Published in Biometrika (2026), asag029
https://doi.org/10.1093/biomet/asag029

Nonparametric estimators over metric graphs

By ALDO CLEMENTE
MOX - Department of Mathematics, Politecnico di Milano,
Piazza Leonardo da Vinci 32, 20133 Milano, Italy
aldo.clemente@polimi.it

ELEONORA ARNONE
Department of Management, Università degli Studi di Torino,
Via Giuseppe Verdi 8, 10124 Torino, Italy
eleonora.arnone@unito.it

JORGE MATEU
Department of Mathematics, Universitat Jaume I,
Avinguda de Vicent Sos Baynat, 12006 Castelló de la Plana, Spain
mateu@uji.es

and LAURA M. SANGALLI*
MOX - Department of Mathematics, Politecnico di Milano,
Piazza Leonardo da Vinci 32, 20133 Milano, Italy
laura.sangalli@polimi.it

Summary

This work discusses a theory of functional spaces over metric graphs, that permits the definition of penalized likelihood methods for data observed over spatial supports that are graphs. Within the considered mathematical framework, we recover classical results in functional analysis, such as a Poincaré-type inequality. This, in turn, enables us to uplift, to the considered setting, the theory of some fundamental penalized likelihood methods. Specifically, we present two important classes of statistical models: nonparametric regression and nonparametric density estimation, here defined for data observed over graphs. We derive theoretical results regarding the well-posedness of the associated estimation problems and the consistency of the estimators. We also demonstrate the performances of the defined estimators with respect to state-of-art alternatives.

Some key words: Smoothing with roughness penalties, nonparametric regression, nonparametric density estimation

1. Introduction

Since the seminal work of Wahba (1990), penalized spline and spline-like estimators have become a standard general-purpose method when dealing with functional estimation. Besides the classical framework of nonparametric regression, explored, e.g., in Craven & Wahba (1978/79) and Eilers & Marx (1996), penalized splines have been widely used in other settings, such as semiparametric and generalized regression (Green & Silverman, 1994; Ruppert et al., 2003), density estimation (Silverman, 1982; Gu & Qiu, 1993), quantile regression (Koenker et al., 1994) and generalized additive models (Hastie & Tibshirani, 1986, 1990; Wood, 2017). The extension of spline estimators from one-dimensional to multidimensional domains has been extensively explored since the 1990s. The most classical approaches are based on tensor-product splines or thin plate splines, discussed, for instance, by Cox (1984), Duchon (1977) and Wahba (1990), respectively. More recently, penalized estimators have been extended to deal with planar domains with non-convex geometries. Ramsay (2002) has introduced FELsplines, Wood et al. (2008) have proposed soap film smoothing, Sangalli et al. (2013) and Azzimonti et al. (2015) have introduced spatial regression with partial differential equation regularization, Lai & Schumaker (2007) and Lai & Wang (2013) have described splines over triangulations, Wilhelm & Sangalli (2016) have discussed generalized spatial regression, and Ferraccioli et al. (2021) have considered nonparametric density estimation. Curved domains and non-convex three-dimensional domains have also been addressed by, e.g., Wahba (1981); Baramidze et al. (2006); Ettinger et al. (2016); Arnone et al. (2023b).

On a different line of literature, recent years have seen a growing interest in models capable of analyzing data observed over linear networks, such as road or rail networks. Networks have also become a central object of study in machine learning, where both graph kernel methods (Borgwardt & Kriegel, 2005; Shervashidze et al., 2011) and Graph Neural Networks (Scarselli et al., 2008; Wu et al., 2020) have become widely adopted tools. However, in the machine learning literature, the focus is typically on learning from graphs as data structures. In the framework of this paper, instead, the graph is regarded as a physical subset of $\mathbb{R}^d$, such as indeed a road or rail network in $\mathbb{R}^2$, and constitutes the support over which data are observed. Within this latter line of research, several statistical approaches have been recently developed. For instance, Lu et al. (2014) have considered Geographically Weighted Regression, while Ladle et al. (2017) have examined the use of Kriging to model pedestrian activities in non-urban areas, considering network distance instead of the Euclidean distance. It is worth noting that using Euclidean distance on linear networks may result in non-positive covariance matrices and, consequently, invalid covariance models. To address this limitation, Ver Hoef (2018) have introduced a Kriging method specifically designed for data observed over undirected linear networks.

More recently, Anderes et al. (2020) have developed covariance functions on graphs with Euclidean edges, which are graphs whose edges are treated as sets equipped with a coordinate system, allowing points on the graph to be identified either as vertices or as locations along an edge, and Porcu et al. (2023) have proposed a nonseparable spatio-temporal covariance function on the same kind of graphs. Finally Bolin et al. (2024) have used the notion of metric graph (see, e.g., Berkolaiko & Kuchment, 2013) and defined Whittle-Matérn fields as Gaussian random fields on such graphs. Moreover, several authors have proposed models for point patterns observed on linear networks. State-of-the-art approaches to tackling this problem have mainly relied on kernel smoothing methods. In particular, Boots et al. (2009) and Okabe & Sugihara (2012) have presented the Equal Split Discontinuous Kernel Density Estimation as an algorithm assigning values to each part of the network in such a way that the kernel has total mass one. McSwiggan et al. (2017) have considered a Diffusion Estimator, or Heat Kernel Estimator, by exploiting the connection between kernel density estimation and diffusion. Moradi et al. (2019) have proposed a nonparametric adaptive estimator that relies on Voronoi tessellation over the network. Rakshit et al. (2019) have introduced a technique for intensity estimation over linear networks based on two-dimensional kernels. Schneble & Kauermann (2022) have proposed a nonparametric method based on splines. Finally, Yin & Sang (2021) have presented an intensity estimation algorithm based on a graph regularization technique to estimate the spatially varying intensity function.

In this paper, we present a mathematical framework which enables the development of advanced nonparametric regression and maximum likelihood methods, starting from the definition of appropriate functional spaces over metric graphs. Metric graphs, indeed, do not enjoy the analytical characteristics of topological domains in $\mathbb{R}^d$, and the classical definitions of functional spaces do not hold. Building on Berkolaiko & Kuchment (2013), we introduce appropriate functional spaces and differential operators over metric graphs, leveraging the notion of quantum graphs, and provide some essential analytical results in the considered spaces, such as a Poincaré inequality. Other functional analysis results on metric graphs are available, e.g., in Haeseler (2011); Berkolaiko & Kuchment (2013); Bolin et al. (2024). The use of differential operators on graphs and the introduced functional embeddings permit the extensions to graphs of nonparametric and maximum likelihood estimators with roughness penalties, similar to those extensively investigated in the simpler setting of one-dimensional and regular multi-dimensional domains, in the vast literature briefly recalled above. Specifically, we here present nonparametric regression and nonparametric density estimation over graphs, and exploits the considered analytical framework to derive the asymptotic properties of the associated estimators, proving results like those obtained for smoothing splines by Cox & O'Sullivan (1990), for nonparametric regression over $\mathbb{R}^d$ by Van De Geer & Wegkamp (1996) and Geer (2000), for polynomial splines by Huang (2003), and for smoothing spline density estimators by Silverman (1982) and Gu & Qiu (1993). Differently from simpler one-dimensional and regular multidimensional domains, where the solution to the penalized nonparametric estimation problem possesses a closed-form analytical solution, the penalized estimators over graphs do not enjoy a closed-form solution. It is thus necessary to solve the considered estimation problems through appropriate numerical discretization, and we here discuss a discretization based on finite elements over metric graphs (see, e.g., Arioli & Benzi, 2018). The proposed methods are implemented in the library R package `fdaPDE` (Arnone et al., 2025).

2. Metric graphs and Sobolev spaces on graphs

2.1. *Metric graphs*

We consider a finite graph G as a subset of the space $\mathbb{R}^d$, made by a set of points, connected by piecewise differentiable curves. In particular, let $V = \{v_i\}_{i=1}^N \subset \mathbb{R}^d$ be the set of N points, called *vertices* of the graph, and let $E = \{e_i\}_{i=1}^K \subset V \times V$ be the set of K couples of connected vertices, called *edges*, where $e = (u_e, v_e)$ denotes the edge connecting the two vertices $u_e, v_e \in V$. We consider undirected graphs, therefore the couple (u_e, v_e) is equivalent to the couple (v_e, u_e). We say that the edge e is incident to $v \in V$ if $v = v_e$ or $v = u_e$. Let E_v denote the set of the edges incident to vertex v; the degree of a vertex v, denoted by $deg(v)$, is the cardinality of E_v. For each edge $e \in E$, let $r_e : [0, \ell_e] \to \mathbb{R}^d$ be the arc-length parametrization of the piecewise differentiable curve associated with e, where ℓ_e denotes the length of this curve, $r_e(0) = u_e$ and $r_e(\ell_e) = v_e$. We here consider the graph G as the subset of $\mathbb{R}^d$ composed by all the vertices V, together with all the points in the image set of the piecewise differentiable curves r_e, for all $e \in E$. We denote by V_I the set of interior vertices, i.e., vertices having $deg(v) > 1$, and by V_B the set of boundary vertices, i.e., vertices having $deg(v) = 1$, so that $V = V_B \cup V_I$. We assume $V_B \neq \emptyset$, i.e., we consider graphs that have at least one boundary node, as common in real data applications. Figure 1 shows an example of a graph in $\mathbb{R}^2$.

In order to define a distance on the graph G, we start defining the distance between any two points x_1, x_2 on the same edge e, i.e., belonging to the image set of the curve r_e. For such x_1, x_2, there exist $s_1, s_2 \in [0, \ell_e]$ such that $x_1 = r_e(s_1)$ and $x_2 = r_e(s_2)$. Since r_e is the arc-length parametrization, we can simply define the distance as $d_e(x_1, x_2) := |s_1 - s_2|$. We point out that any other parametrization can be used, and in that case the distance is defined as

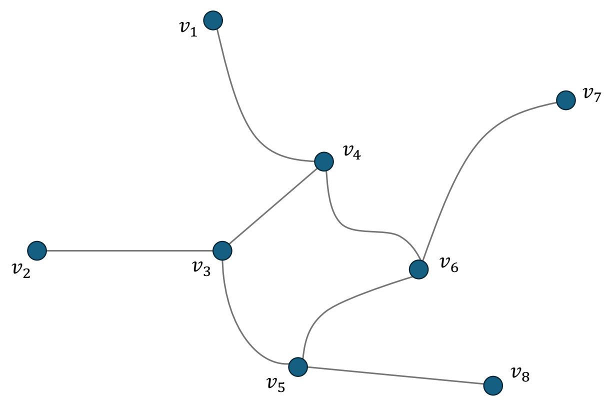

Fig. 1: Example of a graph in $\mathbb{R}^2$. The graph G is the subset of $\mathbb{R}^2$ composed by all the vertices $V = \{v_1, \ldots, v_8\}$, together with all the points in the image set of the piecewise differentiable curves r_e for all edges $e \in E$, indicated by grey lines. For the graph in the figure, we have internal vertices $V_I = \{v_3, v_4, v_5, v_6\}$ and boundary vertices $V_B = \{v_1, v_2, v_7, v_8\}$.

$d_e(x_1, x_2) := \int_{s_1}^{s_2} \|r_e'(t)\|_2 \, dt$. Using the distance d_e, we can now define the shortest path distance d_G on G as follows: for any $x, y \in G$, $d_G(x, y)$ is the minimum length, among the lengths of all the paths connecting x and y. Endowed with this distance, (G, d_G) is a metric space, called *metric graph*.

2.2. *Functional spaces on graphs*

In nonparametric regression and density estimation problems, the estimator is a function belonging to an appropriate functional space. In the setting here considered, this function is defined on a graph. It is therefore essential to define appropriate spaces of functions over graphs, in which to embed the considered estimation problems. Thanks to the metric space structure introduced in Section 2.1, we can now define the space of continuous functions over the graph $C^0(G)$ in the usual way. In the following we identify each edge $e \in E$ with the associated interval $[0, \ell_e]$. Thus, for any edge $e \in E$ of the graph G, we denote with $L^p(e)$ the standard Lebesgue L^p space over $[0, \ell_e]$, that is, $L^p(e) = \left\{ f : e \to \mathbb{R} \ : \left(\int_e |f|^p \right)^{\frac{1}{p}} < \infty \right\}$, equipped with the norm $\|\cdot\|_{L^p(e)}$. Denote by

$$H^s(e) = \{ f : e \to \mathbb{R} \ : f^{(k)} \in L^2(e) \ \forall k = 0, \ldots, s \},$$

the Sobolev space on e, equipped with the norm $\|\cdot\|_{H^s(e)}$ induced by the standard inner product $(f, g)_{H^s(e)} = \sum_{k=0}^{s} \left(f^{(k)}, g^{(k)} \right)_{L^2(e)}$. We recall that the spaces $(L^2(e), (\cdot, \cdot)_{L^2(e)})$ and $(H^s(e), (\cdot, \cdot)_{H^s(e)})$ are Hilbert spaces when equipped with the inner products $(\cdot, \cdot)_{L^2(e)}$ and $(\cdot, \cdot)_{H^s(e)}$ respectively (see, e.g., Brezis & Brézis, 2011).

We can now introduce the appropriate functional spaces defined over the graph. We define the L^p space over the graph as follows.

Definition 1. *Define the functional space*

$$L^p(G) := \{ \phi : G \to \mathbb{R} \ \textit{such that} \ \phi|_e \in L^p(e), \ \forall e \in E \} = \bigoplus_{e \in E} L^p(e)$$

which consists of the functions that are p-integrable over each edge $e \in E$. The space $L^p(G)$ is equipped with the following norm:

$$\|f\|_{L^p(G)}^p := \sum_{e \in E} \|f\|_{L^p(e)}^p \,. \tag{1}$$

The introduction of the Sobolev space H^s over graphs, for general $s \in \mathbb{N}$, is not straightforward. We start by recalling the definition of the space $\tilde{H}^s$ provided by Berkolaiko & Kuchment (2013).

Definition 2. *Define the functional space*

$$\tilde{H}^s(G) := \bigoplus_{e \in E} H^s(e)$$

which consists of the functions whose restriction to $e \in E$ belongs to $H^s(e)$, $\forall e \in E$. The space $\tilde{H}^s(G)$ is equipped with the following norm:

$$\|f\|_{\tilde{H}^s(G)}^2 := \sum_{e \in E} \|f\|_{H^s(e)}^2 \,. \tag{2}$$

The spaces $L^p(G)$ and $\tilde{H}^s(G)$ are defined as the direct sum of the local spaces over the edges. Similarly, we can introduce differential operators acting on functions belonging to such spaces. For instance, we can define the first and second weak derivatives of a function ϕ over the graph, as those functions that, once restricted to any edge of the graph, are the first and second weak derivatives of ϕ. Specifically, for any point x on the edge e, they are

$$\mathcal{D} : \phi(x) \to \frac{d\phi}{dx_e}(x) \qquad \text{and} \qquad \mathcal{H} : \phi(x) \to \frac{d^2\phi}{dx_e^2}(x) \tag{3}$$

where $\frac{d}{dx_e}$ denotes the derivative along that edge. Note that the operator $\mathcal{D}$ is well defined from $\tilde{H}^1(G)$ to $L^2(G)$, while $\mathcal{H}$ is well defined from $\tilde{H}^2(G)$ to $L^2(G)$. Differential operators are particularly useful in nonparametric statistical modeling, since they are commonly employed to measure the regularity of the function.

