Journal of the Royal Statistical Society Series B: Statistical Methodology, 2023, **85**, 1417–1440
https://doi.org/10.1093/jrsssb/qkad082
Advance access publication 8 September 2023
Original Article

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series B
Statistical Methodology

Stationary nonseparable space-time covariance functions on networks

Emilio Porcu[1], Philip A. White[2,3] and Marc G. Genton[4]

[1]Department of Mathematics, Khalifa University, Abu Dhabi, United Arab Emirates
[2]Berry Consultants, Austin, Texas, United States
[3]Department of Statistics, Brigham Young University, Provo, Utah, United States
[4]Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Address for correspondence: Marc G. Genton, Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. Email: marc.genton@kaust.edu.sa

Abstract

The advent of data science has provided an increasing number of challenges with high data complexity. This paper addresses the challenge of space-time data where the spatial domain is not a planar surface, a sphere, or a linear network, but a generalised network (termed a graph with Euclidean edges). Additionally, data are repeatedly measured over different temporal instants. We provide new classes of stationary nonseparable space-time covariance functions where space can be a generalised network, a Euclidean tree, or a linear network, and where time can be linear or circular (seasonal). Because the construction principles are technical, we focus on illustrations that guide the reader through the construction of statistically interpretable examples. A simulation study demonstrates that the correct model can be recovered when compared to misspecified models. In addition, our simulation studies show that we effectively recover simulation parameters. In our data analysis, we consider a traffic accident dataset that shows improved model performance based on covariance specifications and network-based metrics.

Keywords: circular time, covariance function, dynamical support, generalised network, linear time, spatio-temporal statistics

1 Introduction

1.1 Context and state of the art

The data science revolution has introduced many challenges with data complexity. Amongst many, challenges related to domain complexity in georeferenced data and point pattern data are an important part of the literature, and the reader is referred to Anderes et al. (2020), Moradi and Mateu (2020), Baddeley et al. (2021), and Rakshit et al. (2017) for recent contributions.

This paper focuses on the development of covariance functions for space-time Gaussian random fields, where space is represented either as a generalised (or linear) network or a Euclidean tree, and time is modelled as either a linear (the real line) or circular (the unit circle) process. The use of stochastic processes defined over generalised networks has become increasingly important in spatial statistics, as evidenced by a growing number of applications in this field. Therefore, this paper aims to provide insights into the development of appropriate covariance functions for these types of spatial data (Cressie et al., 2006; Gardner et al., 2003; Montembeault et al., 2012; Peterson et al., 2007, 2013; Ver Hoef et al., 2006), point processes (Baddeley et al., 2017; Deng et al., 2014; Perry and Wolfe, 2013; Xiao et al., 2017), and machine learning (Alsheikh et al., 2014; Borovitskiy et al., 2022; Georgopoulos and Hasler, 2014; Hamilton et al., 2017; Pinder et al., 2021).

Covariance functions are positive definite, and this requirement is nontrivial to verify. Building positive definite functions is even more challenging for stochastic processes defined over networks,

Received: September 11, 2022. **Accepted:** July 30, 2023

and we refer the reader to Ver Hoef et al. (2006) and Peterson et al. (2007) for efforts in this direction.

Spectral techniques are often used to check a candidate function for positive definiteness. Unfortunately, no spectral representations are available for the cases studied in this paper, making the problem more challenging. Even the definition of stationarity over networks is controversial (Baddeley et al., 2017). These facts motivated Anderes et al. (2020) to consider isotropic covariance functions over generalised networks. Hence, the covariance function depends on the distance between any two points located over the network.

To generalise a linear network, Anderes et al. (2020) proposed graphs with Euclidean edges: graphs where each edge is associated with an abstract set in bijective correspondence with a segment of the real line. This provides each edge with a Cartesian coordinate system to measure distances between any two points on that edge.

The connections between Gaussian processes and covariance functions allow complete determination of finite-dimensional distributions (Stein, 1999). Hence, the likelihood (or a related approximation) becomes analytically tractable. Alternatively, one can pursue approaches based on conditional distributions through the construction of Gaussian Markov random fields (Lindgren et al., 2011).

Recent approaches have addressed the challenge of covariance functions over networks through the perspective of Gaussian Markov random fields. Bolin et al. (2022) consider fractional stochastic partial differential equations on a compact metric graph. Their approach is the analogue of the approach taken by Lindgren et al. (2011) when working on Euclidean spaces or on manifolds. Processes constructed using these approaches are once differentiable over metric graphs, while the processes constructed by Ver Hoef and Peterson (2010) are not. A subclass of the processes considered by Bolin et al. (2022) has Markov properties, and, thus, the evaluation of the finite-dimensional distributions of the process can be done exactly and in a computationally efficient manner. We are not aware of any attempt to extend the work of Bolin et al. (2022) to space-time, and such an extension seems nontrivial.

It is difficult to compare approaches based on direct construction (i.e. on the covariance function) with those based on Gaussian Markov random fields. A large role is played by the graph topology. Over a linear network, it is possible to build a wealth of covariance models that allow for arbitrary level of mean square differentiability of the associated process. Further, over a linear network, it is possible to build models that parameterise the fractal dimension in a continuous fashion. We doubt this is so for the case of graphs with Euclidean edges, for which Anderes et al. (2020) provide models that are continuous but not differentiable at the origin. For the specific choice of a Matérn one-time differentiable model, the strategy provided by Bolin et al. (2022) is very successful. Our paper pursues the alternative path of building parametric classes of space-time covariance functions for the topology of graphs with Euclidean edges cross time.

1.2 The problems and our contributions

Anderes et al. (2020) constructed generalised networks using two alternative metrics for graphs with Euclidean edges. Further, they provided sufficient conditions for function classes to be positive definite when composed with these metrics.

Generalisations to the setting where generalised networks are considered as topological structures that do not evolve over time are provided by Tang and Zimmerman (2020). In particular, Tang and Zimmerman (2020) adapted a version of the Gneiting class (Gneiting, 2002b) to generalised networks cross linear time. In their formulation, the temporal distance is rescaled by spatial component, while Gneiting's original class proposed the opposite (rescaling spatial distance with temporal variables). Hence, further investigations are needed.

The existing literature on space-time covariance functions that incorporate a network as a spatial component is limited. Currently available options for covariance functions are insufficient in capturing the diverse range of interactions between space and time. This paper aims to address this gap by presenting a very general class of stationary nonseparable space-time covariance functions. Our proposed class of covariance functions accommodates a range of network structures, including generalised and linear networks, as well as Euclidean trees. This new class of covariance

functions allows for either the geodesic or resistance metrics in space (details given subsequently). We further allow for compactly supported covariance functions when working on Euclidean trees. While our primary focus is on linear time, we also provide a contribution that extends to circular (seasonal) time, subject to mild regularity conditions. Further, the proposed structure is general enough to describe several types of interactions between space and time.

Specifically, we provide the following contributions:

(a) We describe a procedure to compose two parametric classes of functions describing, respectively, spatial and temporal dependence. This combination provides a new class of stationary nonseparable space-time covariance functions.
(b) A special case of this class results in an adaptation of the Gneiting class of covariance functions, that had been originally proposed over planar surfaces (Gneiting, 2002b; Porcu and Zastavnyi, 2011).
(c) We consider covariance functions that are dynamically compactly supported over generalised networks with linear or circular time. This means that for every fixed temporal lag the covariance functions are compactly supported over balls with given radii. This allows for important computational gains.
(d) A simulation study addresses three aspects. We work on a generalised network and compare correctly specified and misspecified models. First, we show the impact of using the incorrect distance metric or covariance function in terms of likelihood estimation and predictive performance. Second, we assess parameter estimation error. Lastly, identifiability problems are inspected under a correct choice of the spatial metric and correct covariance function.
(e) We analyse a traffic accident point pattern dataset. We compare various models that differ in terms of distance metric and probability mass function. We find that the best model uses the network distance. Using this model, we explore posterior summaries, the space-time correlation function of random effects, and the posterior mean and standard deviation over space and time.

We further clarify how our contributions integrate or diverge with respect to previous literature:

1. *Generality of the approach*. Our approach is general concerning both the topology (graphs with Euclidean edges) and models (a wealth of covariance functions available). The contribution by Bolin et al. (2022) provides a more general topology (metric graphs) but time is not considered, and the method is pursued for the case of the Matérn covariance only.
2. *Dynamical embedding of compact supports on trees*. Tang and Zimmerman (2020) provide covariance functions that are compactly supported over a Euclidean tree in $\mathbb{R}^n$ embedded in $\mathbb{R}^{n+1}$. Our manuscript provides a constructive criticism of this construction and proposes covariance functions that are dynamically supported on the Euclidean tree cross time. Specifically, for every fixed temporal lag u_0 there exists a positive function ψ such that the *spatial* margin on the tree is compactly supported over a ball embedded in $\mathbb{R}^n$ (for n being linked to the number of leaves) with radius $\psi(u_0)$.
3. *Special topologies involving flows*. Ver Hoef and Peterson (2010) developed covariance models for stream flows, by invoking the moving average representation (in a continuous fashion) of a given random field. Such a representation is a shifted integral of a deterministic function, g, against a white noise, W. The resulting covariance function is the auto-convolution of the function g. The stream network topology is additionally complicated by the flow (which induces the so-called *tail up* or *tail down* models). The last fact is especially relevant because the resulting models are not, in general, isotropic, while our modelling strategy attains isotropic models. At the same time, the models provided in our paper can be used as building blocks for more sophisticated scenarios, such as nonisotropy and nonstationarity. What makes the modelling strategy in Ver Hoef and Peterson (2010) challenging is that one must know the *root g* that allows for an explicit algebraically closed form of the resulting covariance. A technical approach to this problem is taken by Ehm et al. (2004). On the other hand, Tang and Zimmerman (2020) note that the topology provided by Ver Hoef and Peterson (2010)

corresponds to a directed Euclidean tree. Accordingly, they take a scale-mixture approach on directed Euclidean trees and prove that an exponential tail-down model as presented in Ver Hoef and Peterson (2010) is the one and only one that is directionless (i.e. isotropic) and is thereby a bridge between models presented in our paper as well as those in Tang and Zimmerman (2020). To generalise the approach in Ver Hoef and Peterson (2010) to space-time, Tang and Zimmerman (2020) use the analogue of the product-sum model (De Cesare et al., 2001) with space represented by a directed Euclidean tree.

