, L.M., 2021. Nonparametric density estimation over complicated domains. J. R. Stat. Soc. Ser. B Stat. Methodol. 83 (2), 346–368.
Fishman, P., Snyder, D., 1976. The statistical analysis of space-time point processes. IEEE Trans. Inform. Theory 22 (3), 257–274.
Fuglstad, G.A., Simpson, D., Lindgren, F., Rue, H., 2019. Constructing priors that penalize the complexity of Gaussian random fields. J. Amer. Statist. Assoc. 114 (525), 445–452.
Furrer, E.M., Nychka, D.W., 2007. A framework to understand the asymptotic properties of kriging and splines. J. Korean Statist. Soc. 36 (1), 57–76.
Gabriel, E., 2014. Estimating second-order characteristics of inhomogeneous spatio-temporal point processes: influence of edge correction methods and intensity estimates. Methodol. Comput. Appl. Probab. 16, 411–431.
Genovese, C., Wasserman, L., 2008. Adaptive conficence bands. Ann. Statist. 36 (2), 875–905.
Gervini, D., 2022. Doubly stochastic models for spatio-temporal covariation of replicated point processes. Canad. J. Statist. 50 (1), 287–303.
Ghorbani, M., Vafaei, N., Dvořák, J., Myllymäki, M., 2021. Testing the first-order separability hypothesis for spatio-temporal point patterns. Comput. Statist. Data Anal. 161.
Gómez-Rubio, V., 2020. Bayesian Inference with INLA. CRC Press.
González, J.A., Rodríguez-Cortés, F.J., Cronie, O., Mateu, J., 2016. Spatio-temporal point process statistics: a review. Spatial Stat. 18, 505–544.
Goodd, I.J., Gaskins, R.A., 1971. Nonparametric roughness penalties for probability densities. Biometrika 58 (2), 255–277.
Gu, C., 1993. Smoothing spline density estimation: A dimensionless automatic algorithm. J. Amer. Statist. Assoc. 88 (422), 495–504.
Gu, C., Jeon, Y., Lin, Y., 2013. Nonparametric density estimation in high-dimensions. Statist. Sinica 1131–1153.
Gu, C., Qiu, C., 1993. Smoothing spline density estimation: Theory. Ann. Statist. 21 (1), 217–234.
Gu, C., Wang, J., 2003. Penalized likelihood density estimation: Direct cross-validation and scalable approximation. Statist. Sinica 811–826.
Guan, Y., Jalilian, A., Waagepetersen, R., 2015. Quasi-likelihood for spatial point processes. J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (3), 677–697.
Hall, P., Horowitz, J., 2013. A simple bootstrap method for constructing nonparametric confidence bands for functions. Ann. Statist. 1892–1921.
Hering, A.S., Bell, C.L., Genton, M.G., 2009. Modeling spatio-temporal wildfire ignition point patterns. Environ. Ecol. Stat. 16, 225–250.

Hjelle, Ø., Dæhlen, M., 2006. Triangulations and Applications. Springer Science & Business Media.
Hu, Y., Wang, F., Guin, C., Zhu, H., 2018. A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation. Appl. Geogr. 99, 89–97.
Illian, J.B., Hendrichsen, D.K., 2010. Gibbs point process models with mixed effects. Environmetrics Off. J. Int. Environmetrics Soc. 21 (3–4), 341–353.
Jost, J., 2008. Riemannian Geometry and Geometric Analysis, vol. 42005, Springer.
Kaimi, I., Diggle, P., 2011. A hierarchical model for real-time monitoring of variation in risk of non-specific gastrointestinal infections. Epidemiol. Infect. 139 (12), 1854–1862.
Kent, J.T., 1982. The Fisher-Bingham distribution on the sphere. J. R. Stat. Soc. Ser. B Stat. Methodol. 44 (1), 71–80.
Kimeldorf, G.S., Wahba, G., 1970. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. Ann. Math. Stat. 41 (2), 495–502.
Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., Rue, H., 2018. Advanced Spatial Modeling with Stochastic Partial Differential Equations using R and INLA. Chapman and Hall/CRC.
Lange, K., 2013. Optimization, vol. 95, Springer Science & Business Media.
Laslett, G.M., 1994. Kriging and splines: an empirical comparison of their predictive performance in some applications. J. Amer. Statist. Assoc. 89 (426), 391–400.
Leonard, T., 1982. An Empirical Bayesian Approach to the Smooth Estimation of Unknown Functions. University of Wisconsin-Madison. Mathematics Research Center.
Li, Y., Guan, Y., 2014. Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance. J. Amer. Statist. Assoc. 109 (507), 1205–1215.
Lindgren, F., Bachl, F.E., 2023. inlabru: Bayesian latent Gaussian modelling using INLA and extensions. R package version 2.10.1, https://CRAN.R-project.org/package=inlabru.
Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73 (4), 423–498.
Liu, D.C., Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. Math. Program. 45 (1–3), 503–528.
Loecher, M., 2023. RgoogleMaps: overlays on static maps. R package version 1.5.1, https://CRAN.R-project.org/package=RgoogleMaps.
Mack, Y., Rosenblatt, M., 1979. Multivariate k-nearest neighbor density estimates. J. Multivariate Anal. 9 (1), 1–15.
Mardia, K.V., Jupp, P.E., 2000. Directional Statistics, vol. 2, Wiley Online Library.
Marra, G., Miller, D.L., Zanin, L., 2012. Modelling the spatiotemporal distribution of the incidence of resident foreign population. Stat. Neerl. 66 (2), 133–160.
Maz'ya, V., 2008. Sobolev Spaces in Mathematics II: Applications in Analysis and Partial Differential Equations, vol. 9, Springer Science & Business Media.
Maz'ya, V., 2013. Sobolev Spaces. Springer.
Medinas, D., Marques, J.T., Costa, P., Santos, S., Rebelo, H., Barbosa, A.M., Mira, A., 2021. Spatiotemporal persistence of bat roadkill hotspots in response to dynamics of habitat suitability and activity patterns. J. Environ. Manag. 277.
Menafoglio, A., Guadagnini, A., Secchi, P., 2014. A kriging approach based on aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers. Stoch. Environ. Res. Risk Assess. 28, 1835–1851.
Menafoglio, A., Secchi, P., Guadagnini, A., 2016. A class-kriging predictor for functional compositions with application to particle-size curves in heterogeneous aquifers. Math. Geosci. 48, 463–485.
Møller, J., Waagepetersen, R.P., 2007. Modern statistics for spatial point processes. Scand. J. Stat. 34 (4), 643–684.
Møller, J., Waagepetersen, R., 2017. Some recent developments in statistics for spatial point patterns. Annu. Rev. Stat. Appl. 4, 317–342.
Neal, R.M., 2003. Slice sampling. Ann. Statist. 31 (3), 705–767.
Nocedal, J., 1980. Updating quasi-Newton matrices with limited storage. Math. Comp. 35 (151), 773–782.
Nocedal, J., Wright, S.J., 1999. Numerical Optimization. Springer.
O'Sullivan, F., 1988. Fast computation of fully automated log-density and log-hazard estimators. SIAM J. Sci. Stat. Comput. 9 (2), 363–379.
Pollard, D., 2002. A User's Guide to Measure Theoretic Probability. (8), Cambridge University Press.
Python, A., Illian, J.B., Jones-Todd, C.M., Blangiardo, M., 2019. A Bayesian approach to modelling subnational spatial dynamics of worldwide non-state terrorism, 2010–2016. J. R. Stat. Soc. Ser. A Stat. Soc. 182 (1), 323–344.
Quarteroni, A., 2009. Numerical Models for Differential Problems, vol. 2, Springer.
Ramsay, J.O., 2000. Differential equation models for statistical functions. Canad. J. Statist. 28 (2), 225–240.