It should now be noted that the spaces $\tilde{H}^s$ do not possess the properties of Sobolev spaces over the whole graph, but only locally on the edges. In particular, they do not satisfy the classical Sobolev embedding property $\tilde{H}^s \subset C^0(G)$ (see, e.g., Brezis & Brézis, 2011), due to lack of sufficient regularity at the graph vertices. To define global Sobolev spaces over (G, d_G) it is thus necessary to introduce appropriate regularity conditions at the vertices. Such *vertex conditions*, C_V, may concern the value or the function or of its derivative at vertices $v \in V$. A metric graph, equipped with a differential operator O, and accompanied by vertex conditions C_V, is called a *quantum graph*. Hence, a quantum graph is the triple:

$$((G, d_G), O, C_V) \,.$$

The notion of quantum graph is inherently linked to the definition of a specific Sobolev space on the graph. Consider, for example, the first weak derivative $\mathcal{D}$, defined in (3), and consider, as vertex conditions, the continuity at the vertices:

$$C_V^0(\phi) : \;\; \phi \text{ is continuous } \forall v \in V.$$

The quantum graph $\left((G, d_G), \mathcal{D}, C_V^0\right)$ characterizes the space $H^1(G)$, defined as

$$H^1(G) := \left\{\phi \in \tilde{H}^1(G) : \phi \;\; \text{satisfies} \;\; C_V^0\right\} .$$

This space satisfies the Sobolev embedding $H^1(G) \subset C^0(G)$, since it is naturally satisfied along each edge, and it also holds at vertices by definition.

Consider now as differential operator the second weak derivative $\mathcal{H} : \tilde{H}^2(G) \to L^2(G)$ defined in (3). Several vertex conditions can be considered in order to define an H^2-Sobolev space over the graph. In this work, we consider the so-called Neumann-Kirchoff conditions,

$$NK_V(\phi) : \ \phi \text{ satisfies } C_V^0 \text{ and } \sum_{e \in E_\ell} \frac{d\phi}{dx_e}(v_\ell) = 0 \qquad \forall v_\ell \in V$$

and the Dirichlet boundary conditions

$$D_{V_D,\gamma}(\phi) : \ \phi(v_\ell) = \gamma(v_\ell), \qquad v_\ell \in V_D$$

where $V_D \subseteq V_B$ is a subset of boundary vertices and $\gamma(\cdot)$ is a function on V_D. The Neumann-Kirchoff conditions arise naturally by considering the operator $\mathcal{H}$ in (3), and they ensure continuity and the conservation of mass at graph vertices, while the Dirichlet conditions impose the value at the boundary vertices in V_D and are particularly interesting when problem-specific information is available concerning the behavior of the function at boundary vertices. Other possible types of vertex conditions may involve, for instance, the non-homogeneous Neumann-Kirchhoff conditions or the fully Dirichlet conditions, that impose the values of the function also at the interior vertices (see, e.g., Berkolaiko & Kuchment, 2013).

Specifically, we here consider two functional spaces, associated with the quantum graphs $((G, d_G), \mathcal{H}, NK_V)$ and $((G, d_G), \mathcal{H}, NK_{V \setminus V_D} \cap D_{V_D})$, that provide the natural functional embeddings for the nonparametric estimation problems that will be considered in the subsequent sections.

Definition 3. *We define the space $H^2_{NK}(G)$ as*

$$H^2_{NK}(G) := \left\{ \phi \in \tilde{H}^2(G), \ \phi \textit{ satisfies } NK_V \right\}.$$

Definition 4. *We define the space $H^2_{NK,\gamma}(G)$ as*

$$H^2_{NK,\gamma}(G) := \{ \phi \in \tilde{H}^2(G), \ \phi \textit{ satisfies } NK_{V \setminus V_D} \textit{ and } D_{V_D,\gamma} \}.$$

Both $H^2_{NK}(G)$ and $H^2_{NK,\gamma}(G)$ are Sobolev spaces equipped with the norm $\|\cdot\|_{\tilde{H}^2(G)}$ defined in (2). Moreover, notice that the point-wise evaluation of functions belonging to the spaces $H^2_{NK}(G)$ or $H^2_{NK,\gamma}(G)$ is well-defined. Indeed, thanks to the Sobolev embedding on each edge $H^2(e) \subset C^0(e)$ and the vertex conditions NK_V, the functions in $H^2_{NK}(G)$ or $H^2_{NK,\gamma}(G)$ are globally continuous, so that $H^2_{NK}(G)$ and $H^2_{NK,\gamma}(G)$ are sub-spaces of $C^0(G)$. This is a crucial feature to enable the definition of appropriate nonparametric statistical models on graphs. In particular, the space $H^2_{NK}(G)$ introduced in Definition 3 provides the functional embedding for the density estimation problem presented in Section 4, and is used to prove the well-posedness of the estimator and its asymptotic properties. Similarly, the space $H^2_{NK,\gamma}(G)$ in Definition 4 provides the functional embedding for the nonparametric regression estimator in Section 3, and it is used to prove its properties.

2.3. *Poincaré inequality in functional spaces over graphs*

The proofs of the theoretical results presented in this section are available in section **??** of the Supplementary Material. We first prove a Poincaré-type inequality, which is useful to establish the consistency of the nonparametric estimators showcased in Sections 3 and 4. We use the following notation:

$$\int_G u := \sum_{e\in E} \int_e u\, dx_e\ , \qquad \left\| \frac{d^k u}{dx^k} \right\|^2_{L^2(G)} := \sum_{e\in E} \left\| \frac{d^k u}{dx_e^k} \right\|^2_{L^2(e)} .$$

We first state the following Hölder inequality on metric graphs.

Lemma 1 (Hölder inequality). *For any $u \in L^p(G)$ and $v \in L^q(G)$, where $p, q \in \mathbb{R}$ are such that $\frac{1}{p} + \frac{1}{q} = 1$, the following inequality holds*

$$\|u\, v\|_{L^1(G)} = \sum_{e\in E} \int_e |u\, v| \le \|u\|_{L^p(G)} \|v\|_{L^q(G)} .$$

As in more classical settings, when $p = q = 2$ we refer to the inequality of Lemma 1 as *Cauchy-Schwarz inequality*. Define the space $H^1_\gamma(G) := \{\phi \in H^1(G), \phi \text{ satisfies } D_{V_D,\gamma}\}$. We can then state the following Poincaré Inequality.

Theorem 1 (Poincaré Inequality). *For any $u \in H^1_0(G)$, there exists a constant $c_p > 0$ such that*

$$\|u\|_{L^2(G)} \le\ c_p \left\| \frac{du}{dx} \right\|_{L^2(G)} .$$

Remark 1. It should be noted that Haeseler (2011) provides an alternative Poincaré inequality over graphs, for functions $u \in H^2_{NK}(G)$ such that $\int_G u = 0$, i.e., in the space of functions with null mean over the graph, whilst Theorem 1 is stated in $H^1_0(G)$ and thus hold in the space $H^2_{NK,0}(G)$, which provides a more natural functional embedding for the estimation problems here considered. Indeed, the space of functions with null mean over graph does not offer an appropriate functional embedding for our nonparametric estimation problems: if we assumed that the nonparametric estimator have a null mean, then we would need to estimate the mean separately. The Poincaré inequality in Theorem 1 and that in Haeseler (2011) can nevertheless be shown to be equivalent, exploiting classical analytical results; see, e.g., Chapter 7 of Salsa (2016).

Additionally, it is possible to prove the equivalence between the norms $\|\cdot\|_{\tilde{H}^2(G)}$ and $\left\|\frac{d^2}{dx^2}\cdot\right\|_{L^2(G)}$ on $H^2_{NK,\gamma}(G)$. In particular, Corollary 1 provides an upper bound for the norm $\|\frac{d}{dx}\cdot\|_{L^2(G)}$ in terms of the norm $\left\|\frac{d^2}{dx^2}\cdot\right\|_{L^2(G)}$, and Corollary 2 establishes the equivalence between the two aforementioned norms.

Corollary 1. *For any $u \in H^2_{NK,0}(G)$, there exists a constant $C > 0$ such that*

$$\|u\|_{\tilde{H}^2(G)} \le C \left\| \frac{d^2 u}{dx^2} \right\|_{L^2(G)} .$$

Corollary 2. *The norms $\|\cdot\|_{\tilde{H}^2(G)}$ and $\left\|\frac{d^2}{dx^2}\cdot\right\|_{L^2(G)}$ are equivalent in the space $H^2_{NK,0}(G)$. In particular, there exists a constant $c_L > 0$ such that the following inequality holds:*

$$c_L \|u\|_{\tilde{H}^2(G)} \le \left\| \frac{d^2 u}{dx^2} \right\|_{L^2(G)} \le \|u\|_{\tilde{H}^2(G)} .$$

Remark 2. Since Theorem 1 also holds for any function $u \in H^2_{NK}(G)$ such that $\int_G u = 0$ it follows that Corollary 1 and Corollary 2 also hold in this case.

3. Nonparametric Regression over Graphs

3.1. *Statistical Model*

This section defines a penalized nonparametric regression for data observed over graphs; the proofs of the theoretical results here presented are available in Section **??** of the Supplementary Material. Consider n locations $\boldsymbol{p}_1, \ldots \boldsymbol{p}_n$ over the graph G. Let $y_i \in \mathbb{R}$ be the value taken by a random variable of interest, observed at points $\boldsymbol{p}_i$. We assume the nonparametric model:

$$y_i = f(\boldsymbol{p}_i) + \varepsilon_i \qquad i = 1, ..., n,$$

where $f \in H^2_{NK,\gamma}(G)$ is an unknown function, and $\varepsilon_1, ..., \varepsilon_n$ are uncorrelated noises with zero mean and finite constant variance σ^2. We consider

$$\hat{f} := \underset{f \in H^2_{NK,\gamma}(G)}{\arg\min} \left\{ \frac{1}{n} \sum_{i=1}^{n} (y_i - f(\boldsymbol{p}_i))^2 + \lambda \left\| \frac{d^2 f}{dx^2} \right\|^2_{L^2(G)} \right\}, \tag{4}$$

where λ is a positive smoothing parameter, that balances the data-adaptation and the smoothness of the estimate. For any function $f \in H^2_{NK,\gamma}(G)$, denote by $\boldsymbol{f}_n = [f(\boldsymbol{p}_1), \ldots, f(\boldsymbol{p}_n)]^\top$ the vector of the evaluations of f at the n locations of the observations. Recall that the point-wise evaluation is well-defined for any function belonging to $H^2_{NK,\gamma}(G)$, since $H^2_{NK,\gamma}(G) \subset C^0(G)$, as commented in Section 2.2. Finally, let $\boldsymbol{y} = [y_1, \ldots, y_n]^\top \in \mathbb{R}^n$ be the vector of the observations. The following theorem states that the estimation problem (4) has a unique solution.

Theorem 2. *The estimator $\hat{f} \in H^2_{NK,\gamma}(G)$, that solves the nonparametric regression problem* (4)*, exists and is unique. Moreover, $\hat{f}$ satisfies*

$$\frac{1}{n} \boldsymbol{u}_n^\top \hat{\boldsymbol{f}}_n + \lambda \sum_{e \in E} \int_e \frac{d^2 \hat{f}}{dx_e^2} \frac{d^2 u}{dx_e^2} = \frac{1}{n} \boldsymbol{u}_n^\top \boldsymbol{y} \quad \forall u \in H^2_{NK,0}(G). \tag{5}$$

Remark 3. Although, for simplicity of exposition, Theorem 2 is presented for functions satisfying Dirichlet boundary conditions, the same result also holds in the case of $f \in H^2_{NK}(G)$, thanks to the characterization of the kernel of the differential operator detailed in the proof of Theorem 6.

Differently from the classical settings of smoothing splines and thin-plate splines, where the nonparametric regression problem possesses a closed form analytical solution, that is an appropriate spline, the estimator $\hat{f}$ in (4) does not enjoy a closed form solution. This issue, coupled with the complexity of graphs with respect to classical topological domains, makes it impossible to derive the consistency of the nonparametric regression estimator on graphs by straightforward extensions of the arguments used, e.g., by Cox (1983, 1984); Cucker & Zhou (2007); Györfi et al. (2002); Huang (2003) to prove the consistency of smoothing splines and thin-plate-splines. Nevertheless, consistency of the estimator in (4) can be established as detailed in the following section.

3.2. *Consistency of Nonparametric Regression Estimator over Graphs*

We consider the estimation problem in (4), denoting the smoothing parameter by λ_n to highlight the dependence on the sample size n. We denote by f_0 the true unknown function and assume $f_0 \in H^2_{NK,\gamma}(G)$. Let $F_n(\boldsymbol{p})$ be the empirical cumulative distribution function that assigns mass n^{-1} to each point $\boldsymbol{p}_i$. Specifically, F_n is the restriction to G of the empirical cumulative distribution function defined over $\mathbb{R}^d$. We make the following assumptions, which are standard in the literature of nonparametric regression over simple one-dimensional or multi-dimensional domains (see, e.g., Cox, 1983, 1984).

Assumption 1. $\{F_n\}$ converges uniformly to a cumulative distribution function F with associated density $f \in C^{\infty}(G)$, with respect to the Lebesgue measure on G, where f is such that, for all $\boldsymbol{p} \in G$, $0 < k_1 \leq f(\boldsymbol{p}) \leq k_2 < \infty$ for some constant k_1, k_2.

Assumption 2. Set λ_n such that $\lim_{n\to\infty} d_n \lambda_n^{-1} = 0$, where $d_n := \sup_{\boldsymbol{p}\in G} |F(\boldsymbol{p}) - F_n(\boldsymbol{p})|$.

The following result extends Lemma 4.2 of Cox (1984) to nonparametric regression estimation on graphs.

THEOREM 3. *Let (G, d_G) be a metric graph. Under Assumptions 1 and 2, $\forall h, g \in H^1(G)$, there exists a constant $c > 0$ such that*

$$\left| \sum_{e\in E} \int_e hg \; d(F - F_n) \right| \leq c \; d_n |E| \; \|h\|_{H^1(G)} \; \|g\|_{H^1(G)}, \tag{6}$$

where $|E|$ denotes the number of edges of G.

Theorem 4 below gives an upper bound for the $L^2(G)$-norm of the bias of the estimator, denoted by $\mathcal{B} := f_0 - \mathbb{E}[\hat{f}]$. Theorem 5 gives an upper bound for the $L^2(G)$-norm of the variance of the estimator. Finally, Corollary 3 states the consistency of the nonparametric regression estimator.

THEOREM 4. *Under Assumptions 1 and 2, for $f_0 \in H^2_{NK,\gamma}(G)$ and n large enough,*

$$\|\mathcal{B}\|^2_{L^2(G)} \leq C\lambda_n.$$

THEOREM 5. *Under Assumptions 1 and 2, for $f_0 \in H^2_{NK,\gamma}(G)$ and n large enough,*

$$Var(\hat{f}) = O\left(\lambda_n^{-1/2} \, \frac{\sigma^2}{n}\right).$$

COROLLARY 3. *Under Assumptions 1 and 2, for $f_0 \in H^2_{NK,\gamma}(G)$, setting $\lambda_n = n^{-2/3}$ we have*

$$MSE_{L^2(G)}(\hat{f}) = \mathbb{E}\left[\|f_0 - \hat{f}\|^2_{L^2(G)}\right] = O\left(n^{-2/3}\right).$$

4. NONPARAMETRIC DENSITY ESTIMATION OVER GRAPHS

4.1. *Statistical Model*

This section introduces a nonparametric density estimation method for point patterns observed over graphs; the proofs of the theoretical results here presented are available in Section **??** of the Supplementary Material. As commented in the Introduction, the study of point patterns on graphs has recently attracted a strong interest in the statistical literature, stimulated by several applications, among which the analysis of road accidents. Building on the analytical framework introduced in Section 2, we extend the nonparametric density estimators proposed by Silverman (1982) and Gu & Qiu (1993), originally developed for simple one- and two-dimensional domains, to the more complex setting of point patterns on graphs. Let f denote an unknown density on the graph G, and let $\boldsymbol{p}_1, ..., \boldsymbol{p}_n$ be n random points over G, independently drawn from the unknown density f. Given the observations $\boldsymbol{p}_1, ..., \boldsymbol{p}_n$, our objective is to estimate the log-density $g = log(f)$, solving the following minimization problem:

$$\hat{g} := \underset{g\in H^2_{NK}(G)}{\arg\min} \left\{ -\frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{p}_i) + \int_G \exp\{g\} + \lambda \left\| \frac{d^2 g}{dx^2} \right\|^2_{L^2(G)} \right\}, \tag{7}$$

where the first term represents the negative log-likelihood, the second term ensures that the function $f = \exp\{g\}$ integrates to one over G, and the third term is the regularization. The following theorem states the well-posedness of this estimation problem.