The results obtained in this paper are technical and require substantial background on quasi-metric spaces, isometric embeddings, graphs with Euclidean edges, and harmonic analysis. Hence, the exposition focuses on describing the main ideas and illustrating their implications through practical examples, while keeping a statistical language and deferring the technical part to the Online Supplementary Material. The plan of the paper is the following. Section 2 contains a succinct statistical background and an illustration about graphs with Euclidean edges. Section 3 describes the main idea and construction for the new class of space-time covariance functions. Section 3.1 guides the reader through practical examples. We also discuss previously proposed examples in concert with models that are not valid. A simulation study in Section 4 illustrates the practical implementation of our model, explores the effects of model misspecification on estimation and prediction, and studies the identifiability of model parameters under the correctly specified model. Section 5 analyses a traffic accident point pattern dataset and compares various models that differ in terms of distance metric and probability mass function. A short discussion concludes the paper.

The Supplemental Material contains the following. All theorems referenced are stated and proved in the Supplemental Material. All Supplemental Material sections have an 'S' prefix, while tables and equations are given number labels as if they followed subsequently in the manuscript. Online Supplementary Material, Section S1 presents a mathematical background needed to understand the proofs. Online Supplementary Material, Section S2 provides formal statements and their proofs to justify the general ideas illustrated in Section 3. Online Supplementary Material, Section S3 reports tables for constructing a wealth of practical examples of new covariance functions.

2 Background and notation

A linear network is commonly understood as the union of finitely many line segments in the plane, where different edges only possibly intersect with each other at one of their vertices. The upper-left part of Figure 1 depicts an abstract drawing for a linear network. More sophisticated pictures are available in the literature, but a simplified version is provided here. Both random fields and point processes over linear networks have been considered in the literature. For continuously indexed random fields (not evolving over time) the reader is referred to Anderes et al. (2020). For point processes over networks, a standard reference is Baddeley et al. (2021).

A network (or equivalently, a graph), $\mathcal{G}$, is comprised of two components: a set of nodes (or vertices) denoted by $\mathcal{V}$, and a set of edges denoted by $\mathcal{E}$. Clearly, a linear network as defined above is a special case of network. Generalised networks as defined by Anderes et al. (2020) allow for non-linear edges. The problem with nonlinear edges stands mainly in how to measure distances between any pair of points belonging either to the set of vertices or to the edges. The problem is solved by Anderes et al. (2020) who propose graphs with Euclidean edges: those are sophisticated topological structures that allow *distances* in the following way. Each graph has a collection of bijection mappings, such that each edge in the graph is mapped into an open interval (see bottom-right of Figure 1), and every pair of vertices connected by the edge is mapped into two points at the extremes of the same interval. This creates a Euclidean system with an orientation and a suitable way to measure distances. Rigorous definitions of these topological structures are given in Online Supplementary Material, Section S1. The bottom-right part of Figure 1 depicts a typical graph with Euclidean edges. We note that the edge e is mapped into an open interval and that the two vertices (u, v) are mapped into the endpoints of the interval, denoted $(\underline{e}, \overline{e})$. The path merging $\underline{e}$ and $\overline{e}$ is highlighted in orange in the picture. The geodesic distance is the length of such a path and is denoted d_G throughout the manuscript. More accurately, the geodesic distance, d_G, is

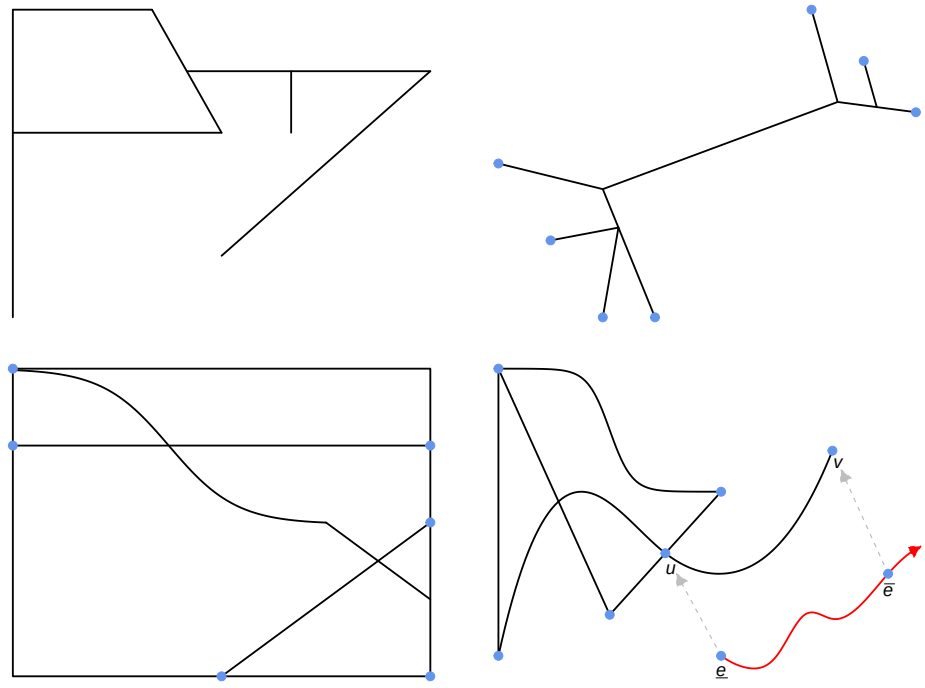

Figure 1. *Upper-Left*: an example of linear network; *Upper-Right*: A Euclidean tree with seven leaves (blue dots), which may represent a stream network; *Bottom-Left*: A graph with Euclidean edges that may represent a road traffic network; crosses between edges with no vertices represent bridges or tunnels. *Bottom-Right*: another graph with Euclidean edges; the bijection mapping the vertices u and v into $\underline{e}$ and $\overline{e}$ and the edge e into the open interval $(\underline{e}, \overline{e})$ gives a Euclidean system with orientation and a way to measure distances.

the length of the shortest path merging any pair of points belonging to $\mathcal{G}$. Another example of graph with Euclidean edges is provided in the bottom-left part of the same figure.

Figure 2 further illustrates how distances are computed over a graph. We note that to each edge, e and e', are associated two (possibly) different mappings, φ_e and $\varphi_{e'}$. To calculate the geodesic distance between a vertex, v_2 being one extreme of the edge e', and a point v_0 lying somewhere on the edge e, we sum the length of two paths. The first path is highlighted in green from the purple curve. Note: (a) the purple curve is a Euclidean coordinate system, with an orientation; (b) for each edge, e', a bijection $\varphi_{e'}$ is assigned; (c) to the vertices v_2 and v_1 we assign respectively the points $\overline{e'} = \varphi_{e'}(v_1)$ and $\underline{e'} = \varphi_{e'}(v_2)$; the length of the path is measured through $|\varphi_{e'}(v_1) - \varphi_{e'}(v_2)|$; as a result, we have that $d_G(v_0, v_2) = |\varphi_e(v_0) - \varphi_e(v_1)| + |\varphi_{e'}(v_1) - \varphi_{e'}(v_2)|$.

Anderes et al. (2020) provided an accurate description of the technical conditions on a graph to have Euclidean edges. Distance consistency is one of those, and we refer the reader to Figure 1 in Anderes et al. (2020) for an illustration. Some graphs are *forbidden* with respect to the geodesic distance. Examples are provided by Anderes et al. (2020), and Figure 3 shows on the left side an example of a forbidden graph. Distance is inconsistent because the geodesic distance (length of the green arc) is different than the within-edge distance (length of the orange arc), which is calculated through $|\varphi_e(v_1) - \varphi_e(v_2)|$. On the other hand, the right-hand side shows an example where the geodesic and the within-edge distances coincide. Such a graph is indeed consistent with respect to the geodesic metric.

The upper-right part of Figure 1 depicts a Euclidean tree, a special case of a graph with Euclidean edges. It is a tree-like graph (which is planar). Vertices of a Euclidean tree that are connected to one edge only are called leaves. As noted by Tang and Zimmerman (2020), an arbitrary point x belongs to $\mathcal{G}$ when $x \in \mathcal{V} \cup \bigcup_{e \in \mathcal{E}}$. As in their paper, we assume that the topological structure of $\mathcal{G}$ does not evolve over time.