Ramsay, T., 2002. Spline smoothing over difficult regions. J. R. Stat. Soc. Ser. B Stat. Methodol. 64 (2), 307–319.
Rodrigues, A., Diggle, P.J., 2012. Bayesian estimation and prediction for inhomogeneous spatiotemporal log-Gaussian Cox processes using low-rank models, with application to criminal surveillance. J. Amer. Statist. Assoc. 107 (497), 93–101.
Rodríguez de Rivera, Ó., López-Quílez, A., Blangiardo, M., 2018. Assessing the spatial and spatio-temporal distribution of forest species via Bayesian hierarchical modeling. Forests 9 (9), 573.
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2), 319–392.
Sangalli, L.M., 2021. Spatial regression with partial differential equation regularisation. Internat. Statist. Rev. 89 (3), 505–531.
Sangalli, L.M., Ramsay, J.O., Ramsay, T.O., 2013. Spatial spline regression models. J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (4), 681–703.
Scimone, R., Menafoglio, A., Sangalli, L.M., Secchi, P., 2022. A look at the spatio-temporal mortality patterns in Italy during the COVID-19 pandemic through the lens of mortality densities. Spatial Stat. 49.
Serra, L., Saez, M., Mateu, J., Varga, D., Juan, P., Díaz-Ávalos, C., Rue, H., 2014. Spatio-temporal log-Gaussian Cox processes for modelling wildfire occurrence: the case of Catalonia, 1994–2008. Environ. Ecol. Stat. 21, 531–563.
Shirota, S., Gelfand, A.E., 2017. Space and circular time log Gaussian Cox processes with application to crime event data. Ann. Appl. Stat. 481–503.
Silverman, B.W., 1982. On the estimation of a probability density function by the maximum penalized likelihood method. Ann. Statist. 795–810.
Streit, R.L., Streit, R.L., 2010. The Poisson Point Process. Springer.
Tapia, R.A., Thompson, J.R., 1978. Nonparametric Probability Density Estimation. Johns Hopkins University Press.
Taylor, B.M., Davies, T.M., Rowlingson, B.S., Diggle, P.J., 2013. lgcp: an R package for inference with spatial and spatio-temporal log-Gaussian Cox processes. J. Stat. Softw. 52, 1–40.
Taylor, B.M., Davies, T.M., Rowlingson, B.S., Diggle, P.J., 2023. lgcp: log-Gaussian Cox process. R package version 2.0, https://CRAN.R-project.org/package=lgcp.
Ugarte, M.D., Adin, A., Goicoa, T., Militino, A.F., 2014. On fitting spatio-temporal disease mapping models using approximate Bayesian inference. Stat. Methods Med. Res. 23 (6), 507–530.
Valente, F., Laurini, M., 2020. Tornado occurrences in the United States: a spatio-temporal point process approach. Econometrics 8 (2).
Waagepetersen, R., Guan, Y., Jalilian, A., Mateu, J., 2016. Analysis of multispecies point patterns by using multivariate log-Gaussian Cox processes. J. R. Stat. Soc. Ser. C. Appl. Stat. 65 (1), 77–96.
Wang, K., Ling, C., Chen, Y., Zhang, Z., 2023. Spatio-temporal joint modelling on moderate and extreme air pollution in Spain. arXiv preprint arXiv:2302.06059.
Yuan, Y., Bachl, F., Lindgren, F., Borchers, D.L., Illian, J.B., Buckland, S.T., Rue, H., Gerrodette, T., 2017. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. Ann. Appl. Stat. 11 (4), 2270–2297.
Zhang, Z., 2010. stam: spatio-temporal analysis and modelling. R package version 0.1-1, https://CRAN.R-project.org/package=stam.
Zhang, Z., Chen, D., Liu, W., Racine, J.S., Ong, S., Chen, Y., Zhao, G., Jiang, Q., 2011. Nonparametric evaluation of dynamic disease risk: a spatio-temporal kernel approach. PLoS One 6 (3).