THEOREM 6. *The density estimation minimization problem* (7) *has a unique minimizer in* $H^2_{NK}(G)$.

In the next section, we establish the consistency of the nonparametric density estimator.

4.2. *Consistency of Nonparametric Density Estimator over Graphs*

Let $g_0 = log(f_0)$ be the true log-density , and assume that $f_0(\boldsymbol{p}) > 0$ for all $\boldsymbol{p} \in G$. Let g^* be an approximation of $\hat{g}$, which is the minimizer of

$$L_*(g) = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{p}_i) + 1 + \mu_{g_0}(g) + \frac{1}{2}Var_{g_0}(g - g_0) + \lambda\left\|\frac{d^2 g}{dx^2}\right\|^2_{L^2(G)},$$

where $\mu_{g_0}(g) = \int_G g\exp\{g_0\}$ and Var_{g_0} denotes the variance of $g_0\,(X)$ when X has log-density g_0. Let $D_sKL(\cdot,\cdot)$ denote the symmetrized Kullback-Leibler distance. The functional L_* and its minimizer g^* are introduced to decompose the distance between $\hat{g}$ and g_0 in two parts, namely $D_sKL(\hat{g}, g^*)$ and $D_sKL(g^*, g_0)$, whose asymptotic properties are easier to derive compared to those of $D_sKL(\hat{g}, g_0)$. We consider the following assumptions on g_0 and g_*, that extend to the setting here considered the analogous assumptions made by Silverman (1982) and Gu & Qiu (1993) for density estimators over regular one and two-dimensional domains.

Assumption 3. The true log-density g_0 is bounded above and below, and is such that $\left\|\frac{d^2 g_0}{dx^2}\right\|_{L^2(G)} < \infty$.

Assumption 4. Let B_0 be a convex set centered at g_0, containing $\hat{g}$ and g_*. There exists a positive constant c such that $c\,Var_{g_0} < Var_g$ uniformly with respect to g.

In particular, Assumption 3 ensures that the weighted $L^2(G)$-norm, where the true density $f_0 = e^{g_0}$ is used as the weight function, is equivalent to the standard $L^2(G)$-norm. Assumption 4 requires Assumption 3 to hold also for functions near g_0, and it is satisfied whenever the elements of B_0 are bounded from above and below.

THEOREM 7. *Under Assumptions 3 and 4, as* $\lambda \to 0$ *and* $n\lambda^{1/2} \to \infty$ *the estimator* $\hat{g}$*, that solves* (7)*, is consistent and*

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

We recall that the symmetrized Kullback-Leibler distance also governs other metrics commonly used for density functions, such as the Hellinger and total variation distances (see, e.g., Pollard, 2002). As a result, the proposed estimator is also consistent with respect to such distances.

5. DISCRETIZATION VIA FINITE ELEMENTS OVER GRAPHS

The nonparametric regression estimator in (4) and the nonparametric density estimator over graphs in (7) do not enjoy closed-form analytical solutions and must therefore be computed through numerical discretization. In this section, we introduce a discretization strategy based on the Finite Elements method over graphs (Arioli & Benzi, 2018). We discretize the graph G by subdividing the edges of the original graph G into smaller edges, setting along the graph the nodes $V_h = \{\xi_1, \ldots, \xi_{N_h}\}$, where $V \subset V_h$. Let $G_h = (V_h, E_h)$ denote the discretization of

G, where $V_h = \{\xi_1, \ldots, \xi_{N_h}\}$ is the set of mesh nodes, and $E_h = \{e_1, \ldots, e_{K_h}\}$ is the set of edges. The finite element space of order r is defined as the space of functions $C^0(G_h)$ that are polynomials of order r once restricted to any edge e_h. In particular, in the numerical studies and test-bed applications in Sections 6 and 7, we use linear finite elements, that are finite elements of order 1. It is possible to define a set of N_h basis functions $\psi_1, \ldots, \psi_{N_h}$ associated with the nodes $V_h = \{\xi_1, \ldots, \xi_{N_h}\}$ spanning the finite element space of order 1. Let $\boldsymbol{\psi} = (\psi_1, ..., \psi_{N_h})$ be the N_h vector of the basis functions. Any function u_h in the finite element space can be written as the linear combination of the basis functions, that is, $u_h(\boldsymbol{p}) = \sum_{j=1}^{N_h} u_j \; \psi_j(\boldsymbol{p}) = \boldsymbol{u}^\top \boldsymbol{\psi}(\boldsymbol{p})$, where $\boldsymbol{u}$ is the vector of coefficients of the basis expansion. Such a vector of coefficients coincides with the vector of the evaluations of the function at the N_h nodes of the mesh, i.e., $\boldsymbol{u} = (u(\xi_1), \ldots, u(\xi_{N_h}))$. Denote by Ψ the $n \times N_h$ matrix, with its entries representing the evaluations of the N_h basis functions at the n data locations, that is $\Psi_{ij} := \psi_j(\boldsymbol{p}_i)$. Finally, let R_0 and R_1 denote the mass and stiffness matrices, defined as the following $N_h \times N_h$ matrices:

$$(R_0)_{ij} := \sum_{e_h \in E_h} \int_{e_h} \psi_j \psi_i \qquad (R_1)_{ij} := \sum_{e_h \in E_h} \int_{e_h} \frac{d\psi_j}{dx_{e_h}} \frac{d\psi_i}{dx_{e_h}}.$$

The matrices Ψ, R_0, and R_1 are employed in the discretization of estimation problems considered in Sections 3 and 4. In particular, the nonparametric regression problem discussed in Section 3 can be solved via a so-called mixed finite element approach, as detailed in Azzimonti et al. (2014) for the case of nonparametric regression over two dimensional domains, leading to the nonparametric regression estimator $\hat{\boldsymbol{f}}^\top \boldsymbol{\psi}(\boldsymbol{p})$ where $\hat{\boldsymbol{f}} = (\Psi^\top \Psi + \lambda R_1^\top R_0^{-1} R_1)^{-1} \Psi^\top \boldsymbol{y}$. The solution is obtained efficiently without explicitly computing R_0^{-1} or $(\Psi^\top \Psi + \lambda R_1^\top R_0^{-1} R_1)^{-1}$. As detailed in Arnone et al. (2023a) for the estimators over multidimensional domains, the computational cost depends both on the number of observations n and on the mesh size N_h. In particular, for a fixed mesh (i.e., fixed N_h), the computational complexity is of order $O(n^{1/2})$. On the other hand, for a fixed sample size n, the computational complexity is of order $O(N_h)$. The convergence properties of the finite element regression estimator can be derived by appropriately combining the results in Arioli & Benzi (2018), concerning the bounds for the error of the finite element solution of a partial differential equation defined over a graph, and those in Azzimonti et al. (2014), concerning the discrete regression estimator defined over two-dimensional domains.

Concerning the nonparametric density estimation problem discussed in Section 4, the discrete counterpart of the estimation problem (7) is given by

$$L_h(\mathbf{g}) = -\mathbf{1}^\top \Psi \mathbf{g} + n \sum_{e_h \in G_h} \mathbf{w}^\top \exp\{\Psi_h \mathbf{g}\} + \lambda \mathbf{g}^\top R_1 R_0^{-1} R_1 \mathbf{g}. \tag{8}$$

Here, the first term is the evaluation of the negative log-likelihood; the second term approximates the integral of the density that appears as the second term in (7) via an opportune quadrature rule, where Ψ_h is the matrix of the basis functions evaluated at the quadrature nodes and $\boldsymbol{w}$ is the weights vector of the considered quadrature rule; finally, the last term is the discretization of the penalty term in (7), obtained in analogy with the mixed finite element solution to the nonparametric regression problem. The minimization of the discrete problem (8) can be performed using classical steepest descent algorithms, such as the gradient descent algorithm or the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Such approaches are guaranteed to converge, thanks to the strict convexity of the functional. We remark that, for both models, the graph complexity does not pose any theoretical issue, since the well-posedness of the estimators is ensured by the presence of the penalty term. Of course, higher graph complexities imply a higher mesh sizes, and thus higher computational cost. In that respect, it should be noted that the mesh should be chosen fine enough to capture the possibly localized features of the signal; however, overly fine meshes, i.e., meshes with a resolution much higher than the data resolution, are unnecessary, and lead to increased computational cost for the same level of accuracy of the estimates (see, e.g., Azzimonti et al., 2014, for a detailed study in the context of estimators over a two-dimensional domain). Data-driven meshes, with finer resolution where data locations are denser, can be used to save computational costs, while preserving accuracy, as illustrated by, e.g., Tomasetto et al. (2024) in the context of nonparametric regression over two-dimensional domains, and by Ferraccioli et al. (2021) in the context of nonparametric density estimation over two-dimensional domains. Data-driven meshes are a standard approach to ensure good asymptotic and inferential properties of the estimators, also in the context of classical functional data analysis problems (see, e.g., Ramsay & Silverman, 2005). For both models, the value of the smoothing parameter λ may be selected using cross-validation. In particular, for the regression model, we consider Generalized Cross Validation (see, e.g., Sangalli et al., 2013), while for the density estimation model, we consider K-fold validation (see, e.g., Ferraccioli et al., 2021). This ensures empirical robustness, as shown in extensive simulation studies for estimators over two and three-dimensional domains; see, e.g., Sangalli et al. (2013); Arnone et al. (2023a); Ferraccioli et al. (2021). The implementation of the proposed methods, for both first-order and second-order finite elements, is available in the R package `fdaPDE` (Arnone et al., 2025), together with utilities to refine the mesh.

6. Simulation studies

6.1. *Simulation 1: nonparametric regression - clustered locations and Whittle-Matérn noise*

This section presents a simulation study dealing with a nonparametric regression problem over the `simplenet` network available with the R package `spatstat` (Baddeley et al., 2015); see Figure 3. The graph is discretized into a mesh consisting of 123 nodes and 123 equally spaced edges; this mesh is used by all the competing methods presented below. We consider the test

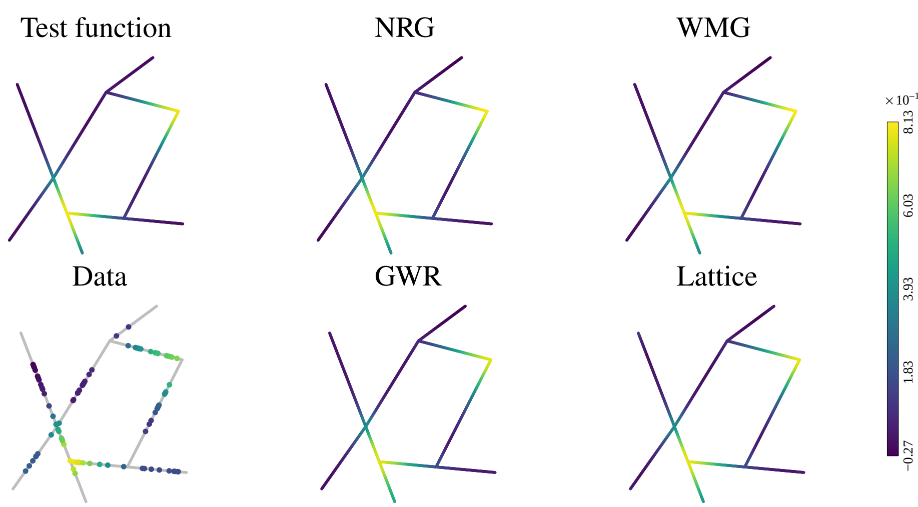

Fig. 2: Simulation 1 (Section 6.1): nonparametric regression problem - clustered locations and Whittle-Matérn noise. Top left: test function. Bottom left: data obtained from the test function in the top left panel, evaluated at locations clustered along the graph, with the addition of Whittle-Matern noise. Center and right: mean estimates obtained by the competing methods over 30 simulation repetitions with sample size $n = 100$.

function shown in the top left panel of Figure 2, and sample the data at n locations, clustered along the network, and add correlated noise, generated as a Whittle-Matérn field on the graph using the R package `MetricGraph` (Bolin et al., 2023). Data generation (both locations and values) is repeated 30 times, for different sample sizes $n = 100, 150, 250, 500$. We compare the Nonparametric Regression estimator over Graph (NRG), presented in Section 3 and implemented in the R package `fdaPDE` (Arnone et al., 2025), to three competing methods available in the literature. The first competitor is the Whittle-Matérn field (WMG), introduced by Bolin et al. (2024), that is based on a stochastic partial differential approach and constitutes an extension to metric graphs of the Gaussian fields with Matérn covariance function originally introduced by Lindgren et al. (2011). This method is implemented in the R package `MetricGraph` (Bolin et al., 2023). The second competitor we consider is the Geographically Weighted Regression (GWR), which consists of local techniques for exploring spatial heterogeneity in data relationships, by assuming spatial variation of the regression coefficients; see Fotheringham et al. (2003). The implementation for linear networks is available in the R package `GWmodel` (Gollini et al., 2015), and uses the shortest path distance on the graph, that we have formally defined in Section 2. The method considers a fixed kernel bandwidth, selected through minimization of AIC_c index. The third method belongs to the class of Lattice-based models (Lattice), that use diffusion kernels to estimate smooth regression functions defined on complicated domains, see McIntyre & Barry (2018). Such an approach can be applied to any situation where it is possible to represent the spatial region, over which the estimator is defined, by a set of nodes connected by neighboring relationships. The implementation for the linear network is based on the R package `latticeDensity`(Barry, 2021), using the functions provided in the supplementary material of Barry & McIntyre (2020); the adjacency matrix is computed assuming an isotropic model between neighboring nodes. The central and right column panels of Figure 2 show the mean estimates provided by the competing methods, over 30 simulated repetitions with 100 observations. From a qualitative point of view, it seems that all the competing methods are capable of capturing the main features of the signal. A quantitative comparison of the estimates provided by the competing methods is then performed on the basis of the Root Mean Square Error (RMSE), computed on a fine grid of points along the network. For $n = 100$, the mean RMSEs are 0.0062 for the proposed NRG method (sd = 0.0026), 0.0062 for the WMG method (sd = 0.0025), 0.0103 for the GWR method (sd = 0.0125), and 0.0084 for the Lattice method (sd = 0.0025). These results indicate that the proposed NRG method, presented in Section 3, and the WMG method, introduced by Bolin et al. (2024), provide the most accurate estimates when the number of observations is low. As the number of observations increases, the other competing methods tend to yield similar RMSE values, as shown by the boxplots of RMSE in the left panel of Figure **??**; see Section **??** of the Supplementary Material. In addition, Section **??** of the Supplementary Material presents a further simulation study conducted under uncorrelated noise. R scripts reproducing all simulation studies, including that presented in the next section, as well as the test-bed applications in Section 7, are available in the GitHub repository: https://github.com/fdaPDE/case-studies.