Our paper considers weakly stationary random fields $\{Z(\boldsymbol{x}, t), (\boldsymbol{x}, t) \in \mathcal{G} \times \mathcal{T}\}$, with $\mathcal{G}$ being a graph (with Euclidean edges, an Euclidean tree, or a linear network) and $\mathcal{T}$ describing *time*,

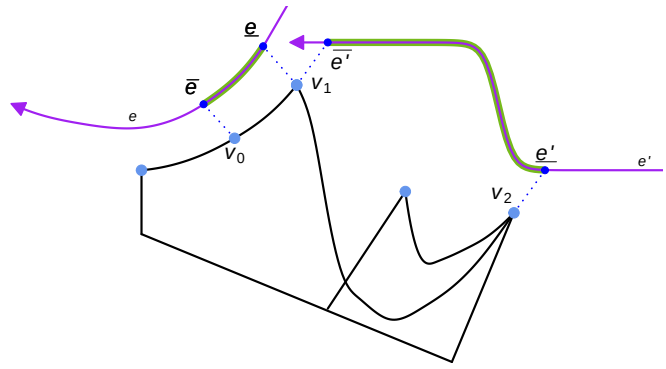

Figure 2. An illustration on how to compute the geodesic distance over a graph with Euclidean edges, namely $d_G(v_0, v_2) = |\varphi_e(v_0) - \varphi_e(v_1)| + |\varphi_{e'}(v_1) - \varphi_{e'}(v_2)|$.

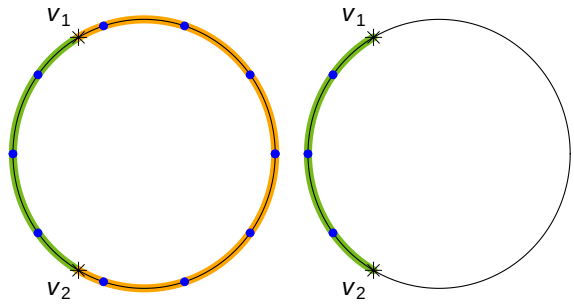

Figure 3. An example of a forbidden (left) and allowed (right) Euclidean cycle.

that can be either linear (time is the whole real line, $\mathbb{R}$), or circular (seasonal time on the circle, denoted $\mathbb{S}$). This work focuses on the second-order properties of Z, with special emphasis on the covariance function, being a linear measure of association between the random variable $Z(\boldsymbol{x}, t)$ at point $\boldsymbol{x} \in \mathcal{G}$ and time $t \in \mathcal{T}$, and the random variable $Z(\boldsymbol{x}', t')$ at point $\boldsymbol{x}' \in \mathcal{G}$ and time $t' \in \mathcal{T}$. We assume that

$$\text{cov}\{Z(\boldsymbol{x}, t), Z(\boldsymbol{x}', t')\} = G(\text{distance}_{\mathcal{G}}(\boldsymbol{x}, \boldsymbol{x}'), \text{separation}_{\mathcal{T}}(t, t')), \tag{2.1}$$

for some suitable function, G. Covariance functions are positive definite, that is $\sum_{i,j=1}^{N} a_i \text{cov}\{Z(\boldsymbol{x}_i, t_i), Z(\boldsymbol{x}_j, t_j)\}a_j \geq 0$ for any arbitrary system $\{a_i\}_{i=1}^{N}$ of real constants and any arbitrary choice of points $\{(\boldsymbol{x}_i, t_i)\}_{i=1}^{N}$. For a random field defined over $\mathcal{G}$ and not evolving over time, we define the variogram (denoted γ throughout) as the variance of the increments of $Z(\boldsymbol{x})$ with respect to $Z(\boldsymbol{x}')$, for $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{G}$: $\gamma(\boldsymbol{x}, \boldsymbol{x}') = \text{var}\{Z(\boldsymbol{x}) - Z(\boldsymbol{x}')\}$. The variogram can also be used as an alternative metric, termed *resistance* metric in Anderes et al. (2020), and denoted d_R throughout. The resistance metric is actually the variogram of a special class of random fields (see Anderes et al., 2020, for details). The function G will be equipped with either the geodesic or the resistance distance, and with the difference (if linear time) or the geodesic over the circle (if circular time).

3 Main results

We discuss general construction here. We consider two parametric classes of functions. Let p and k be two positive integers. Then, we define

$$\mathcal{D}_{\boldsymbol{\theta}} := \{\varphi(x \mid \boldsymbol{\theta}), x \geq 0, \boldsymbol{\theta} \in \mathbb{R}^p\} \quad \text{and} \quad \mathcal{H}_{\boldsymbol{\vartheta}} := \{\psi(t \mid \boldsymbol{\vartheta}), t \geq 0, \boldsymbol{\vartheta} \in \mathbb{R}^k\}, \tag{3.1}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ are parameter vectors. Our purpose is to use elements of the classes $\mathcal{D}_{\boldsymbol{\theta}}$ and $\mathcal{H}_{\boldsymbol{\vartheta}}$ to create a new class of space-time covariance functions with a wealth of practical examples and interactions between space and time. Specifically, we propose the following construction. For two given functions φ and ψ belonging to the class $\mathcal{D}_{\boldsymbol{\theta}}$ and $\mathcal{H}_{\boldsymbol{\vartheta}}$, respectively, we define the function $G_{\alpha,\beta}(\cdot,\cdot \mid \boldsymbol{\theta},\boldsymbol{\vartheta}):[0,\infty)^2 \to \mathbb{R}$:

$$G_{\alpha,\beta}(x,t \mid \boldsymbol{\theta},\boldsymbol{\vartheta}) = \frac{1}{\psi(t \mid \boldsymbol{\vartheta})^{\alpha}} \varphi\left(\frac{x}{\psi(t \mid \boldsymbol{\vartheta})^{\beta}} \mid \boldsymbol{\theta}\right), \quad x,t \geq 0. \tag{3.2}$$

The parameters α and β have been left intentionally outside the vectors $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ because of their physical interpretation, clarified subsequently. Hence, the problem is to find conditions on the functions φ, ψ, and on the values of α and β such that

$$\operatorname{cov}\{Z(\boldsymbol{x},t), Z(\boldsymbol{x}',t')\} = G_{\alpha,\beta}(d_{\cdot}(\boldsymbol{x},\boldsymbol{x}'), |t-t'| \mid \boldsymbol{\theta},\boldsymbol{\vartheta}), \quad \boldsymbol{x},\boldsymbol{x}' \in \mathcal{G},\ t,t' \in \mathbb{R} \tag{3.3}$$

is a positive definite function. Here, we use $d_{\cdot}$ whenever we do not wish to specify any choice between d_G and d_R. We now clarify the role of the parameters α and β. When α and β are both positive, $G_{\alpha,\beta}$ corresponds to a functional form that was originally proposed by Gneiting (2002b) for *space* being the d-dimensional Euclidean space. Several generalisations of this class are summarised in Porcu, Furrer, et al. (2020). When β is positive, the spatial distance is rescaled by temporal dependence. When β is negative, then the function acting on temporal dependence multiplies the spatial distance.

The functional and parametric conditions ensuring $G_{\alpha,\beta}$ to become a covariance function are carefully explored and justified in Online Supplementary Material, Section S2. As an example, the function $\varphi(x \mid \boldsymbol{\theta}) = \sigma^2 \mathrm{e}^{-x/a}$ is suitable, with $\boldsymbol{\theta} = (\sigma^2, a)^{\top} \in \mathbb{R}_+^2 \subset \mathbb{R}^2$, and with $\top$ denoting the transpose operator. A simple choice from the class $\mathcal{H}_{\boldsymbol{\vartheta}}$ is the function

$$\psi(t \mid \boldsymbol{\vartheta}) = (1+t^a)^b, \quad t \geq 0, \tag{3.4}$$

with $\boldsymbol{\vartheta} = (a,b)^{\top}$. Here, $a \in (0,2]$ and b belongs to the interval $(0,1]$. We provide a simplified exposition here. For the sake of simplicity, we only report one of the results that can be found in the Online Supplementary Material. The relevant classes of functions that are used in the result below are carefully described in Online Supplementary Material, Section S1.

Theorem 1 Let $\mathcal{G}$ be a graph with Euclidean edges. Let $G_{\alpha,\beta}$ be the mapping defined through equation (3.2). Let $\varphi(\cdot \mid \boldsymbol{\theta})$ be a parametric family of Stieltjes functions. Let $\alpha \geq 1$ and $\beta \in (0,1]$. Then,

(a) if time is linear ($\mathcal{T} = \mathbb{R}$) then $G_{\alpha,\beta}(d_R(\cdot,\cdot), |\cdot|^2)$ is positive definite provided $\psi(\cdot \mid \boldsymbol{\vartheta})$ is a parametric family of Bernstein functions;
(b) if time is circular ($\mathcal{T} = \mathbb{S}$) then $G_{\alpha,\beta}(d_R(\cdot,\cdot), d_G(\cdot,\cdot))$ is positive definite provided $\psi(\cdot \mid \boldsymbol{\vartheta})$ is the restriction to the interval $[0,\pi]$ of a parametric family of Bernstein functions;
(c) if $\mathcal{G}$ is a graph with Euclidean edges that forms a finite sequential 1-sum of Euclidean cycles and trees, then for both cases above the resistance metric, d_R, can be replaced by the geodesic distance, d_G.