6.2. *Simulation study 2: density estimation*

This section presents a simulation study dealing with a density estimation problem. We generate samples of independent and identically distributed observations, from the true density, on the `simplenet` graph, shown in the top left panel of Figure 3. We consider different sample sizes, $n = 50, 100, 150, 250$, and repeat the data generation 30 times for each sample size n.

We compare the nonparametric Density Estimator on Graphs (DEG), presented in Section 4 and implemented in the R package `fdaPDE` (Arnone et al., 2025), to three competing methods available in the literature. The first competing method is Kernel Density Estimation based on

n	DEG	KDE-HEAT	KDE-2D	VORONOI
50	**0.02838** (0.01353)	0.03346 (0.00969)	0.04037 (0.00915)	0.14594 (0.14230)
100	**0.01427** (0.00507)	0.02179 (0.00420)	0.02704 (0.00405)	0.15331 (0.32537)
150	**0.01368** (0.00527)	0.01903 (0.00473)	0.02273 (0.00463)	0.13434 (0.1946)
250	**0.00795** (0.00252)	0.01395 (0.00282)	0.01689 (0.00268)	0.42064 (1.16095)

Table 1: The table reports the mean (standard deviation) of the Root Mean Square Error over 30 simulation repetitions of Simulation 2 (Section 6.2), for different sample sizes. Same methods as in Figure 3. In bold, the smallest RMSE.

the heat equation on the network (KDE-HEAT); see, e.g., McSwiggan et al. (2017). The second competing method is Kernel Density Estimation based on a two-dimensional kernel (KDE-2D); see, e.g., Rakshit et al. (2019). The last competing method is the nonparametric adaptive estimation based on Dirichlet-Voronoi tessellation over the network (VORONOI), introduced by Moradi et al. (2019). The implementation of these three methods is available in the R package `spatstat` (Baddeley et al., 2015). Figure 3 shows, in the top left panel, the true density, in the bottom left panel, a point pattern of size $n = 100$ from this density, and, in the center and right panels, the mean density estimates provided by the competing methods over 30 simulation repetitions. From a qualitative point of view, the VORONOI provides rougher estimates than the competitors. We quantify the estimation error via the RMSE, computed on a fine grid of points over the network. Table 1 reports the mean (standard deviation) of the RMSE over 30 simulation repetitions, for each sample size n. The nonparametric Density Estimator on Graphs, presented in Section 4, provides the best estimates under all settings.

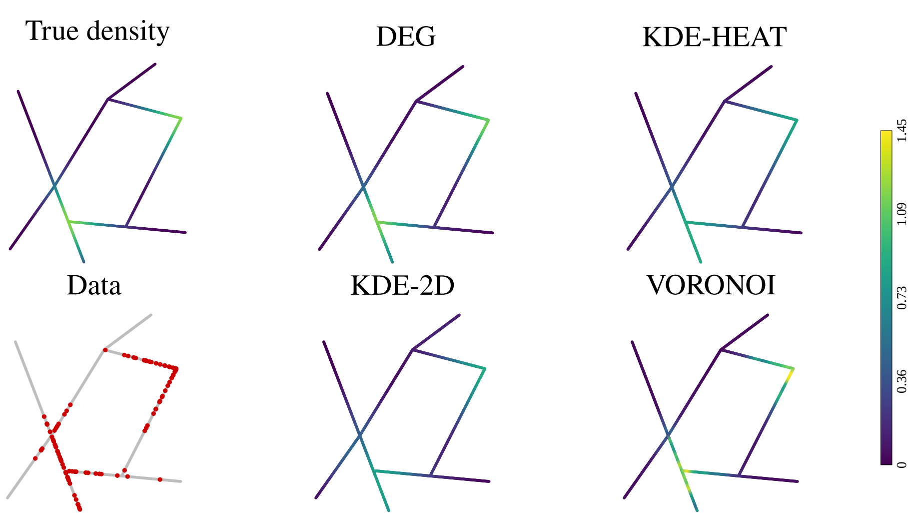

Fig. 3: Simulation 2 (Section 6.2): density estimation problem. Top left: true density; Bottom left: point pattern generated sampling independent observations from the density on the top left panel. Center and right: mean density estimates provided by the competing methods over 30 simulation repetitions with sample size $n = 100$.

7. Applications

7.1. *Test-bed regression problem: London House Price*

In this section, we illustrate a semiparametric regression problem, concerning the London House Prices dataset (see, e.g., Lu et al., 2014). This dataset consists of the selling prices of 1601 properties in London (UK), which have been sold during 2001, together with some characteristics of the properties and neighborhood. Specifically, we consider as the response variable the logarithm of the overall price of the house, and as explanatory variables: the size of the property, the presence of at least two bathrooms, and the percentage of the workforce in professional or managerial occupations in the census enumeration in which the house is located. The nonparametric regression estimator considered in this work can indeed be extended to include space-varying covariates, by considering a semiparametric structure, similar to that described in Sangalli et al. (2013) for spatial regression over simpler two-dimensional domains. The left panel of Figure 4 shows the locations of these properties, with the colour of the point marker referring to the selling price of the house (in thousands of sterlings), and, in the right panel, a zoomed-in view of the road network in the center of London, highlighting the complexity of the spatial support, which comprises the whole municipality. Specifically, the graph contains approximately 75k nodes and 96k edges. Such dense spatial support presents a more challenging computational problem than those addressed in the simulation studies in Section 6. Regarding the competing method considered in this application, we compare the GWR estimates in Lu et al. (2014), based on the network distance metric, with those obtained by the semiparametric version of the spatial regressor discussed in Section 3 of this work. We do not consider the `Lattice` method due to its similar performance to the `GWR` method in the simulation studies presented in Section 6. Additionally, we were unable to include a comparison with the `WMG` method proposed by Bolin et al. (2024) because initializing the data structure that stores the graph, using the `MetricGraph` package, requires more RAM than was available on the machine used for both the simulations and the application studies (a standard laptop with Intel Core i7-8565U, 1.80GHz, quad-core, 16 GB RAM). The smoothing parameter of the proposed method is selected through generalized cross-validation. The GWR model is calibrated using fixed kernel bandwidths, optimally selected via a minimized AIC_c approach; see Lu et al. (2014) for details. Both models estimate a positive effect of the considered covariates on the house prices, aligning with economic intuition. The two methods are compared on the basis of RMSE, computed at test locations, on a 10-fold cross-validation scheme. The mean cross-validated RMSEs are 0.0426 for the proposed method (sd 0.0051) and 0.0794 for the GWR method (sd 0.0085), indicating an advantage of the proposed estimator also in this semiparametric setting.

7.2. *Test-bed density estimation problem: Chicago Crimes*

In this section, we provide a test-bed application of the considered Density Estimator over Graphs (DEG), considering the Chicago crimes dataset, available in the R package `spatstat`, and considered, e.g., by Rakshit et al. (2019); Yin & Sang (2021). This dataset records the locations (nearest street address) of crimes reported in the neighborhood of the University of Chicago, in the city of Chicago (Illinois, USA), between 25 April 2002 and 8 May 2002. The top left panel of Figure 5 shows the locations of the crimes and the road network. We compare the estimate provided by the proposed estimator, DEG, with those obtained by the competing techniques KDE-HEAT and KDE-2D, described in Section 6.2. Figure 5 shows the density estimates provided by the competing methods; the proposed DEG method appears to capture the peaks in the density distribution better than the kernel-based alternatives, which appear to suffer from limited data availability. We quantitatively compare the estimates in 10-fold cross-

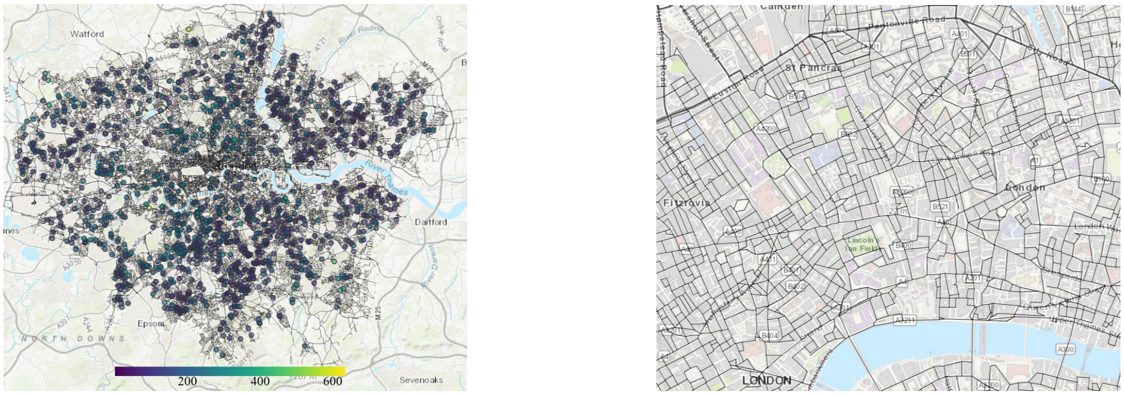

Fig. 4: London House Price (Section 7.1). On the left, the response variable: the price of the houses sold during 2001. On the right, a zoomed-in view of the road-network in the city center of London.

validation, computing the error $R = \int_G (\hat{f}^{-[k]})^2 - \frac{2}{\#x^{[k]}} \sum_{i\in[k]} \hat{f}^{-[k]}(x^{[k]})$, where $\hat{f}^{-[k]}$ is the density estimated without considering the k-th fold, $x^{[k]}$ is the subset of observations of the k-th fold and $\#x^{[k]}$ is its cardinality. We point out that this error, considered among others by, e.g., Marron (1987), can take negative values, with smaller values indicating better results. The 10-fold cross-validation errors of the competing methods are -0.1260 for DEG (sd 0.0229), -0.1084 for KDE-HEAT (sd 0.0123), and -0.1017 for KDE-2D (sd 0.0110), highlighting the superiority of the proposed estimator with respect to the alternatives also in this test-bed application.

8. Discussion

In this work, we have extended to graphs two fundamental nonparametric estimation problems, nonparametric regression and nonparametric density estimation, classically investigated for data observed over simply connected open domains of $\mathbb{R}^d$. We have specifically focused on the properties of the infinite-dimensional estimators, considering appropriate functional embeddings. Following the lines of Ferraccioli et al. (2022), Arnone et al. (2023b), and Begu et al. (2024), one could additionally derive the properties of the corresponding discrete estimators, defined in Section 5. In particular, using the arguments in Ferraccioli et al. (2022) and Arnone et al. (2023b), it is possible to prove the asymptotic normality of the discrete regression estimator over graphs. This justifies the use in this context of Wald-type tests, and other nonparametric inferential tools, presented by Ferraccioli et al. (2022) and Cavazzutti et al. (2024). Regarding the density estimation problem, instead, uncertainty quantification can be derived using a Bayesian approach, as detailed in Begu et al. (2024) for the density estimator over two-dimensional domains.

A challenging direction for future developments concerns nonparametric regression problems featuring penalization terms that involve the generic k-th derivative, and the study of the asymptotic properties of the associated estimators, as done in classical contexts by, e.g., Cox (1983), Cox (1984) and Stone (1982). However, such a study would require the definition of Sobolev spaces of higher order, with appropriate vertex conditions, and would entail various analytical complexities raised by the non-trivial forms of the regularizing terms, and of the geometry of the graph, which prevents direct extension of the arguments used by Cox and Stone. Another possible extension involves incorporating physical knowledge on the problem under study, such as flow

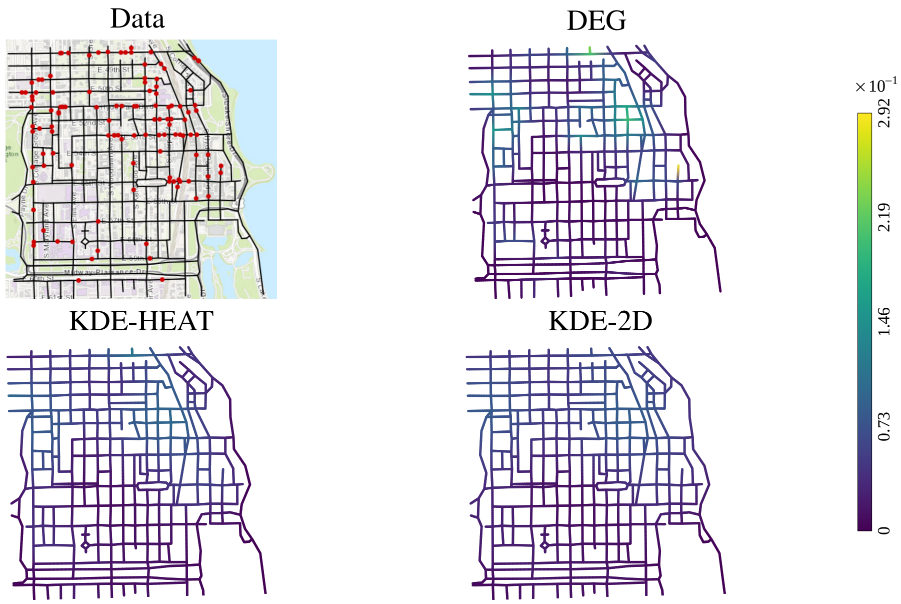

Fig. 5: Chicago Crimes (Section 7.2). The upper-left panel shows the locations of crimes recorded over the Chicago road network, between April 25, 2002, and May 8, 2002, in the neighborhood of the University of Chicago (Illinois, USA). The remaining panels display the density estimates provided by the competing methods: the proposed nonparametric Density Estimation on Graphs (DEG), presented in Section 4; Kernel Density Estimation based on the heat equation on the network (KDE-HEAT); Kernel Density Estimation based on a two-dimensional kernel (KDE-2D).

along the graph. This extension would be of high interest for the modeling of data observed over river networks, offering an alternative to kriging over directed graphs, presented in Ver Hoef et al. (2006), Ver Hoef & Peterson (2010), Ver Hoef et al. (2019) and Barbi et al. (2023).

Finally, we point out that, using the results of this work, it is possible to define broad classes of nonparametric and semiparametric methods for data observed on graphs. This includes, for instance, nonparametric generalized regression, nonparametric space-time (linear and generalized) regression, and their semiparametric analogues (see, e.g., Sangalli, 2021). Additionally, more complex scenarios for the error distribution, such as correlated or heteroskedastic noise, could be addressed by considering nonparametric quantile regression, substituting the sum-of-squares-errors in the regression functional (4) with the pinball loss, as recently done by Castiglione et al. (2025) for data observed over two-dimensional domains.

Acknowledgments. We are deeply grateful to L. Formaggia for insightful discussions on the numerical aspects of the proposed methods. A. Clemente and L.M. Sangalli acknowledge the PRIN research project CoEnv - Complex Environmental Data and Modeling (CUP 2022E3RY23), funded by the Italian Ministry for University and Research (MUR), and the European Union NextGenerationEU programme. A. Clemente acknowledges funding by the European Union - NextGenerationEU programme. A. Clemente and L.M. Sangalli also acknowledge the MUR research project Dipartimento di Eccellenza 2023–2027, Dipartimento di Matematica, Politecnico di Milano, Italy. J. Mateu is partially funded by grants CIAICO/2022/191 (from Generalitat Valenciana), and PID2022-141555OB-I00 (from Ministry of Science and Innovation).