3.1 Examples from the new class of space-time covariance functions

We show here how the class of space-time covariance functions proposed in (3.3) can be adapted for a wealth of practical situations and different interactions between space and time. To select any example from this class, the practitioner should take into account:

- The reference space: a linear network, Euclidean tree, or graph with Euclidean edges;
- The temporal component: linear or circular time;

- The type of spatial distance: the geodesic, d_G, or the resistance metric, d_R. In turn, this choice depends on the reference space and on the function φ used for the construction (3.2). Details are in Theorem 1 and Online Supplementary Material, Theorem 2(S);
- The fact that the function φ from the class $\mathcal{D}_{\boldsymbol{\theta}}$ is strictly positive on the positive real line, or compactly supported.

Some examples follow.

Example 1 We consider a graph with Euclidean edges, $\mathcal{G}$, equipped with the resistance metric, d_R. We consider the function $x \mapsto \varphi(x \mid \boldsymbol{\theta}) = \sigma^2\{1 + (x/c_S)^{b_S}\}^{-\delta_S}$, for $x \geq 0$. Here, the positive parameter c_S rescales spatial distance, while the parameters $b_S \in (0, 1]$ and $\delta_S > 0$ are related to fractal dimension and long memory of the associated random process. We can use a rescaled version of the function $\psi(\cdot/c_T)$, for $c_T > 0$ a temporal scale with parameter vector $\boldsymbol{\vartheta} = (a_T, b_T)^\top$ as defined through (3.4). Throughout, we fix $b_T = 1$ with no loss of generality. Hence, a direct application of Theorem 1 ensures that

$$G_{\alpha,\beta}(d_R(\boldsymbol{x},\boldsymbol{x}'), |t-t'| \mid \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \frac{\sigma^2}{\left\{1+\left(\frac{|t-t'|}{c_T}\right)^{a_T}\right\}^{\alpha}} \left(1 + \left[\frac{d_R(\boldsymbol{x},\boldsymbol{x}')}{c_S\left\{1+\left(\frac{|t-t'|}{c_T}\right)^{a_T}\right\}^{\beta}}\right]^{b_S}\right)^{-\delta_S}, \tag{3.5}$$

for $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{G}$ and $t, t' \in \mathbb{R}$, is a valid covariance function provided $\alpha > 0$ and $\beta \in (0, 1]$. A similar example can be created by replacing the function φ above with the function $\varphi(x \mid \boldsymbol{\theta}) = \sigma^2\{1 - x^{b_S\delta_S}(1 + x^{b_S})^{-\delta_S}\}$, $x \geq 0$. Using the function $\psi(t \mid \boldsymbol{\vartheta}) = (\eta + t^{a_T})$, $\eta > 0$, and after appropriate rescaling over space and time, we get

$$G_{\alpha,\beta}(d_R(\boldsymbol{x},\boldsymbol{x}'), |t-t'| \mid \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \frac{\sigma^2}{\left\{\eta+\left(\frac{|t-t'|}{c_T}\right)^{a_T}\right\}^{\alpha}} \times \left\{1 - \left[\frac{d_R(\boldsymbol{x},\boldsymbol{x}')}{c_S\left\{\eta+\left(\frac{|t-t'|}{c_T}\right)^{a_T}\right\}^{\beta}}\right]^{b_S\delta_S} \left(1 + \left[\frac{d_R(\boldsymbol{x},\boldsymbol{x}')}{c_S\left\{\eta+\left(\frac{|t-t'|}{c_T}\right)^{a_T}\right\}^{\beta}}\right]^{b_S}\right)^{-\delta_S}\right\}, \tag{3.6}$$

for $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{G}$ and $t, t' \in \mathbb{R}$, is a valid covariance function provided $\alpha > 0$ and $\beta \in (0, 1]$. For both examples, the resistance metric might be replaced with the geodesic, and the reader is referred to point 3 in Theorem 1 for details.

Example 2 Covariance functions with compact support play an important role when the graph is a Euclidean tree with a given number of leaves. Details are provided through Online Supplementary Material, Theorem 2(S) in Section S2. An illustration is provided below. We consider the function

$$\varphi(x \mid \boldsymbol{\theta}) = \sigma^2\left(1 - \frac{x}{c_S}\right)_+^{\nu_S}, \quad x \geq 0,$$

where $c_S > 0$, ν is positive and has a lower bound that is specified through Online Supplementary Material, Theorem 2(S). Here, $(x)_+$ stands for the positive part of the real number x. The parameter c_S determines the support of the function, because φ is identically equal to zero whenever $x \geq c_S$. Specifically, we have for example that

$$G_{\alpha,\beta}(d_\cdot(x, x'), |t - t'| \mid \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \frac{\sigma^2}{\left\{1 + \left(\frac{|t - t'|}{c_T}\right)^{a_T}\right\}^{\alpha}} \left[1 - \frac{d_\cdot(x, x')}{c_S\left\{1 + \left(\frac{|t - t'|}{c_T}\right)^{a_T}\right\}^{\beta}}\right]_+^{\nu_s} \tag{3.7}$$

is a valid construction. Indeed, this is possible if the graph $\mathcal{G}$ is a Euclidean tree with a given number of leaves, m. The parameters ν_S and α depend linearly on m (see Online Supplementary Material, Theorem 2(S)), and β belongs to the interval [0, 1). Also, the geodesic distance can be replaced by the resistance metric with no harm (see Online Supplementary Material, Theorem 2(S)). An important feature of the covariance function in equation (3.7) is that it is dynamically compactly supported. That is, for every fixed time t, t', the function $G_{\alpha,\beta}$ is compactly supported over a ball embedded in $\mathcal{G}$ with radius $\psi(|t - t'| \mid \boldsymbol{\vartheta})$. This feature has been well studied in spatial statistics, and the reader is referred to Porcu, Bevilacqua, et al. (2020) for a modelling perspective, as well as to Bevilacqua et al. (2019) for the implications of using compact support for modelling, estimation and prediction under the so-called infill asymptotic framework (Stein, 1999).

Example 3 When the parameter β in the covariance $G_{\alpha,\beta}$ is negative, the spatial distance $d_\cdot$ is multiplied (no longer rescaled) by temporal dependence. For instance, the function

$$G_{\alpha,-1}(d_\cdot(\boldsymbol{x}, \boldsymbol{x}'), |t - t'| \mid \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \frac{\sigma^2}{\psi(|t - t'| \mid \boldsymbol{\vartheta})^{\alpha}} \exp\left\{-\frac{d_\cdot(\boldsymbol{x}, \boldsymbol{x}')}{c_S} \psi(|t - t'| \mid \boldsymbol{\vartheta})\right\} \tag{3.8}$$

is a valid covariance function if $\alpha \geq 1$ and provided ψ satisfies the conditions in Online Supplementary Material, Theorem 3(S). Many other examples of this kind can be obtained using Online Supplementary Material, Theorem 3(S) in concert with suitable choices from Online Supplementary Material, Tables 1(S) and 2(S).

3.2 When time is circular

If we assume time to be circular (as in Mastrantonio et al., 2019; Shirota and Gelfand, 2017; White and Porcu, 2019), then the Euclidean distance needs to be replaced by the geodesic distance over the circle. Theorem 1, Online Supplementary Material, Theorems 2(S) and 3(S) provide technical conditions for when the geodesic distance can be used in the function $\psi(\cdot \mid \boldsymbol{\vartheta})$. It is worth mentioning that the geodesic distance over the circle has range $[0, \pi]$, so that the function ψ is restricted to this interval. This is stated accurately in the relevant propositions.

Online Supplementary Material, Theorem 4(S) provides a different construction that is based on half-spectral inversion. We do not enter mathematical details, but note here that such a construction allows for examples that cannot be covered through Theorem 1, Online Supplementary Material, Theorems 2(S) and 3(S). In particular, all the examples that have been previously introduced do not allow for negative spatial dependencies. The function

$$G_{1,1}(d_R(\boldsymbol{x}, \boldsymbol{x}'), d_G(t, t') \mid \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \sigma^2 \left\{ \frac{1-\varepsilon}{1-\varepsilon\psi(d_R(\boldsymbol{x}, \boldsymbol{x}') \mid \boldsymbol{\vartheta}) \cos d_G(t, t')} \right\}^{\tau},$$

for $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{G}$, $t, t' \in \mathbb{S}$, is a valid covariance function for any graph with Euclidean edges cross circular time. Here, ψ needs to be a positive definite function over the circle with the additional requirement that $\psi(0 \mid \boldsymbol{\vartheta}) = 1$. The parameter ε belongs to the open interval $(0, 1)$. More examples are reported in Online Supplementary Material, Table 3(S).

3.3 Previously proposed models

Menegatto et al. (2020) considered the more general setting of quasi-metric spaces and provided sufficient conditions for the structure $G_{\alpha,\beta}(|\cdot|, \sigma(\cdot, \cdot))$ (notice that the arguments are exchanged here) to be positive definite. Here, σ is an arbitrary quasi-metric. Tang and Zimmerman (2020) noticed this fact and considered the pair $(\mathcal{G}, d_R)$ as a quasi-metric space. As a result, the Menegatto-Porcu-Oliveira construction can be adapted to a covariance function $G_{\alpha,\beta}(|\cdot|, d_R(\cdot, \cdot))$ where the temporal separation is rescaled by spatial dependence. This is unusual in spatial statistics, and for a constructive criticism the reader is referred to Porcu et al. (2018), with the references therein.