References

Anderes, E., Møller, J. & Rasmussen, J. G. (2020). Isotropic covariance functions on graphs and their edges. *The Annals of Statistics* **48**.

Arioli, M. & Benzi, M. (2018). A finite element method for quantum graphs. *IMA Journal of Numerical Analysis* **38**, 1119–1163.

Arnone, E., Clemente, A., Sangalli, L. M., Lila, E., Ramsay, J. & Formaggia, L. (2025). *fdaPDE: Physics-Informed Spatial and Functional Data Analysis*. R package version 1.1-21.

Arnone, E., De Falco, C., Formaggia, L., Meretti, G. & Sangalli, L. M. (2023a). Computationally efficient techniques for spatial regression with differential regularization. *International Journal of Computer Mathematics* **100**, 1971–1991.

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* **79**, 3510–3521.

Azzimonti, L., Nobile, F., Sangalli, L. M. & Secchi, P. (2014). Mixed finite elements for spatial regression with pde penalization. *SIAM/ASA Journal on Uncertainty Quantification* **2**, 305–335.

Azzimonti, L., Sangalli, L. M., Secchi, P., Domanin, M. & Nobile, F. (2015). Blood flow velocity field estimation via spatial regression with pde penalization. *Journal of the American Statistical Association* **110**, 1057–1071.

Baddeley, A., Rubak, E. & Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. London: Chapman and Hall/CRC Press.

Baramidze, V., Lai, M.-J. & Shum, C. (2006). Spherical splines for data interpolation and fitting. *SIAM Journal on Scientific Computing* **28**, 241–259.

Barbi, C., Menafoglio, A. & Secchi, P. (2023). An object-oriented approach to the analysis of spatial complex data over stream-network domains. *Spatial Statistics* , 100784.

Barry, R. (2021). *latticeDensity: Density Estimation and Nonparametric Regression on Irregular Regions*. R package version 1.2.6.

Barry, R. P. & McIntyre, J. (2020). Lattice-based methods for regression and density estimation on complicated multidimensional regions. *Environmental and Ecological Statistics* **27**, 571–589.

Begu, B., Panzeri, S., Arnone, E., Carey, M. & Sangalli, L. M. (2024). A nonparametric penalized likelihood approach to density estimation of space–time point patterns. *Spatial Statistics* **61**, 100824.

Berkolaiko, G. & Kuchment, P. (2013). *Introduction to quantum graphs*. No. 186. American Mathematical Soc.

Bolin, D., Simas, A. B. & Wallin, J. (2023). *MetricGraph: Random fields on metric graphs*. R package version 1.4.0.

Bolin, D., Simas, A. B. & Wallin, J. (2024). Gaussian whittle–matérn fields on metric graphs. *Bernoulli* **30**, 1611–1639.

Boots, B., Sugihara, K., Chiu, S. N. & Okabe, A. (2009). Spatial tessellations: concepts and applications of voronoi diagrams .

Borgwardt, K. M. & Kriegel, H.-P. (2005). Shortest-path kernels on graphs. In *Fifth IEEE international conference on data mining (ICDM'05)*. IEEE.

Brezis, H. & Brézis, H. (2011). *Functional analysis, Sobolev spaces and partial differential equations*, vol. 2. Springer.

Castiglione, C., Arnone, E., Bernardi, M., Farcomeni, A. & Sangalli, L. M. (2025). Pde-regularised spatial quantile regression. *Journal of Multivariate Analysis* **205**, 105381.

Cavazzutti, M., Arnone, E., Ferraccioli, F., Galimberti, C., Finos, L. & Sangalli, L. M. (2024). Sign-flip inference for spatial regression with differential regularisation. *Stat* **13**, e711.

Cox, D. D. (1983). Asymptotics for m-type smoothing splines. *The Annals of Statistics* , 530–551.

Cox, D. D. (1984). Multivariate smoothing spline functions. *SIAM Journal on Numerical Analysis* **21**, 789–813.

Cox, D. D. & O'Sullivan, F. (1990). Asymptotic analysis of penalized likelihood and related estimators. *The Annals of Statistics* , 1676–1695.

Craven, P. & Wahba, G. (1978/79). Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **31**, 377–403.

Cucker, F. & Zhou, D. X. (2007). *Learning theory: an approximation theory viewpoint*, vol. 24. Cambridge University Press.

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces , 85–100. Lecture Notes in Math., Vol. 571.

Eilers, P. H. C. & Marx, B. D. (1996). Flexible smoothing with *B*-splines and penalties. *Statist. Sci.* **11**, 89–121. With comments and a rejoinder by the authors.

Ettinger, B., Perotto, S. & Sangalli, L. M. (2016). Spatial regression models over two-dimensional manifolds. *Biometrika* **103**, 71–88.

Ferraccioli, F., Arnone, E., Finos, L., Ramsay, J. O. & Sangalli, L. M. (2021). Nonparametric density estimation over complicated domains. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **83**, 346–368.

Ferraccioli, F., Sangalli, L. M. & Finos, L. (2022). Some first inferential tools for spatial regression with differential regularization. *Journal of Multivariate Analysis* **189**, 104866.

Fotheringham, A. S., Brunsdon, C. & Charlton, M. (2003). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

Geer, S. A. (2000). *Empirical Processes in M-estimation*, vol. 6. Cambridge university press.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C. & Harris, P. (2015). GWmodel: An R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software* **63**, 1–50.

Green, P. J. & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*, vol. 58 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London. A roughness penalty approach.

Gu, C. & Qiu, C. (1993). Smoothing spline density estimation: Theory. *The Annals of Statistics* , 217–234.

Györfi, L., Kohler, M., Krzyzak, A., Walk, H. et al. (2002). *A distribution-free theory of nonparametric regression*, vol. 1. Springer.

Haeseler, S. (2011). Heat kernel estimates and related inequalities on metric graphs. *arXiv preprint arXiv:1101.3010* .

Hastie, T. & Tibshirani, R. (1986). Generalized additive models. *Statist. Sci.* **1**, 297–318. With discussion.

Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized additive models*, vol. 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics* **31**, 1600–1635.

Koenker, R., Ng, P. & Portnoy, S. (1994). Quantile smoothing splines. *Biometrika* **81**, 673–680.

Ladle, A., Avgar, T., Wheatley, M. & Boyce, M. S. (2017). Predictive modelling of ecological patterns along linear-feature networks. *Methods in Ecology and Evolution* **8**, 329–338.

Lai, M.-J. & Schumaker, L. L. (2007). *Spline functions on triangulations*. No. 110. Cambridge University Press.

Lai, M.-J. & Wang, L. (2013). Bivariate penalized splines for regression. *Statistica sinica* , 1399–1417.

Lindgren, F., Rue, H. & Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **73**, 423–498.

Lu, B., Charlton, M., Harris, P. & Fotheringham, A. S. (2014). Geographically weighted regression with a non-euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science* **28**, 660–681.

Marron, J. (1987). A comparison of cross-validation techniques in density estimation. *The Annals of Statistics* , 152–162.

McIntyre, J. & Barry, R. P. (2018). A lattice-based smoother for regions with irregular boundaries and holes. *Journal of Computational and Graphical Statistics* **27**, 360–367.

McSwiggan, G., Baddeley, A. & Nair, G. (2017). Kernel density estimation on a linear network. *Scandinavian Journal of Statistics* **44**, 324–345.

Moradi, M. M., Cronie, O., Rubak, E., Lachieze-Rey, R., Mateu, J. & Baddeley, A. (2019). Resample-smoothing of voronoi intensity estimators. *Statistics and computing* **29**, 995–1010.

Okabe, A. & Sugihara, K. (2012). *Spatial analysis along networks: statistical and computational methods*. John Wiley & Sons.

Pollard, D. (2002). *A user's guide to measure theoretic probability*. No. 8. Cambridge University Press.

Porcu, E., White, P. A. & Genton, M. G. (2023). Stationary nonseparable space-time covariance functions on networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **85**, 1417–1440.

Rakshit, S., Davies, T., Moradi, M. M., McSwiggan, G., Nair, G., Mateu, J. & Baddeley, A. (2019). Fast kernel smoothing of point patterns on a large network using two-dimensional convolution. *International Statistical Review* **87**, 531–556.

Ramsay, J. & Silverman, B. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer.

Ramsay, T. (2002). Spline smoothing over difficult regions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**, 307–319.

Ruppert, D., Wand, M. P. & Carroll, R. J. (2003). *Semiparametric regression*. No. 12. Cambridge university press.

Salsa, S. (2016). *Partial differential equations in action: from modelling to theory*, vol. 99. Springer.

Sangalli, L. M. (2021). Spatial regression with partial differential equation regularisation. *International Statistical Review* **89**, 505–531.

Sangalli, L. M., Ramsay, J. O. & Ramsay, T. O. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 681–703.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M. & Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks* **20**, 61–80.

Schneble, M. & Kauermann, G. (2022). Intensity estimation on geometric networks with penalized splines. *The Annals of Applied Statistics* **16**, 843–865.

Shervashidze, N., Schweitzer, P., Van Leeuwen, E. J., Mehlhorn, K. & Borgwardt, K. M. (2011). Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research* **12**.

Silverman, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *The Annals of Statistics* , 795–810.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The annals of statistics* , 1040–1053.

TOMASETTO, M., ARNONE, E. & SANGALLI, L. M. (2024). Modeling anisotropy and non-stationarity through physics-informed spatial regression. *Environmetrics* **35**, e2889.

VAN DE GEER, S. & WEGKAMP, M. (1996). Consistency for the least squares estimator in nonparametric regression. *The Annals of Statistics* , 2513–2523.

VER HOEF, J. M. (2018). Kriging models for linear networks and non-euclidean distances: Cautions and solutions. *Methods in Ecology and Evolution* **9**, 1600–1613.

VER HOEF, J. M., PETERSON, E. & THEOBALD, D. (2006). Spatial statistical models that use flow and stream distance. *Environmental and Ecological statistics* **13**, 449–464.

VER HOEF, J. M. & PETERSON, E. E. (2010). A moving average approach for spatial statistical models of stream networks. *Journal of the American Statistical Association* **105**, 6–18.

VER HOEF, J. M., PETERSON, E. E. & ISAAK, D. J. (2019). Spatial statistical models for stream networks. In *Handbook of Environmental and Ecological Statistics*. Chapman and Hall/CRC, pp. 421–444.

WAHBA, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing* **2**, 5–16.

WAHBA, G. (1990). *Spline models for observational data*, vol. 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

WILHELM, M. & SANGALLI, L. M. (2016). Generalized spatial regression with differential regularization. *Journal of Statistical Computation and Simulation* **86**, 2497–2518.

WOOD, S. N. (2017). *Generalized additive models*. Texts in Statistical Science Series. CRC Press, Boca Raton, FL. An introduction with R.

WOOD, S. N., BRAVINGTON, M. V. & HEDLEY, S. L. (2008). Soap film smoothing. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **70**, 931–955.

WU, Z., PAN, S., CHEN, F., LONG, G., ZHANG, C. & YU, P. S. (2020). A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems* **32**, 4–24.

YIN, L. & SANG, H. (2021). Fused spatial point process intensity estimation with varying coefficients on complex constrained domains. *Spatial Statistics* **46**, 100547.

Supplementary Material for Nonparametric estimators over metric graphs

By ALDO CLEMENTE
Department of Mathematics, Politecnico di Milano,
Piazza Leonardo da Vinci 32, 20133 Milano, Italy
aldo.clemente@polimi.it

ELEONORA ARNONE
Department of Management, Università degli Studi di Torino,
Via Giuseppe Verdi 8, 10124 Torino, Italy
eleonora.arnone@unito.it

JORGE MATEU
Department of Mathematics, Universitat Jaume I,
Avinguda de Vicent Sos Baynat, 12006 Castelló de la Plana, Spain
mateu@uji.es

AND LAURA M. SANGALLI*
Department of Mathematics, Politecnico di Milano,
Piazza Leonardo da Vinci 32, 20133 Milano, Italy
laura.sangalli@polimi.it

A. Proofs of Section *2*

A.1. *Proof of Theorem 1*

The proof mimic the one for the classical Poincaré inequality, for topological domains that are subsets of $\mathbb{R}^d$ in Salsa (2016) once noted that the embedding of $H^1(G)$ into $L^2(G)$ is compact (see, e.g., Lemma 3.27 Mugnolo, 2014) and that the functions in $H^1(G)$ are continuous. We report it here for completeness.

By contradiction, suppose the theorem is not true. This means that $\forall j > 1$ there exists $u_j \in H_0^1(G)$ such that

$$\|u_j\|_{L^2(G)} > j \left\| \frac{du_j}{dx} \right\|_{L^2(G)}. \tag{A.1}$$

Define a new sequence $w_j := u_j / \|u_j\|_{L^2(G)}$, such that, by construction $\|w_j\|_{L^2(G)} = 1$ for all $j > 1$. From equation A.1 we have

$$\left\| \frac{dw_j}{dx} \right\|_{L^2(G)} < \frac{1}{j} \leq 1.$$

Thus the sequence $\{w_j\}$ is bounded in $H^1(G)$, and using the compactness of the embedding of $H^1(G)$ in $L^2(G)$ there exists a subsequence $\{w_{j_k}\}$ and $w \in H_0^1(G)$ such that $w_{j_k} \to w$ in $L^2(G)$ and $\frac{dw_{j_k}}{dx} \rightharpoonup \frac{dw}{dx}$ in $L^2(G)$.
Using the continuity and the weak sequential lower semicontinuity of the norm we have $\|w\|_{L^2(G)} = 1$ and

$$\left\| \frac{dw}{dx} \right\|_{L^2(G)} \leq \liminf \left\| \frac{dw_{j_k}}{dx} \right\|_{L^2(G)} = 0.$$

Thus, w is constant on each edge of the graph, that is, $\forall e \in E$, there exists $c_e \in \mathbb{R}$ such that $w|_e = c_e$. Moreover, $w \in H^1(G)$ imply that w is continuous, that is $c_e = c$ for all $e \in E$. Finally, since we assume that there is at list a Dirichlet boundary node, w must satisfy the homogeneous Dirichlet boundary condition, therefore $w = 0$, in contradiction to $\|w\|_{L^2(G)} = 1$.