For the case of Euclidean trees with a given number of leaves, Tang and Zimmerman (2020) proposed what they term *metric models*. Let $\rho_{1,n}$ be the ℓ_1 distance in $\mathbb{R}^n$. For a function C such that $C(\rho_{1,n+1}(\cdot, \cdot))$ is positive definite, arguments in Theorem 4 of Anderes et al. (2020) show that $C(d_{\cdot}(\cdot, \cdot) + |\cdot|)$ is positive definite over a Euclidean tree cross the real line. The construction is clearly reminiscent of zonal anisotropy in geostatistics, whose adaptation to the space-time setting has received constructive criticism, and we refer the reader to Chilès and Delfiner (2012), Gneiting (2002a), Stein (2005) and Porcu and Zastavnyi (2011) amongst others.

3.4 Related constructions and forbidden models

A *partially forbidden* model. The Matérn class of functions has been the cornerstone of spatial statistics for a long time. We refer the reader to Stein (1999) and more recently to Bevilacqua et al. (2022) for a thorough account. For a very recent review that embraces several disciplines, the reader is referred to Porcu et al. (2023). We define it here through

$$\mathcal{M}_{\nu}(x) = \frac{2^{1-\nu}}{\Gamma(\nu)} x^{\nu} \mathcal{K}_{\nu}(x), \quad x \geq 0, \tag{3.9}$$

where $\mathcal{K}_{\nu}$ is a modified Bessel function of the second kind of order $\nu > 0$. The parameter ν serves as an index of mean square differentiability for the associated random process. Given the massive use of the Matérn family in spatial statistics, one might be tempted to choose

$$\varphi(x \mid \boldsymbol{\theta}) = \sigma^2 \mathcal{M}_{\nu_S}\left(\frac{x}{c_S}\right), \quad x \geq 0,$$

with $\boldsymbol{\theta} = (\sigma^2, c_S, \nu_S)^{\top}$, where the three parameters index the variance, the spatial scale, and the smoothness, respectively. A similar calculation as in Example 1 yields

$$G_{\alpha,\beta}(d_R(\boldsymbol{x}, \boldsymbol{x}'), |t-t'| \mid \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \frac{\sigma^2}{\left\{1 + \left(\frac{|t-t'|}{c_T}\right)^{a_T}\right\}^{\alpha}} \mathcal{M}_{\nu_S}\left(\frac{d_R(\boldsymbol{x}, \boldsymbol{x}')}{c_S \left\{1 + \left(\frac{|t-t'|}{c_T}\right)^{a_T}\right\}^{\beta}} \right). \tag{3.10}$$

Unfortunately, arguments in Theorem 1 in Anderes et al. (2020) in concert with Theorem 1 show that the parameter ν_S is restricted to the interval $(0, 1/2]$ to ensure positive definiteness. For such an interval, the associated process is continuous but not mean square differentiable. Hence, this covariance is not suitable to index spatial smoothness.

Forbidden models and unclear cases. For $\mathcal{G}$ being any general graph with Euclidean edges, Online Supplementary Material, Section S2 shows that the functions φ and ψ involved in the composition $G_{\alpha,\beta}$ as in equation (3.2) are both strictly positive. While Online Supplementary Material, Theorem 2(S) shows that covariance functions can attain negative values for a relatively small number of leaves, Corollary 3 in Anderes et al. (2020) proves that covariance functions on trees with *any* number of leaves must be strictly positive, or identically equal to zero after a given lag. Further, the condition of non-negativity is only necessary. Sufficient conditions based on isometric embeddings from the metric space $(\mathbb{R}^n, \rho_{1,n})$ into the quasi-metric space $(\mathcal{G}, d_\cdot)$, where the number of leaves is related to the dimension n where the original space is defined can be inferred from Zastavnyi (2000).

An important implication for data analysis is that the model in equation (3.7) is definitely not suitable for Euclidean trees with a large number of leaves. In fact, the condition $v \geq 2n-1$ from Online Supplementary Material, Theorem 2(S), with $n = \lceil m/2 \rceil$ and m being the number of leaves, implies that for m large enough the kernel almost vanishes except at the origin. Such an inconvenience is clearly shared by the metric models that have been introduced by Tang and Zimmerman (2020) to analyse data over a tree with a given number of leaves. Example (3.8) overcomes this inconvenience, as the function φ in the composition is not related to the number of leaves. Hence, the example in equation (3.8) is more recommendable to deal with Euclidean trees with a large number of leaves.

Finding covariances with negative values on a general graph with Euclidean edges is elusive. Theorem 1 in Anderes et al. (2020) provides a sufficient condition, and all the functions satisfying such a condition are strictly positive. It is unclear whether any choice of function φ attaining negative values can preserve positive definiteness. Emery and Porcu (2022) show that this is doable when *space* is the metric space $(\mathbb{R}^n, \rho_{2,n})$, with $\rho_{2,n}$ denoting the Euclidean distance in $\mathbb{R}^n$. Apparently, the elegant isometric argument in Anderes et al. (2020) cannot be used in this case, and this remains an open problem.

New models commuted from old literature. The proofs of Theorem 1, Online Supplementary Material, Theorems 2(S) and 3(S) show that typical scale mixtures arguments can be used to adapt space-time covariance functions that have been proposed for the setting of the metric space $(\mathbb{R}^n \times \mathbb{R}, \rho^2_{2,n}, \rho_{1,n})$. Here we list the most prominent constructions:

1. The *quasi-arithmetic* class (Porcu et al., 2010);
2. The scale mixtures as in Fonseca and Steel (2011), Apanasovich and Genton (2010), and Schlather (2010);
3. Other scale-mixture-based constructions as in Porcu et al. (2006), Porcu et al. (2007), Porcu and Mateu (2007), and Alegría et al. (2019).

Other popular constructions can be adapted from earlier literature. For instance, Peron et al. (2018) proposed linear combinations of products of covariance functions defined over graphs with temporal covariance functions. They provided conditions for at least one weight in the linear combination to be negative.

4 Simulation study

Our simulation study considers data simulated over a river network to analyse the effects of model misspecification in terms of both distance metric and covariance function, as well as the estimation and identifiability of model parameters used to generate data. Specifically, we use a subset of $n_s = 50$ sites on the Clearwater River Basin in Idaho, USA (see Figure 4). These locations are derived from data used by Isaak et al. (2018), available at https://www.researchgate.net/publication/325933910_Principal_components_of_thermal_regimes_in_mountain_river_networks. To emphasise the importance of how distance is defined in such settings, we calculate the Euclidean distance and plot it against the distance over the river network in Figure 4. Because river networks generally present unique challenges with flow direction and river connections (Ver Hoef et al., 2006), we emphasise that we only use this network structure as an illustration. For each location, we simulate 10 random time points, distributed uniformly between 0 and 1, yielding 500 data points.

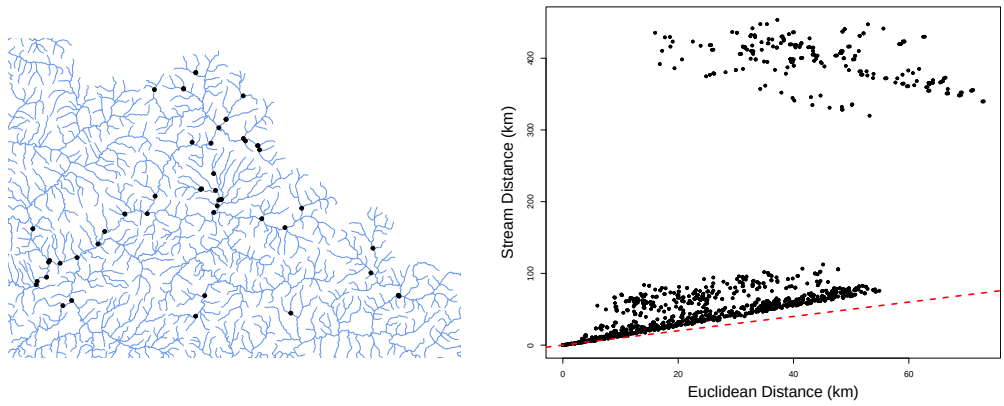

Figure 4. (Left) Locations of sites (black points) on the river network (blue colour) used for the simulation study. (Right) Comparison of Euclidean distance and network (stream) distance for the sites used in the simulation study. The red dashed line has a slope of one and marks equality.