A.2. *Proof of Corollary 1*

Let u belong to $H^2_{NK,0}(G)$. Using the definition of $\|\cdot\|_{\tilde{H}^2(G)}$according to (2), we have that:

$$\|u\|^2_{\tilde{H}^2(G)} = \left\| \frac{d^2u}{dx^2} \right\|^2_{L^2(G)} + \left\| \frac{du}{dx} \right\|^2_{L^2(G)} + \|u\|^2_{L^2(G)}. \tag{A.2}$$

We first derive a bound for the second term of the equation (A.2). Note that

$$\begin{aligned} \left\| \frac{du}{dx} \right\|^2_{L^2(G)} &= \sum_{e \in E} \int_e \frac{du}{dx_e} \frac{du}{dx_e} = \sum_{e \in E} \left(\left[u \frac{du}{dx_e} \right]_e - \int_e \frac{d^2u}{dx_e^2} u \right) \\ &= \sum_{v_\ell \in H^2_{NK}(G)} u(v_\ell) \left(\sum_{e \in E_\ell} \frac{du}{dx_e}(v_\ell) \right) - \sum_{e \in E} \int_e \frac{d^2u}{dx_e^2} u. \end{aligned}$$

The first term in the last equation is null, due to the condition imposed at vertices on $H^2_{NK,0}(G)$. Finally, we can exploit the standard Cauchy-Schwarz inequality and the Young inequality, obtaining the following inequality:

$$\left\| \frac{du}{dx} \right\|^2_{L^2(G)} \leq \sum_{e \in E} \left(\frac{c}{2} \left\| \frac{d^2u}{dx^2} \right\|^2_{L^2(e)} + \frac{1}{2c} \|u\|^2_{L^2(e)} \right).$$

Setting $c = c_p^2$ and using the definition of L^2-norm according to (1) and the Poincare-type inequality in Theorem 1, we get

$$\left\| \frac{du}{dx} \right\|^2_{L^2(G)} \leq \frac{c_p^2}{2} \left\| \frac{d^2u}{dx^2} \right\|^2_{L^2(G)} + \frac{1}{2} \left\| \frac{du}{dx} \right\|^2_{L^2(G)}.$$

Hence,

$$\left\| \frac{du}{dx} \right\|^2_{L^2(G)} \leq c_p^2 \left\| \frac{d^2u}{dx^2} \right\|^2_{L^2(G)}. \tag{A.3}$$

The first term on the right hand side of equation (A.2) is bounded as follows:

$$\|u\|^2_{L^2(G)} \leq c_p^2 \left\| \frac{du}{dx} \right\|^2_{L^2(G)} \leq c_p^4 \left\| \frac{d^2u}{dx^2} \right\|^2_{L^2(G)},$$

thanks to the Poincaré-type inequality in Theorem 1 and equation (A.3). The left hand side of equation (A.2) is thus bounded by $(1 + c_p^2 + c_p^4) \left\| \frac{d^2u}{dx^2} \right\|^2_{L^2(G)}$, yielding the desired result.

A.3. *Proof of Corollay 2*

The inequality $c_L \|u\|_{\tilde{H}^2(G)} \leq \left\| \frac{d^2u}{dx^2} \right\|_{L^2(G)}$, for some constant c_L, follows form Corollary 1. The inequality $\left\| \frac{d^2u}{dx^2} \right\|^2_{L^2(G)} \leq \|u\|^2_{\tilde{H}^2(G)}$ is trivially satisfied, being inherited by same inequality on each graph edge, thanks to the definition of the $L^2(G)$-norm and $\tilde{H}^2(G)$-norm according to (1) and (2).

B. Proofs of Section *3*

B.1. *Proof of Theorem 2*

We start by considering the homogeneous case, $f \in H^2_{NK,0}(G)$. Note that (4) is equivalent to the following minimization problem:

$$\hat{f} = \underset{f \in H^2_{NK,0}(G)}{\arg\min} \left\{ \frac{1}{n} \boldsymbol{f}_n^\top \boldsymbol{f}_n + \lambda \left\| \frac{d^2 f}{dx^2} \right\|^2_{L^2(G)} - \frac{2}{n} \boldsymbol{f}_n \boldsymbol{y} \right\}, \tag{B.1}$$

since the term $\boldsymbol{y}^\top \boldsymbol{y}/n$ does not depend on f.

Existence and uniqueness of the estimator $\hat{f}$ in (4) can be shown by resorting to a classical characterization theorem (see, e.g., Chapter 2 of Braess, 2007), which we report below:

Theorem 8 (Braess (2007)). *If the functional $J(g)$ has the form*

$$J(g) = \mathcal{A}(g, g) + \mathcal{L}g + c,$$

where $A : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ is a continuous, coercive and symmetric bilinear form in $\mathcal{G}$, $L : \mathcal{G} \to \mathbb{R}$ is a linear operator, c is a constant and $\mathcal{G}$ is a Hilbert space, then there exists a unique $\hat{g} \in \mathcal{G}$ such that

$$J(\hat{g}) = \inf_{g \in \mathcal{G}} J(g).$$

Moreover, $\hat{g}$ satisfies the following Euler–Lagrange equation:

$$(J'(\hat{g}), \varphi) = 2A(\hat{g}, \varphi) + L\varphi = 0 \qquad \forall \varphi \in \mathcal{G}.$$

We apply this theorem to our estimation problem by writing the functional (B.1) that we want to minimize in terms of $\mathcal{A}$ and $\mathcal{L}$ that, for any $u, v \in H^2_{NK,0}(G)$, are defined as follows:

$$\mathcal{A}(u, v) := \frac{1}{n} \boldsymbol{u}_n^\top \boldsymbol{v}_n + \lambda \sum_{e \in E} \int_e \frac{d^2u}{dx_e^2} \frac{d^2v}{dx_e^2} \tag{B.2}$$

$$\mathcal{L}(v) := -\frac{2}{n}\boldsymbol{v}_n^\top \boldsymbol{y} \tag{B.3}$$

To exploit the theorem presented above, we show that the form $\mathcal{A}$ in (B.2) is continuous, coercive and symmetric on $H^2_{NK,0}(G) \times H^2_{NK,0}(G)$, and $\mathcal{L}$ in (B.3) is continuous on $H^2_{NK,0}(G)$. Exploiting the Sobolev embedding theorem over each graph edge (see, e.g., Brezis & Brézis, 2011), and the definitions of norms in (1) and (2), we obtain

$$\begin{aligned} |\mathcal{A}(u,v)| &\leq c_1 \|u\|_{C^0(G)} \|v\|_{C^0(G)} + \lambda |E| \left\| \frac{d^2u}{dx^2} \right\|_{L^2(G)} \left\| \frac{d^2v}{dx^2} \right\|_{L^2(G)} \\ &\leq (c_1 + \lambda|E|) \|u\|_{\tilde{H}^2(G)} \|v\|_{\tilde{H}^2(G)}, \end{aligned}$$

where $|E|$ is the number of edges of the graph. Moreover,

$$|\mathcal{L}(v)| = |2\boldsymbol{v}_n^\top \boldsymbol{y}| \leq c_2 \|v\|_{C^0(G)} \leq c_2 \|v\|_{\tilde{H}^2(G)}.$$

This proves the continuity of $\mathcal{A}(u,v)$ on $H^2_{NK,0}(G) \times H^2_{NK,0}(G)$ and the continuity of $\mathcal{L}(v)$ on $H^2_{NK,0}(G)$. Moreover, $\mathcal{A}(u,v)$ is symmetric and coercive, since

$$\mathcal{A}(u,u) = \frac{1}{n}\boldsymbol{u}_n^\top \boldsymbol{u} + \lambda \sum_{e\in E} \int_e \left| \frac{d^2u}{dx^2} \right|^2 \geq \lambda \|u\|^2_{L^2(G)} \geq \lambda c_L^2 \|u\|^2_{\tilde{H}^2(G)},$$

where the last inequality follows from Corollary 2. This concludes the proof in the case of homogeneous boundary conditions.

The non-homogeneous boundary conditions can be handled by introducing a lift of the solution, as detailed e.g. in Azzimonti et al. (2014) for a nonparametric estimator over two-dimensional domains, leveraging the results in Berkolaiko & Kuchment (2013) on PDEs with non-homogeneous boundary conditions. Specifically, we define a fixed function $f_{\text{lift}} \in H^2_{NK,\gamma}(G)$ satisfying the prescribed non-homogeneous boundary conditions on V_D, which is the subset of the boundary vertices of the graph where Dirichlet boundary conditions are imposed; see Section 2 of the main article. This function f_{lift} does not depend on the data and can be any smooth function on G satisfying the prescribed non-homogeneous boundary conditions on V_D. We then write $f = \tilde{f} + f_{\text{lift}}$, where the unknown component $\tilde{f}$ belongs to $H^2_{NK,0}(G)$ and satisfies homogeneous boundary conditions. Consequently, the original problem with non-homogeneous boundary conditions reduces to an equivalent problem for $\tilde{f}$ with homogeneous boundary conditions. This concludes the proof of Theorem 2.

B.2. *Proof of Theorem 3*

We report Lemma 4.2 of Cox (1984) for the case of interval domains.

LEMMA 2 (COX (1984)). *Let Assumptions 1 and 2 hold, where the assumptions have to be interpreted over the interval (a,b), for the functions $F, F_n : (a,b) \to \mathbb{R}$. Then, for functions $h, g \in H^1(a,b)$, we have*

$$\left| \int_a^b hg \ d(F - F_n) \right| \leq c d_n \|h\|_{H^1(a,b)} \ \|g\|_{H^1(a,b)}.$$

Applying the triangle inequality to the left-hand side of equation (6) and then Lemma 2 to each edge of the graph, we obtain

$$\left|\sum_{e\in E}\int_e hg\ d(F-F_n)\right| \leq \sum_{e\in E}\left|\int_e hg\ d(F-F_n)\right| \leq \sum_{e\in E} cd_n\|h\|_{H^1(e)}\ \|g\|_{H^1(e)}.$$

Using the norm $\|\cdot\|_{\tilde{H}^1}$ defined according to (2), we obtain:

$$\sum_{e\in E} cd_n\|h\|_{H^1(e)}\ \|g\|_{H^1(e)} \leq \|h\|_{\tilde{H}^1(G)}\ \|g\|_{\tilde{H}^1(G)}\sum_{e\in E} cd_n \leq cd_n|E|\ \|h\|_{\tilde{H}^1(G)}\ \|g\|_{\tilde{H}^1(G)}$$

yielding the desired result.

B.3. *Proof of Theorem 4*

Similarly to the case of two-dimensional domains, the expected value of the estimator $\mathbb{E}[\hat{f}]$ is the solution of the following minimization problem:

$$\mathbb{E}[\hat{f}] = \underset{f\in H^2_{NK,\gamma}(G)}{\arg\min}\ \left\{\frac{1}{n}\sum_{i=1}^{n}(f_0(\boldsymbol{p}_i)-f(\boldsymbol{p}_i))^2 + \lambda_n\left\|\frac{d^2 f}{dx^2}\right\|^2_{L^2(G)}\right\},$$

see, e.g., to Arnone et al. (2022). Moreover, thanks to Theorem 2, $\mathbb{E}[\hat{f}]$ satisfies (5) for any test function belonging to $H^2_{NK,0}(G)$. Specifically, solving the minimization problem (5) is equivalent to finding $\hat{f}^*$ such that

$$\lambda_n\int_\Omega \frac{d^2\hat{f}}{dx_e^2}\frac{d^2 v}{dx_e^2} + \frac{1}{n}\sum_{i=1}^{n}\hat{f}(\boldsymbol{p}_i)\,v(\boldsymbol{p}_i) = \frac{1}{n}\sum_{i=1}^{n} f_0(\boldsymbol{p}_i)\,v(\boldsymbol{p}_i) \qquad \forall v\in H^2_{NK,0}(G).$$

We can rewrite the aforementioned relation in terms of bias $\mathcal{B} = f_0 - \mathbb{E}[\hat{f}]$ by subtraction the quantity $\lambda_n\sum_{e\in E}\int_e \frac{d^2 f_0}{dx_e^2}\frac{d^2 v}{dx_e^2}$ and adding $\sum_{e\in E}\int_e \mathcal{B}v\,dF$ on both sides of the previous equation, thus obtaining

$$\lambda_n\int_\Omega \frac{d^2\mathcal{B}}{dx_e^2}\,\frac{d^2 v}{dx_e^2} + \int_\Omega \mathcal{B}v\,dF = \lambda_n\sum_{e\in E}\int_e \frac{d^2 f_0}{dx_e^2}\,\frac{d^2 v}{dx_e^2} + \sum_{e\in E}\int_e \mathcal{B}v\,dF - \frac{1}{n}\sum_{i=1}^{n}\mathcal{B}(\boldsymbol{p}_i)\,v(\boldsymbol{p}_i).$$

$\forall v\in H^2_{NK,0}(G)$. Since $\mathcal{B}\in H^2_{NK,0}(G)$, setting $v=\mathcal{B}$ we get

$$\lambda_n\left\|\frac{d^2\mathcal{B}}{dx^2}\right\|^2_{L^2(G)} + \sum_{e\in E}\int_e \mathcal{B}^2 dF = \lambda_n\sum_{e\in E}\int_e \frac{d^2 f_0}{dx_e^2}\frac{d^2\mathcal{B}}{dx_e^2} + \sum_{e\in E}\int_e \mathcal{B}^2 d(F-F_n).$$

Note that, by Assumption 1 we have

$$\|v\|^2_{L^2(G)} = \sum_{e\in E}\int_e v^2 dx = \sum_{e\in E}\int_e \frac{v^2}{f} f dx \leq \frac{1}{k_1}\sum_{e\in E}\int_e v^2 dF.$$

Hence, the following inequality holds

$$\lambda_n\left\|\frac{d^2\mathcal{B}}{dx^2}\right\|^2_{L^2(G)} + k_1\|\mathcal{B}\|^2_{L^2(G)} \leq \lambda_n\sum_{e\in E}\int_e \frac{d^2 f_0}{dx_e^2}\frac{d^2\mathcal{B}}{dx_e^2} + \sum_{e\in E}\int_e \mathcal{B}^2 d(F-F_n). \tag{B.4}$$

Thanks to Theorem 3, the second term on the right side of (B.4) is bounded as follows:

$$\sum_{e\in E}\int_e \mathcal{B}^2 d(F-F_n) \leq cd_n|E|\|\mathcal{B}\|^2_{\tilde{H}^1(G)} \leq cd_n|E|\|\mathcal{B}\|^2_{\tilde{H}^2(G)}.$$

For what concerns the first term on the right-hand side of (B.4), we exploit, in order, the Cauchy-Schwarz inequality in Lemma 1, the definition of the norm $\|\cdot\|_{L^2(G)}$ according to (1), and finally the Young inequality, leading to

$$\begin{aligned}\lambda_n \sum_{e\in E}\int_e \frac{d^2 f_0}{dx_e^2}\frac{d^2\mathcal{B}}{dx_e^2} \leq & \lambda_n \left\|\frac{d^2 f_0}{dx^2}\right\|_{L^2(G)} \left\|\frac{d^2\mathcal{B}}{dx^2}\right\|_{L^2(G)} \\ \leq & \lambda_n\left(\frac{1}{2\varepsilon}\left\|\frac{d^2 f_0}{dx^2}\right\|^2_{L^2(G)} + \frac{\varepsilon}{2}\left\|\frac{d^2\mathcal{B}}{dx^2}\right\|^2_{L^2(G)}\right). \qquad \text{(B.5)}\end{aligned}$$

The above inequality holds $\forall\varepsilon > 0$. Setting $\varepsilon = c_L^2$, we obtain

$$\lambda_n \sum_{e\in E}\int_e \frac{d^2 f_0}{dx_e^2}\frac{d^2\mathcal{B}}{dx_e^2} \leq \lambda_n\left(\frac{1}{2c_L^2}\left\|\frac{d^2 f_0}{dx^2}\right\|^2_{L^2(G)} + \frac{c_L^2}{2}\left\|\frac{d^2\mathcal{B}}{dx^2}\right\|^2_{L^2(G)}\right).$$

Hence, the left-hand side of (B.4) is bounded as follows:

$$\lambda_n\left\|\frac{d^2\mathcal{B}}{dx^2}\right\|^2_{L^2(G)} + k_1\|\mathcal{B}\|^2_{L^2(G)} \leq \frac{\lambda_n}{2c_L^2}\left\|\frac{d^2 f_0}{dx^2}\right\|^2_{L^2(G)} + \frac{\lambda_n c_L^2}{2}\left\|\frac{d^2\mathcal{B}}{dx^2}\right\|^2_{L^2(G)} + cd_n|E|\|\mathcal{B}\|^2_{\tilde{H}^2(G)}.$$

Finally, we exploit Corollary 2, which establishes the equivalence between the $\|\cdot\|_{L^2(G)}$-norm and the $\left\|\frac{d^2}{dx^2}(\cdot)\right\|_{L^2(G)}$-norm, together with Assumption 2, according to which, for n sufficiently large, we have $d_n\lambda_n^{-1} \leq \frac{c_L^2}{2c|E|}$. We thus obtain

$$\lambda_n c_L^2\|\mathcal{B}\|^2_{\tilde{H}^2(G)} + k_1\|\mathcal{B}\|^2_{L^2(G)} \leq \frac{\lambda_n}{c_L^2}\left\|\frac{d^2 f_0}{dx^2}\right\|^2_{L^2(G)} + \lambda_n c_L^2\|\mathcal{B}\|^2_{\tilde{H}^2(G)}.$$

Cancelling the common $\tilde{H}^2(G)$-term on both sides yields

$$\|\mathcal{B}\|_{L^2(G)} \leq \sqrt{\frac{\lambda_n}{k_1 c_L^2}}\left\|\frac{d^2 f_0}{dx^2}\right\|_{L^2(G)},$$

which concludes the proof.

B.4. *Proof of Theorem 5*

Thanks to the linearity of (5), the field estimator $\hat{f}$ can be written as $\hat{f} = \mathbb{E}[\hat{f}] + \hat{w}$, where

$$\hat{w} = \underset{w\in H^2_{NK,0}(G)}{\arg\min}\left\{\frac{1}{n}\sum_{i=1}^{n}(w(\boldsymbol{p}_i) - \varepsilon_i)^2 + \lambda_n\left\|\frac{d^2 w}{dx^2}\right\|^2_{L^2(G)}\right\}.$$

Note that $\mathbb{E}[\hat{w}] = 0$ and $Var(\hat{f}) = Var(\hat{w})$. Furthermore, thanks to Theorem 2, $\hat{w}$ satisfies (5), where $\boldsymbol{y}$ is replaced by $\boldsymbol{\varepsilon}$, for any test function belonging to $H^2_{NK,0}(G)$. Therefore, by rearranging equation (5) we have:

$$\lambda_n \sum_e \int_e \frac{d^2\hat{w}}{dx_e^2}\frac{d^2 v}{dx_e^2} + \sum_{e\in E}\int_e \hat{w}v\ dF = \\ \frac{1}{n}\sum_{i=1}^n \varepsilon_i v(\boldsymbol{p}_i) + \sum_{e\in E}\int_e \hat{w}v\ d(F-F_n) \qquad \forall v \in H^2_{\scriptscriptstyle NK,0}(G). \tag{B.6}$$

Denote the left-hand side of the equation above as $(\hat{w}, v)_{\lambda_n}$. Note that because of Corollary 2, which states the equivalence between the $\|\cdot\|_{L^2(G)}$-norm and the $\|\frac{d^2}{dx^2}(\cdot)\|_{L^2(G)}$-norm, we can bound the norm induced by the scalar product $(\cdot,\cdot)_{\lambda_n}$ as the follows:

$$\|w\|^2_{H^2(G)} \le (\tilde{c}\lambda_n)^{-1}\|w\|^2_{\lambda_n}. \tag{B.7}$$

Consider the following operators:

$$T_1(v) := \sum_{e\in E}\int_e \hat{w}v\ d(F-F_n), \quad T_2(v) := \frac{1}{n}\sum_{i=1}^n \varepsilon_i v(\boldsymbol{p}_i), \quad T(v) := T_1(v) + T_2(v).$$

The operator T_1 is bounded on $H^1_0(G)$ due to Theorem 3. Moreover, T_2 belongs to $H^1_0(G)^*$, the dual space of $H^1_0(G)$, due to the one-dimensional Sobolev embedding theorem. Thus also the operator T belongs to $H^1_0(G)^*$. Thus, equation (B.6) can be rewritten as $(\hat{w}, v)_{\lambda_n} = T(v)$, $\forall v \in H^2_{\scriptscriptstyle NK,0}(G)$, and thus

$$\|\hat{w}\|_\lambda = \sup_{v\in H^2_{NK,0}(G)} \frac{T(v)}{\|v\|_{\lambda_n}} \le \sup_{v\in H^2_{NK,0}(G)} \frac{T_1(v)}{\|v\|_{\lambda_n}} + \sup_{v\in H^2_{NK,0}(G)} \frac{T_2(v)}{\|v\|_{\lambda_n}}. \tag{B.8}$$

The end of the proof proceeds along the lines of the proof of Theorem 2 in Arnone et al. (2022). Specifically, we use the operators introduced above, and show that

$$\sup_{v\in H^2_{NK,0}(G)} \frac{T_1(v)}{\|v\|_{\lambda_n}} \ \le\ c_1\, d_n\, \lambda_n^{-1}\, \|\hat{w}\|_{\lambda_n}, \qquad \sup_{v\in H^2_{NK,0}(G)} \frac{T_2(v)}{\|v\|_{\lambda_n}} \ \le\ c_2\, \lambda_n^{-1/4}\, \|T_2\|_{H^1_0(G)^*}. \tag{B.9}$$

The bound regarding T_1 follows by invoking Proposition 3 and equation (B.7). Regarding T_2, note that

$$\sup_{v\in H^2_{NK,0}(G)} \frac{T_2(v)}{\|v\|_{\lambda_n}} \le \sup_{v\in H^2_{NK,0}(G)} \frac{\|T_2\|_{H^1_0(G)^*}\,\|v\|_{H^1(G)}}{\|v\|_{\lambda_n}} \\ \le \sup_{v\in H^2_{NK,0}(G)} \frac{\lambda_n^{-1/4}\,\|T_2\|_{H^1_0(G)^*}\left(\lambda_n^{1/4}\,\|v\|^{1/2}_{H^2(G)}\,\|v\|^{1/2}_{L^2(G)}\right)}{\|v\|_{\lambda_n}}, \tag{B.10}$$

where the last inequality follows from the fact that the inequality $\|v\|_{H^1(e)} \le \|v\|^{1/2}_{H^2(e)}\,\|v\|^{1/2}_{L^2(e)}$, valid on each edge e of the graph, also holds on the whole graph G. Note that, by Assumption 1 and inequality (B.7), we have $\|v\|_{L^2(G)} \le k_1^{-1/2}\,\|v\|_{\lambda_n}$. Indeed,

$$\|v\|^2_{L^2(G)} = \sum_{e\in E}\int_e v^2 dx = \sum_{e\in E}\int_e \frac{v^2}{f} f dx \le \frac{1}{k_1}\sum_{e\in E}\int_e v^2 dF \le \frac{1}{k_1}\|v\|^2_{\lambda_n}.$$

Hence, by the above consideration and applying Young's inequality to the right-hand side of (B.10) and using again (B.7), we obtain the following bound

$$\sup_{v\in H^2_{NK,0}(G)} \frac{T_2(v)}{\|v\|_{\lambda_n}} \le (\tilde{c} + k_1^{-1/2})\lambda_n^{-1/4}\|T_2\|_{H^1_0(G)^*}.$$

This proves the bound regarding T_2 in (B.9), and we have

$$\|w\|_{\lambda_n} \le c_1\, d_n\, \lambda_n^{-1}\, \|\hat{w}\|_{\lambda_n} + c_2\lambda_n^{-1/4}\|T_2\|_{H^1_0(G)^*}.$$

Thanks to Assumption 2 , for n sufficiently large, $d_n\lambda_n = o(1)$. Hence, we can neglect the first term in the right-hand side of the previous equation and obtain

$$\|w\|_{\lambda_n} \le c_2\lambda_n^{-1/4}\|T_2\|_{H^1_0(G)^*}.$$

By squaring and taking the expected values of both terms of the above inequality, we have

$$\mathbb{E}\left(\|w\|^2_{\lambda_n}\right) \le c\lambda_n^{-1/2}\mathbb{E}\left(\|T_2\|^2_{H^1_0(G)^*}\right). \tag{B.11}$$

Finally, note that $\mathbb{E}(||T_2||^2_{H^1(G)^*}) \le c\sigma^2/n$. Indeed,

$$\begin{aligned}\mathbb{E}\Big(\|T_2\|^2_{(H^1)^*}\Big) = \mathbb{E}((T_2,T_2)_{1,*}) &= \mathbb{E}\left(\frac{1}{n^2}\sum_{i,j=1}^n \varepsilon_i\varepsilon_j(\delta_{p_i},\delta_{p_j})_{1,*}\right)\\ &= \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}(\varepsilon_i\varepsilon_j)(\delta_{p_i},\delta_{p_j})_{1,*}\\ &= \frac{1}{n^2}\sum_{i=1}^n \sigma^2\|\delta_{p_i}\|^2_{(H^1)^*} \le \frac{c\sigma^2}{n},\end{aligned}$$

where $c = \max_{i=1,\dots,n}\|\delta_{p_i}\|^2_{(H^1)^*} < \infty$. Hence, the desired result is obtained by combining the previous inequality, inequality (B.11) and noting once again that $\|w\|_{L^2(G)} \le k_1^{-1/2}\|w\|_{\lambda_n}$, where the constant k_1 does not depend on λ_n nor on n. This concludes the proof of Theorem 5.

B.5. *Proof of Corollary 3*

Leveraging Theorem 4 and Theorem 5, we have:

$$MSE_{L^2}(\hat{f}) = \|\mathcal{B}\|^2_{L^2(G)} + Var(\hat{f}) = O(\lambda_n) + O\left(\lambda_n^{-1/2}\frac{\sigma^2}{n}\right),$$

that is minimized when $\lambda_n = n^{-2/3}$.

C. Proofs of Section *4*

C.1. *Proof of Theorem 6*

Set

$$J(g) := -\frac{1}{n}\sum_{i=1}^n g(X_i) + \int_G \exp\{g\}.$$

Using Lemma 1 in Section 2.3 and following the same steps as in Lemma B.1 of Ferraccioli et al. (2021), it is possible to prove that the functional $J(g)$ is continuous and strictly convex on $H^2_{NK}(G)$. Specifically, the continuity of J is immediate, since the first term is linear and both the exponential and the integral operators are continuous. We now prove strict convexity of J. Let $g_1, g_2 \in H^2_{NK}(G)$, $\gamma \in [0,1]$, and define $g = \gamma g_1 + (1-\gamma) g_2$. We show that

$$J(g) \leq \gamma J(g_1) + (1-\gamma) J(g_2),$$

with equality if and only if $g_1 = g_2$. We have

$$\begin{aligned} J(g) &= J(\gamma g_1 + (1-\gamma) g_2) \\ &= -\frac{1}{n}\sum_{i=1}^{n}\big(\gamma g_1(X_i) + (1-\gamma) g_2(X_i)\big) + \int_G \exp\big(\gamma g_1 + (1-\gamma) g_2\big) \\ &= \gamma\left(-\frac{1}{n}\sum_{i=1}^{n} g_1(X_i)\right) + (1-\gamma)\left(-\frac{1}{n}\sum_{i=1}^{n} g_2(X_i)\right) + \int_G \exp(\gamma g_1)\exp((1-\gamma) g_2). \end{aligned}$$

By Hölder's inequality (Lemma 1 of the main paper) with $p = 1/\gamma$ and $q = 1/(1-\gamma)$, we obtain

$$\int_G \exp(\gamma g_1)\exp((1-\gamma) g_2) \leq \left(\int_G \exp(g_1)\right)^{\gamma}\left(\int_G \exp(g_2)\right)^{1-\gamma}.$$

Applying Young's inequality with the same exponents yields

$$\left(\int_G \exp(g_1)\right)^{\gamma}\left(\int_G \exp(g_2)\right)^{1-\gamma} \leq \gamma\int_G \exp(g_1) + (1-\gamma)\int_G \exp(g_2).$$

Combining the above inequalities, we conclude that

$$J(g) \leq \gamma J(g_1) + (1-\gamma) J(g_2),$$

which proves the convexity of J. To prove strict convexity, note that equality in Hölder's inequality holds only if there exist constants $a, b \neq 0$ such that

$$a\exp(g_1) = b\exp(g_2) \quad\iff\quad g_1 = g_2 + \log(b/a).$$

Moreover, in Young's inequality, the equality holds only when

$$\int_G \exp(g_1) = \int_G \exp(g_2).$$

Substituting $g_1 = g_2 + \log(b/a)$ in the equation above, we get $a = b$; this in turn implies $g_1 = g_2$. Thus J is strictly convex in $H^2_{NK}(G)$.

Define the space

$$W := \left\{\phi \in H^2_{NK}(G), \left\|\frac{d^2\phi}{dx^2}\right\|^2_{L^2(G)} = 0\right\}.$$

and denote by $W^{\perp}$ the orthogonal space of W. Hence, W and $W^{\perp}$ are subspaces of $H^2_{NK}(G)$ such that $H^2_{NK}(G) = W \oplus W^{\perp}$. We show that W coincides with the space of constant functions on G. Indeed, let u belong to W, then

$$0 = \left\| \frac{d^2u}{dx^2} \right\|^2_{L^2(G)} = \sum_{e \in E} \left\| \frac{d^2u}{dx^2} \right\|^2_{L^2(e)} \iff \frac{d^2u}{dx^2} = 0 \quad \forall e \in E.$$

Hence, the function $u \in W$ is piece-wise linear on each edge of the domain G and is the solution to the following system:

$$\begin{cases} \sum_{e \in E} \frac{d^2u}{dx^2} = 0 \\ \sum_{e \in E_\ell} \frac{du}{dx_e}(v_\ell) = 0 \quad \forall v_\ell \in H^2_{NK}(G). \end{cases}$$

Moreover, multiplying the first equation in the system by a test function, integrating by parts, applying the Neumann-Kirchoff conditions, and finally setting the test function equal to u, we get:

$$\sum_{e \in E} \left\| \frac{du}{dx} \right\|_{L^2(e)} = 0 \iff \frac{du}{dx} = 0 \quad \forall e \in E.$$

This proves that $u \in W$ is constant on each edge of the domain, and thus is globally constant on G, since $W \subset C^0(G)$. Consequently, $W^\perp$, the orthogonal space of W, is the space of the null mean functions over the graph. Hence, we have:

$$W := \left\{ u \in H^2_{NK}(G), u = c \in \mathbb{R} \right\}$$
$$W^\perp := \left\{ u \in H^2_{NK}(G), \int_G u = 0 \right\}.$$

Moreover, it is worth noting that W is a finite-dimensional space, since it is the space of all constant functions defined on G. Finally, we can prove that the estimation functional in (7) has a unique minimizer in $H^2_{NK}(G)$, by leveraging Theorem 4.1 of Gu & Qiu (1993), where we exploit the fact that $J(g)$ is strictly convex and continuous, it has a unique minimizer in W, and Corollary 2.

C.2. *Proof of Theorem 7*

The argument follows the proofs of Theorem 5.3 of Gu & Qiu (1993), adapted to the estimator over metric graph here considered. For simplicity of notation call

$$L(g) \; := \; -\frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{p}_i) \; + \; \int_G \exp\{g\} \; + \; \lambda \left\| \frac{d^2g}{dx^2} \right\|^2_{L^2(G)},$$

the functional in (7). Moreover, following the notation of Gu & Qiu (1993), call $J(g_1, g_2) = \int_G \frac{d^2g_1}{dx^2} \frac{d^2g_2}{dx^2}$ and $J(g) = J(g, g)$. The proof relies on Theorem 5.2 of Gu & Qiu (1993), which we here report for completeness.