We consider four experiments in this study, and, for each experiment, we simulate $n_{sim} = 1000$ datasets through

$$\boldsymbol{y} \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma} + \tau^2\mathbf{I}). \tag{4.1}$$

Here $\boldsymbol{\Sigma}$ is a covariance matrix with elements determined by equation (3.5) which we subsequently denote by $\mathbb{T}$ for the *true* covariance model. Specifically, we use the network geodesic distance d_G and fix $\alpha = 2$, $a_T = 1$, $\delta_S = 2$. In the simulation experiments, we treat c_T, c_S, σ^2, and τ^2 as unknown parameters to be estimated. The four simulation experiments differ by varying the spatial range parameter, $c_S = 20$, 50, 100, 200 km, depending on the simulation experiment. For all simulations, we use $c_T = 0.2$, $\sigma^2 = 0.9$, and $\tau^2 = 0.1$. For every simulated dataset, we fit models using the *true* model, as well as two competitors, $\mathbb{C}_i$, $i = 1, 2$, defined as

($\mathbb{C}_1$) The same covariance model $\mathbb{T}$, but replacing the network geodesic metric, d_G, with the Euclidean distance;
($\mathbb{C}_2$) The covariance model in equation (3.6), with $\alpha = 2$, $a_T = 1/4$, $\delta_S = 1/2$, $\eta = 1/2$, and using the geodesic distance d_G. As with $\mathbb{T}$, τ^2, σ^2, c_S, and c_T are unknown.

All models are fit using maximum likelihood estimation. Because τ^2, σ^2, c_S, and c_T are strictly positive, we maximise the likelihood as a function of the parameters on the log-scale.

We examine the results of this simulation study in several ways. To verify that model performance is best for $\mathbb{T}$, in Section 4.1, we present the proportion of times each covariance example had the highest likelihood. In addition, we compare models in terms of out-of-sample predictive performance. To verify that using the correct distance metrics improves parameter estimation, in Section 4.2, we compare the parameter estimates under $\mathbb{T}$ and $\mathbb{C}_1$ (these models only differ by the distance metric used for the spatial component). To determine whether we can recapture the true parameters under the correctly specified model, we further compare the estimated model parameters for $\mathbb{T}$ to the true parameters in Section 4.2. The code for these simulations is available at https://github.com/philawhite/network-time.

4.1 Model comparison

For every simulated dataset, we fit the models that differ in terms of covariance function ($\mathbb{T}$, $\mathbb{C}_1$ or $\mathbb{C}_2$) but have the same model form (4.1). Using maximum likelihood estimates, we calculate the log-likelihood for each model to identify the covariance example with the highest likelihood. In Table 1, we present the proportion of simulations where each example had the highest likelihood,

Table 1. The proportion of times each covariance example had the highest likelihood for the simulated datasets

c_S	$\mathbb{T}$	$\mathbb{C}_1$	$\mathbb{C}_2$
20 km	**0.917**	0.083	0.000
50 km	**0.969**	0.031	0.000
100 km	**0.988**	0.012	0.000
200 km	**0.997**	0.003	0.000

Note. Best in bold font.

Table 2. The average predictive mean squared error (PMSE) and the average continuous ranked probability score (CRPS), averaged over all simulations

c_S	Criterion	$\mathbb{T}$	$\mathbb{C}_1$	$\mathbb{C}_2$
20 km	PMSE	**0.567**	0.578	0.651
	CRPS	**0.415**	0.419	0.448
50 km	PMSE	**0.447**	0.465	0.535
	CRPS	**0.369**	0.376	0.406
100 km	PMSE	**0.361**	0.384	0.441
	CRPS	**0.332**	0.342	0.369
200 km	PMSE	**0.288**	0.313	0.353
	CRPS	**0.297**	0.309	0.330

Note. Best in bold font.

calculated for each experiment ($c_S = 20$, 50, 100, 200 km). Even for relatively short spatial range parameters, the model $\mathbb{T}$ was chosen 92% of the time; however, $\mathbb{T}$ was chosen more frequently as the range parameter c_S increased. Thus, we find that as the range parameter increases (i.e. the persistence of spatial correlation increases), it was more important to use the correctly specified network metric.

To compare the predictive performance of $\mathbb{T}$, $\mathbb{C}_1$, and $\mathbb{C}_2$, we randomly select 20% of every simulated dataset to use as a test set. Using the remaining (training) data, we estimate parameters using maximum likelihood estimation. Based on the maximum likelihood estimates, we predict the values of the test data, conditioned on the training data. For every simulation, we calculate predictive mean squared error (PMSE). We also use the average of continuous ranked probability scores (CRPS), $CRPS(F, y_i) = \mathbb{E}_F|Y - y_i| - \frac{1}{2}\mathbb{E}_F|Y - Y'|$, a strictly proper scoring rule that considers the fit of the whole predictive distribution F to an observed value y_i (Epstein, 1969; Gneiting and Raftery, 2007; Matheson and Winkler, 1976). Lastly, we average the performance over all simulations. These results are presented in Table 2.

In all cases, the correct model $\mathbb{T}$ was the best predictive model in terms of CRPS and PMSE. For the smallest spatial range parameter, the correct model $\mathbb{T}$ was only marginally better than the best of the misspecified models. However, for larger spatial range parameters, the relative advantage of the *true* model was larger. As in the likelihood comparison in Tables 1, 2 shows that as the persistence of spatial correlation increases, it was more important to use network geodesic.

4.2 Parameter estimation

In this section, we assess the estimation differences depending on the distance metric used, as well as parameter recovery. As discussed, the data were simulated using $\mathbb{T}$, and we compare the maximum likelihood estimates for $\mathbb{T}$ and $\mathbb{C}_1$ to the true value using distance measures, histograms of estimated parameters, and empirical confidence interval coverage.

For every simulation experiment, we obtain n_{sim} sets of estimated parameters. To assess the estimation error between an estimated set of parameters $\hat{\lambda}$ and true values λ, we use mean absolute

Table 3. Average distances between estimated and true parameters under all four covariance functions from the four simulation experiments

	Covariance	MAE				RMSE			
c_S	Model	$\hat{\sigma}^2$	$\hat{c}_S$	$\hat{c}_T$	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{c}_S$	$\hat{c}_T$	$\hat{\tau}^2$
20 km	$\mathbb{T}$	**0.100**	**3.778**	**0.014**	**0.009**	**0.128**	**4.737**	**0.018**	**0.012**
	$\mathbb{C}_1$	0.103	7.643	0.015	**0.009**	0.133	8.137	0.019	**0.012**
50 km	$\mathbb{T}$	**0.131**	**10.350**	**0.016**	**0.009**	**0.165**	**12.905**	**0.020**	**0.011**
	$\mathbb{C}_1$	0.138	23.360	**0.016**	**0.009**	0.174	24.272	0.021	**0.011**
100 km	$\mathbb{T}$	**0.162**	**22.856**	**0.016**	**0.008**	**0.20**	**29.535**	**0.021**	**0.010**
	$\mathbb{C}_1$	0.172	53.747	0.017	**0.008**	0.218	55.321	0.022	0.011
200 km	$\mathbb{T}$	**0.197**	**53.537**	**0.018**	**0.007**	**0.251**	**68.228**	**0.023**	**0.009**
	$\mathbb{C}_1$	0.213	121.569	0.020	0.008	0.269	124.212	0.025	0.010

Note. Bolded numbers indicate the smallest distance (best performance). MAE and RMSE refer to the absolute and root mean squared error between the estimated and true parameter values

error $\text{MAE}(\lambda, \hat{\lambda}) = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} |\hat{\lambda}_i - \lambda|$ and root mean squared error $\text{RMSE}(\lambda, \hat{\lambda}) = \{\frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} (\hat{\lambda}_i - \lambda)^2\}^{1/2}$. We present the simulation errors for $\mathbb{T}$ and $\mathbb{C}_1$ in Table 3.

In all simulation settings (i.e. for all values of c_S), MAE and RMSE for all parameters were lower or equal under $\mathbb{T}$, compared to $\mathbb{C}_1$. Unsurprisingly, the largest discrepancies between $\mathbb{T}$ and $\mathbb{C}_1$ are for c_S, where the estimation errors for $\mathbb{C}_1$ are nearly twice those for $\mathbb{T}$. The difference in estimation errors for σ^2 is small compared to those for c_S, while the estimation errors for τ^2 and c_T are very small or negligible.

As a simple visualisation of our estimates under the correctly specified model, we plot histograms of estimated parameters under $\mathbb{T}$ against the true values to confirm that the correct parameters can be identified (see Figure 5). The true value used to generate the data is well centred in the span of estimated parameters, suggesting that we are effectively able to recover parameters.

For the true model $\mathbb{T}$, we use the inverse of the Fisher information matrix to obtain standard errors for the maximum likelihood estimates of the natural logs of τ^2, σ^2, c_S, and c_T. We construct 95% confidence intervals for our parameters (on the log scale) using these standard errors and by relying on the asymptotic normality of maximum likelihood estimators. For each simulation i, we obtain lower and upper 95% confidence bounds $l_i^{(\lambda)}$ and $u_i^{(\lambda)}$ for each parameter λ by exponentiating the confidence intervals on the log-scale. We calculate the average 95% confidence interval width $\text{CIW}(l^{(\lambda)}, u^{(\lambda)}) = \frac{1}{n_{sim}} \sum_{i=1}^{n_{\text{sim}}} (u_i^{(\lambda)} - l_i^{(\lambda)})$ and the empirical 95% interval coverage $\text{EIC}(l^{(\lambda)}, u^{(\lambda)}) = \frac{1}{n_{sim}} \sum_{i=1}^{n_{\text{sim}}} \mathbf{1}_{[l_i^{(\lambda)}, u_i^{(\lambda)}]}(\lambda)$, where $\mathbf{1}_A(\cdot)$ is an indicator function over the set A. These quantities for each simulation experiment are given in Table 4.