Theorem 9 (Adapted from Gu & Qiu (1993), Theorem 5.2). *Assume that the eigenvalues $\{\eta_k^2\}_{k \geq 0}$ associated with the roughness penalty $J(\cdot)$ satisfy $\alpha k^2 \leq \eta_k^2 \leq \beta k^2$ for all $k \geq k_0$, for some constants $\alpha, \beta > 0$ and $k_0 \geq 0$, and moreover Assumption 4 is satisfied. Then, as $\lambda \to 0$ and $n\lambda^{1/2} \to \infty$,*

$$Var_{g_0}(\hat{g} - g^*) = o_p\Big(n^{-1}\lambda^{-1/2} + \lambda\Big), \qquad \lambda J(\hat{g} - g^*) = o_p\Big(n^{-1}\lambda^{-1/2} + \lambda\Big).$$

Consequently,

$$Var_{g_0}(\hat{g} - g_0) = O_p\Big(n^{-1}\lambda^{-1/2} + \lambda\Big)\,, \qquad \lambda J(\hat{g} - g_0) = O_p\Big(n^{-1}\lambda^{-1/2} + \lambda\Big)\,.$$

The following lemma ensures that the assumption of Theorem 9 are met for the roughness penalty defined on metric graphs.

LEMMA 3. *There exists an infinite set of functions ϕ_k such that*

$$\text{Cov}(\phi_k, \phi_j) = \delta_{k,j} \text{ and } \int_G \frac{d^2\phi_k}{dx^2}\frac{d^2\phi_j}{dx^2} = \eta_k^2\delta_{k,j}$$

where $\delta_{k,j}$ is the Kronecker delta and $0 \leq \eta_k \to \infty$. In addition, there exist two positive constants α and β such that, for all $k \geq 0$,

$$\eta_k = c_k k, \quad \alpha \leq c_k \leq \beta.$$

Proof. The spectrum of a quantum graph of finite measure is composed of an infinite set of L^2-eigenfunctions (see, e.g., Theorem 3.1.1 of Berkolaiko & Kuchment, 2013). Furthermore, from Theorem 3.1.8 of Berkolaiko & Kuchment (2013) it follows that, denoting by $\mu_k = \eta_k^2 = (c_k k)^2$ the corresponding eigenvalues, there exist two constants $\alpha, \beta > 0$ such that $\alpha \leq c_k \leq \beta,\ \forall k \geq 0$. □

The following lemma establishes a basic variational identity that will be used to control the symmetrized Kullback–Leibler divergence between g_0 and $\hat{g}$.

LEMMA 4. *Let $\hat{g}$ be the minimizer of $L(g)$ in (7) and g_0 be the true log–density. Then*

$$D_{sKL}(g_0, \hat{g}) \;=\; 2\lambda\int_G \frac{d^2\hat{g}}{dx^2}\,\frac{d^2(g_0 - \hat{g})}{dx^2} \;+\; \left[\frac{1}{n}\sum_{i=1}^n(\hat{g} - g_0)(\boldsymbol{p}_i) \;-\; \int_G (\hat{g} - g_0)\,e^{g_0}\right]. \quad \text{(C.1)}$$

Proof. The proof proceeds along the same lines as for equation (5.1) of Gu & Qiu (1993), that carries over to metric graphs. The only graph–specific observation is that all terms in L decompose edgewise and the first variation is linear. Specifically, we want to compute the Gateaux derivative of the functional we aim to minimize. Hence, define

$$A_{g,h}(t) := -\frac{1}{n}\sum_{i=1}^n (g + th)(X_i) + \int_G \exp(g + th)\,dx + \lambda\int_G \left(\frac{d^2}{dx^2}(g + th)\right)^2 dx.$$

Differentiating $A_{g,h}$ with respect to t, we obtain

$$\dot{A}_{g,h}(t) = -\frac{1}{n}\sum_{i=1}^n h(X_i) + \int_G \exp(g + th)\,h\,dx + 2\lambda\int_G \frac{d^2}{dx^2}(g + th)\,\frac{d^2 h}{dx^2}\,dx.$$

By definition of $\hat{g}$ as the minimizer of $L(g)$, we have

$$\dot{A}_{\hat{g},h}(0) = \frac{d}{dt}A_{\hat{g},h}(\hat{g} + th)\bigg|_{t=0} = -\frac{1}{n}\sum_{i=1}^n h(X_i) + \mu_{\hat{g}}(h) + 2\lambda\int_G \frac{d^2\hat{g}}{dx^2}\,\frac{d^2 h}{dx^2}\,dx = 0,$$

for all $h \in H^2_{NK}(G)$. We recover equation (C.1) by setting $h = \hat{g} - g_0$ and adding and subtracting $\mu_{g_0}(\hat{g} - g_0)$ in the equation above. □

Given Theorem 9 and Lemmas 3 and 4 above, we follow the proof of Theorem 5.3 in Gu & Qiu (1993). The first term on the right-hand side of Lemma 4 corresponds to

$2\lambda J(\hat{g}, g_0 - \hat{g})$. Applying in order the Cauchy–Schwarz inequality, the triangle inequality, and the Young's inequality we have:

$$|2\lambda\, J(\hat{g}, g_0 - \hat{g})| \le 2\lambda\sqrt{J(\hat{g})}\sqrt{J(\hat{g} - g_0)} \le \lambda J(g_0) + 3\lambda J(\hat{g} - g_0).$$

Assumption 3 gives $J(g_0) < \infty$ so $\lambda J(g_0) = O(\lambda)$. By Theorem 9 we have $\lambda J(\hat{g} - g_0) = O_p(n^{-1}\lambda^{-1/2} + \lambda)$. Thus, the first term on the right-hand-side of Lemma 4 is $O_p(n^{-1}\lambda^{-1/2} + \lambda)$.

The second term of the right hand side of Lemma 4 can be seen as an error between the norm induced by the true density and a Monte Carlo approximation of the integral based on the points $\boldsymbol{p}_i$. Expanding $\hat{g} - g_0$ in the orthonormal basis $\{\varphi_k\}_{k\ge 0}$ of Lemma 3, we write

$$\hat{g} - g_0 = \sum_{k\ge 0} a_k \varphi_k, \qquad \xi_k := \frac{1}{n}\sum_{i=1}^{n} \varphi_k(\boldsymbol{p}_i) - \int_G \varphi_k\, e^{g_0}.$$

Then the second term in Lemma 4 can be rewritten as

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{g} - g_0)(\boldsymbol{p}_i) - \int_G (\hat{g} - g_0)\, e^{g_0} = \sum_{k\ge 0} a_k\, \xi_k.$$

Following Gu & Qiu (1993), introduce the weights $w_k := 1 + \lambda\eta_k^2$. By the Cauchy–Schwarz inequality,

$$\left|\sum_{k\ge 0} a_k \xi_k\right| \le \left(\sum_{k\ge 0} w_k a_k^2\right)^{1/2} \left(\sum_{k\ge 0} w_k^{-1}\xi_k^2\right)^{1/2}. \tag{C.2}$$

Moreover, by Lemma 3 and the definition of Var_{g_0},

$$\sum_{k\ge 0} w_k a_k^2 = \sum_{k\ge 0}\{a_k^2 + \lambda\eta_k^2 a_k^2\} = \mathrm{Var}_{g_0}(\hat{g} - g_0) + \lambda J(\hat{g} - g_0),$$

which is $O_p(n^{-1}\lambda^{-1/2} + \lambda)$ by Theorem 9. For the second factor of (C.2), note that $\mathbb{E}(\xi_k) = 0$ and $\mathbb{E}(\xi_k^2) = n^{-1}\mathrm{Var}(\varphi_k) = n^{-1}$, hence

$$\mathbb{E}\left[\sum_{k\ge 0} w_k^{-1}\xi_k^2\right] = \frac{1}{n}\sum_{k\ge 0}\frac{1}{1 + \lambda\eta_k^2} = O(n^{-1}\lambda^{-1/2}),$$

where, in the last equality, we used $\eta_k \asymp k$ by Lemma 3. Combining the two factors yields

$$\left|\frac{1}{n}\sum_{i=1}^{n}(\hat{g} - g_0)(\boldsymbol{p}_i) - \int_G (\hat{g} - g_0)\, e^{g_0}\right| = O_p\left(n^{-1}\lambda^{-1/2} + n^{-1/2}\lambda^{1/4}\right). \tag{C.3}$$

Finally, combining the bounds on the two terms in Lemma 4, and noticing that $n^{-1/2}\lambda^{1/4} = (n^{-1}\lambda^{-1/2})^{1/2}\lambda^{1/2} \le n^{-1}\lambda^{-1/2} + \lambda$ we get $D_{sKL}(g_0, \hat{g}) = O_p(n^{-1}\lambda^{-1/2} + \lambda)$. This concludes the proof of Theorem 7.

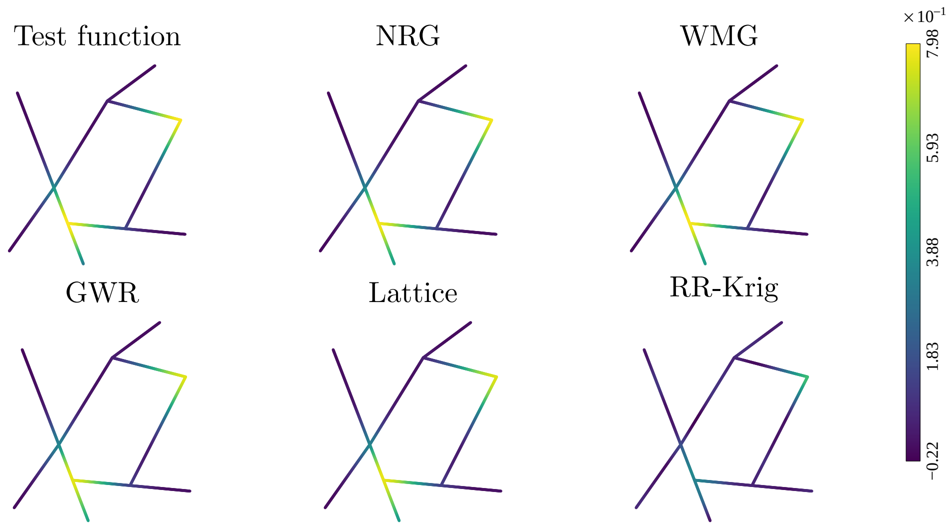

Fig. D.1: Simulation 3 (Section D): nonparametric regression problem - uniform locations and uncorrelated Gaussian noise. Top left: test function. Other panels: mean estimates obtained by the competing methods over 30 simulation repetitions with sample size $n = 100$. The competing methods are: Nonparametric Regression on Graph (NRG) presented in Section 3; Whittle-Matérn field (WMG); Geographically Weighted Regression model (GWR); Lattice-based model (Lattice); Reduced-Rank Kriging (RR-Krig).

D. Simulation study 3: nonparametric regression - uniform locations and uncorrelated Gaussian noise

This section presents an additional simulation study concerning a nonparametric regression problem. Specifically, we consider the same graph, mesh and test function as in Simulation 1 in Section 6.1 of the main paper. We generate the data at n locations uniformly sampled over the graph, by adding independent Gaussian noises to the test function, with variance equal to the 5% of the range of the true function. Likewise for Simulation 6.1, data generation (both locations and values) is repeated 30 times, for different sample sizes $n = 100, 150, 250, 500$.

We consider the same competing methods as in Simulation 6.1 and a Reduced-Rank Kriging (RR-Krig), which has been specifically developed by Ver Hoef (2018) for data observed over undirected linear networks. The method ensures valid spatial covariance matrices, using the shortest-path distance along the graph. It is worth mentioning that such a model was excluded from Simulation 6.1 due to numerical instabilities and poor performance relative to the other methods. The central and right column panels of Figure D.1 show the mean estimates provided by the competing methods, over 30 simulated repetitions with 100 observations. From a qualitative point of view, the Kriging method appears to oversmooth the signal, while the other methods are capable of capturing the main features of the signal. The left-hand panel of Figure D.2 shows the boxplots of the RMSE over 30 repetitions for each level of the sample size n. Also in this setting, the proposed NRG and WMG perform significantly better than the other methods for small n.

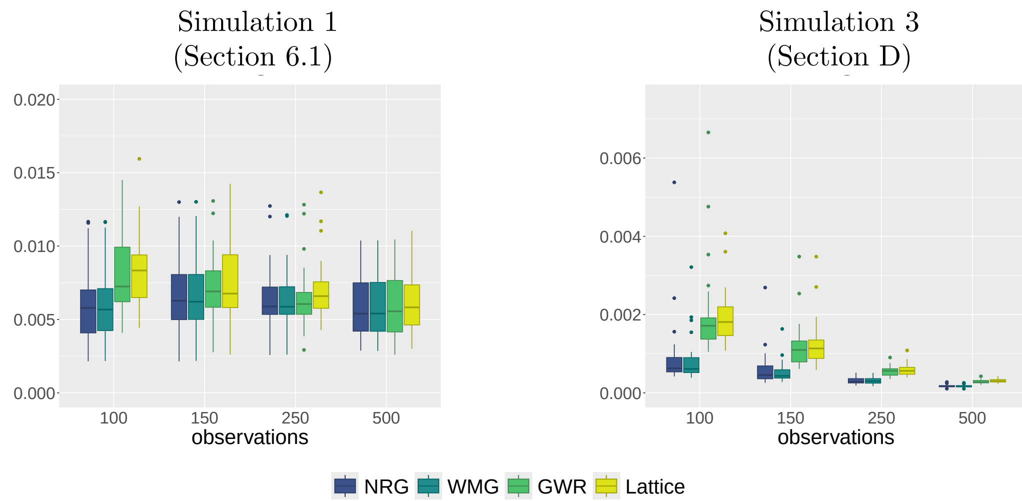

Fig. D.2: Boxplots of the Root Mean Square Error (RMSE) computed over 30 simulation repetitions of Simulation 1 (Section 6.1), on the left, and, Simulation 3 (Section D), on the right, for different sample sizes. Boxplots related to the RR-Krig have not been included, because its poor performance with respect to the other methods would have prevented to clearly appreciate the differences among remaining methods; for the same reason, we removed some extreme outliers observed for the GWR and Lattice methods.

References

Arnone, E., Kneip, A., Nobile, F. & Sangalli, L. M. (2022). Some first results on the consistency of spatial regression with partial differential equation regularization. *Statistica Sinica* **32**, 209–238.

Azzimonti, L., Nobile, F., Sangalli, L. M. & Secchi, P. (2014). Mixed finite elements for spatial regression with pde penalization. *SIAM/ASA Journal on Uncertainty Quantification* **2**, 305–335.

Berkolaiko, G. & Kuchment, P. (2013). *Introduction to quantum graphs*. No. 186. American Mathematical Soc.

Braess, D. (2007). *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, 3rd ed.

Brezis, H. & Brézis, H. (2011). *Functional analysis, Sobolev spaces and partial differential equations*, vol. 2. Springer.

Cox, D. D. (1984). Multivariate smoothing spline functions. *SIAM Journal on Numerical Analysis* **21**, 789–813.

Ferraccioli, F., Arnone, E., Finos, L., Ramsay, J. O. & Sangalli, L. M. (2021). Nonparametric density estimation over complicated domains. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **83**, 346–368.

Gu, C. & Qiu, C. (1993). Smoothing spline density estimation: Theory. *The Annals of Statistics* , 217–234.

Mugnolo, D. (2014). *Semigroup methods for evolution equations on networks*, vol. 20. Springer.

Salsa, S. (2016). *Partial differential equations in action: from modelling to theory*, vol. 99. Springer.

Ver Hoef, J. M. (2018). Kriging models for linear networks and non-euclidean distances: Cautions and solutions. *Methods in Ecology and Evolution* **9**, 1600–1613.