Although there is slightly worse coverage for σ^2 for more persistent spatial correlations (larger c_S), the coverage rates in Table 4 are close to the expected level given by a 95% confidence interval for all parameters in all experiments. The interval widths for c_T and τ^2 are essentially the same in all experiments (i.e. different values of c_S). However, as c_S increases, so does the interval width for σ^2, even though σ^2 does not change. This result may suggest similar identifiability challenges common for scale and range parameters for spatial covariance functions (H. Zhang, 2004). However, overall, these simulation results suggest that, under the configurations considered, model parameters are well identified. Nevertheless, there may still be identifiability challenges that warrant future studies.

5 Data illustration

In this data illustration, we consider traffic accident data from the (approximately) 29-mile I-215 beltway around Salt Lake City, Utah, USA from 2015 to 2020. Crashes are indexed by time and location (mile post, starting at 0 in the west and terminating near 29 in the north). The mile post variable maps the locations to a one-dimensional linear space, making network models

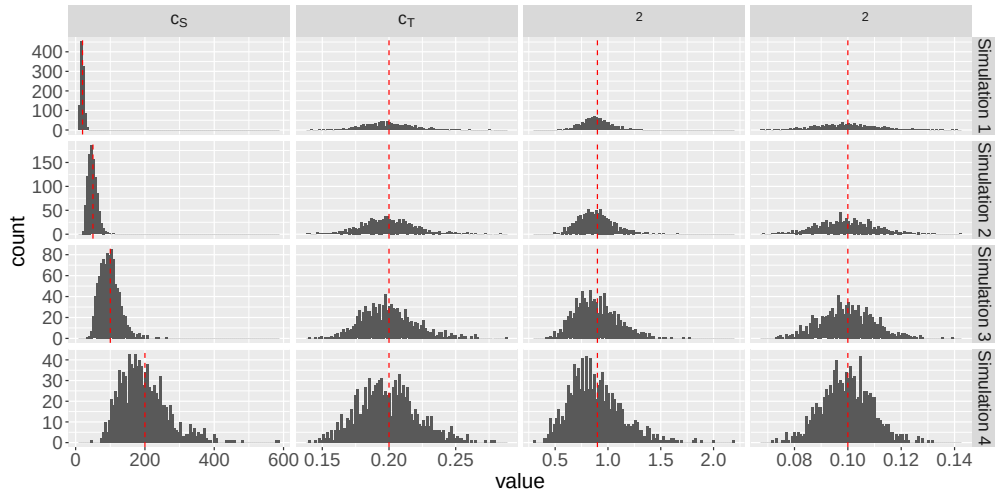

Figure 5. Histogram of model parameters for all simulation experiments. Vertical red dashed lines represent the simulation true values.

Table 4. 95% empirical interval coverage (EIC) and average 95% confidence interval width (CIW) under the correctly specified model

c_S	Criterion	σ^2	c_S	c_T	τ^2
20 km	EIC	0.941	0.958	0.945	0.944
	CIW	0.509	19.5	0.070	0.045
50 km	EIC	0.929	0.944	0.936	0.951
	CIW	0.649	52.7	0.075	0.042
100 km	EIC	0.931	0.937	0.944	0.945
	CIW	0.802	120.6	0.080	0.040
200 km	EIC	0.915	0.94	0.936	0.947
	CIW	0.976	279.9	0.084	0.037

appropriate. In total, we observe 5,027 traffic accidents over these six years. Although we are not licensed to share these data publicly, the data can be requested at https://data-uplan.opendata.arcgis.com/. We have, however, shared relevant code and representative data at https://github.com/philawhite/network-time. To enable simple use of these models and more computationally efficient model fitting, we analyse the data binning over road lengths of 0.5 miles and time windows of approximately one month (30.44 days), giving us counts y_i over $n = 4{,}176$ space-time bins. On average, there are about 1.2 crashes per space-time bin, measured in units of months multiplied by half a mile. Each count y_i is associated with a spatial location $\boldsymbol{s}_i$ and a time t_i, and every location is associated with a milepoint, denoted as $m(\boldsymbol{s}_i)$. We emphasise, however, that our goal is estimating a continuous space-time intensity surface.

In Figure 6, we plot jittered locations of these accidents, a histogram of their occurrence date, and comparison of the Euclidean distance and network geodesic distance between these crashes. These show spatial and temporal heterogeneity, and significant differences between the network geodesic and Euclidean distances. We also point out the drop in accident counts following March 2020 shutdowns due to the COVID-19 pandemic.

In traffic accident modelling, Poisson process models are common (see Jones et al., 1991; Miaou and Lum, 1993, for early examples); however, traffic accident patterns often show overdispersion relative to Poisson processes (see, e.g. Hauer, 2001). In addition, traffic accident patterns often

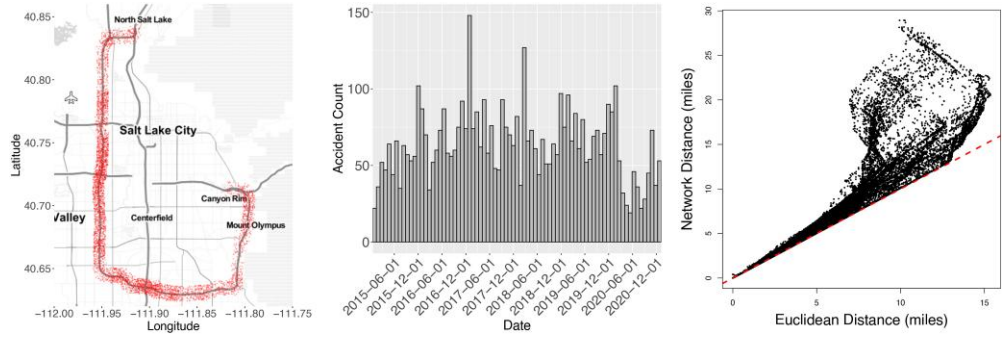

Figure 6. (Left) Jittered accident locations, (Center) a histogram of accident occurrence date, and (Right) a comparison of network and Euclidean distance in miles. The red dashed line has a slope of one and marks equality.

Table 5. Probability mass functions (PMF) used in our analysis of traffic accident counts

PMF	$P(Y_i = y_i), y_i \in \{0, 1, 2, 3, \ldots\}$
Poisson	$\frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}$
Negative Binomial	$\frac{\Gamma(r+y_i)}{\Gamma(r) y_i!} \frac{\mu_i^{y_i} r^r \mu_i^{y_i}}{(r+\mu_i)^{r+y_i}}$
Zero-Inflated Poisson	$\begin{cases} \omega + (1-\omega) e^{-\mu_i} & y_i = 0 \\ (1-\omega) \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} & y_i > 0 \end{cases}$

have higher rates of zeros than Poisson process models support (Shankar et al., 1997). Thus, in addition to Poisson models, we also consider Negative Binomial models and Zero-Inflated Poisson models.[1] We outline these models in Table 5.

For all models, we have a location parameter μ_i which is specified as $\log(\mu_i) = \beta_0 + w_i$, where β_0 is an intercept for the log mean of counts and $(w_1, \ldots, w_n)^\top \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma})$ are zero-mean space-time random effects with covariance $\boldsymbol{\Sigma}$ specified using either $\mathbb{T}$ or $\mathbb{C}_1$ from Section 4. For simplicity in expressing the model, we write $\mathbb{T}$ and $\mathbb{C}_1$ as functions of both distance and time, as well as parameters σ^2, c_S, and c_T. $\mathbb{T}$ uses the network geodesic, while $\mathbb{C}_1$ uses Euclidean distances calculated using Eastings and Northings. Thus, $\text{cov}(w_i, w_j) = \mathbb{T}(|m(\boldsymbol{s}_i) - m(\boldsymbol{s}_j)|, |t_i - t_j|; \sigma^2, c_S, c_T)$ or $\text{cov}(w_i, w_j) = \mathbb{C}_1(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|, |t_i - t_j|; \sigma^2, c_S, c_T)$, depending on the specification. We clarify that we do not include a nugget effect τ^2 (as in Section 4.1) because this parameter is equivalent to a log-normal overdispersion parameter within the generalised model framework. In this case, we prefer to address potential overdispersion directly through the probability mass function (PMF) specification. In total, using three PMFs and two covariance functions, we consider six different models.

We fit all models in a Bayesian framework. We use the following prior distributions for our model parameters (when applicable):

$$\begin{aligned} \beta_0 &\sim \text{Normal}(0, 100), \\ (w_1, \ldots, w_n)^\top &\sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}), \\ \log((\sigma^2, c_S, c_T)^\top) &\sim \text{Normal}(\mathbf{0}, 10\,\mathbf{I}), \\ r &\sim \text{Uniform}(0, 50), \\ \omega &\sim \text{Uniform}(0, 1). \end{aligned} \tag{5.1}$$

These prior distributions are weakly informative, but they provide flexibility given the number of crashes, the spatial range of the dataset (in miles), and the time differences in weeks. For the over-dispersion parameter (r, normally defined as a number of failures) in the Negative Binomial models, we assume that $r \sim \text{Unif}(0, 50)$ because the Negative Binomial resembles the Poisson distribution

[1] We also considered Zero-Inflated Negative Binomial models; however, these models performed poorly.

Table 6. Model comparison results. 'Pois' is the Poisson distribution, 'NB' is the Negative Binomial distribution, and 'ZIP' is the Zero-Inflated Poisson distribution

	Covariance	PMF	WAIC	lppd	p_{WAIC}	RPS	DSS	MAE
1	$\mathbb{T}$	Pois	10730.4	−4675.3	689.9	0.750	1.947	1.162
2	$\mathbb{T}$	NB	10680.4	−4639.8	700.4	0.724	**1.832**	1.108
3	$\mathbb{T}$	ZIP	**10521.9**	**−4609.8**	651.2	**0.699**	1.910	**1.050**
4	$\mathbb{C}_1$	Pois	10633.7	−4626.1	690.8	0.754	2.007	1.170
5	$\mathbb{C}_1$	NB	10799.0	−4694.9	704.5	0.746	1.892	1.154
6	$\mathbb{C}_1$	ZIP	10809.9	−4690.0	715.0	0.783	2.322	1.221

Note. The best performance for each quantity is indicated using bold text, excluding the model complexity.

Table 7. Posterior mean, standard deviation, and 95% central credible intervals (2.5% and 97.5%) for ω, $(1-\omega)e^{\beta_0+\sigma^2/2}$, σ^2, c_S, and c_T

	Mean	Std. Dev.	2.5%	97.5%
ω	0.0020	0.0022	0.0001	0.0075
$(1-\omega)e^{\beta_0+\sigma^2/2}$	1.1724	0.0200	1.1334	1.2129
σ^2	0.6276	0.0538	0.5715	0.7646
c_S	1.1749	0.1257	0.9402	1.3273
c_T	5.8003	0.4444	5.0740	6.3762

for large r. For the Zero-Inflated Poisson distribution, we assume that the zero-inflation probability ω is Uniform(0, 1) and is constant over space as used by Pew et al. (2020) on a similar dataset.

We fit the models using NIMBLE (de Valpine et al., 2017). We sample the log covariance parameters, the intercept, the overdispersion parameter, and zero-inflation parameter using a Normal random walk. We sample the space-time random effects using elliptical slice sampling (Murray et al., 2010). We run this MCMC for 50 000 iterations, discard a burn-in of 20 000 iterations, and, for memory reasons, thin the remaining samples to 5 000 samples.

To compare models, we use the Watanabe-Akaike information criteria (WAIC) (Watanabe 2010). The WAIC approximates cross-validation and is calculated using the computed log pointwise predictive density

$$\text{lppd} = \sum_{i=1}^{n} \log\left\{\frac{1}{M}\sum_{m=1}^{M} p(y_i \mid \theta^m)\right\},$$

as well as a complexity penalty $p_{WAIC} = \sum_{i=1}^{n} \text{var}[\log\{p(y_i \mid \theta)\}]$ (see Gelman et al., 2014, for more discussion). With these components, WAIC is defined as $-2\text{lppd} + 2p_{WAIC}$, and a smaller WAIC represents a better model. In addition, we also consider the out-of-sample predictive performance by carrying out 10-fold cross-validation, where each data point is held out at random exactly one time. We compare competing models using two strictly proper scoring rules for count data (Czado et al., 2009): the ranked probability score (RPS) (Czado et al., 2009; Gneiting and Raftery, 2007), a discrete version of CRPS, and the Dawid-Sebastiani score (DSS) (Dawid and Sebastiani, 1999),

$$DSS(F, y) = \left(\frac{y-\mu_F}{\sigma_F}\right)^2 + 2\log\sigma_F.$$

We estimate the RPS using the empirical CDF approach discussed in Krüger et al. (2021), while the DSS only relies on the first two moments (μ_F and σ_F^2) of the predictive distribution

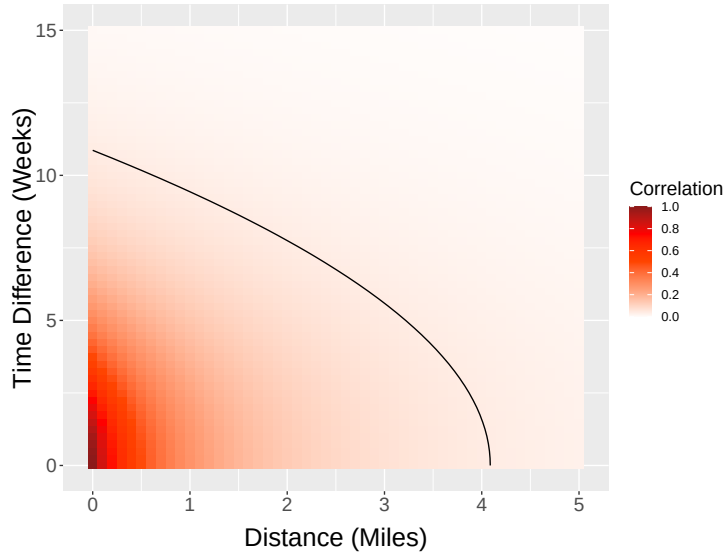

Figure 7. Posterior mean for the correlation function over network distance and difference in time, where the colour scale varies logarithmically.

F. Lastly, we include the mean absolute error (MAE). The results of these comparisons are listed in Table 6.

The results in Table 6 show that the models using the network geodesic are better in terms of out-of-sample prediction; however, they are not always better in terms of WAIC. The PMF was also important in model performance, but the best PMF differed depending on the distance metric and model comparison metric used. Although the negative binomial model using $\mathbb{T}$ was best in terms of DSS, the best model in terms of fit (lppd), WAIC, RPS, and MAE is the zero-inflated Poisson model with random effects using $\mathbb{T}$. We highlight that this covariance function uses the network geodesic to specify space-time random effects and outperforms all models that use the Euclidean distance. We interpret the results based on this model.

For the zero-inflated Poisson model using $\mathbb{T}$, we present posterior summaries for ω, $(1-\omega)e^{\beta_0+\sigma^2/2}$, σ^2, c_S, and c_T in Table 7. We use $(1-\omega)e^{\beta_0+\sigma^2/2}$ because this represents the mean of the model, accounting for the mean-zero Gaussian random effects on the log-scale and the zero-inflation.

Considering the improvement of the zero-inflated Poisson model relative to the Poisson model, the low posterior mean for ω is surprising and suggests that zero-inflation plays a very small role in the model. Nonetheless, the improved prediction under this model was substantial. The 95% credible interval for $(1-\omega)e^{\beta_0+\sigma^2/2}$ suggests that the baseline or average number of crashes is between (1.0330, 1.1903), in units of months multiplied by half a mile. To visualise the effect of c_S and c_T, we plot the posterior mean of the correlation function for the random effects as a function of network distance and time difference (in weeks) in Figure 7. In this plot, we include a contour line marking the effective range (the distance/time difference where the correlation reaches 0.05). Although the correlation decays quickly as a function of distance over the network, correlation persists for many weeks over short distances. On the other hand, for short time differences (e.g. around five weeks) the posterior mean correlation remains above 0.05 for just over 4 miles. On the whole, the posterior mean correlation function suggests rapid spatial changes but more persistent spatial patterns.

Figure 8 displays the posterior mean and standard deviation of the intensity surface, which represents the expected number of crashes per mile and per week. This plot allows us to explore the spatio-temporal patterns in the data, which reveal high variability in the expected number of accidents across space and time. The plot also demonstrates a clear relationship between the mean and standard deviation that we would expect under a zero-inflated Poisson model.

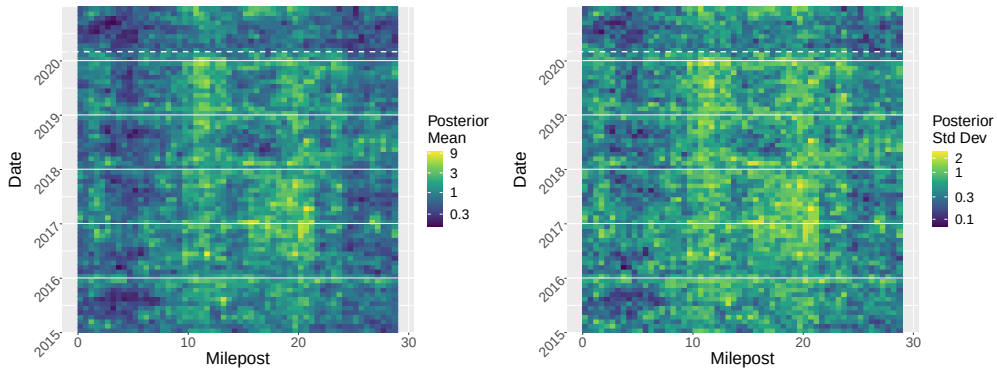

Figure 8. (Left) Posterior mean and (Right) standard deviation for the intensity function or average number of crashes per milexweek, where the colour scale varies logarithmically. Solid horizontal lines indicate new years, while the thicker dashed line represents the timing of shutdowns due to COVID-19.

One of the most interesting external factors during the time span of these data was the COVID-19-related shutdowns in March 2020 in the United States. To visualise the impact of these shutdowns, we have added a dashed line in the figure to indicate when the lockdowns went into effect. The before-and-after pattern is clearly visible along the entire belt route, with the accident intensity dropping rapidly beyond the dashed line, coinciding with decreases in daily commuting. The largest difference in the space-time random effect is near mile post 12, which is the junction between I-215 and I-15, the primary highway in Utah. Overall, these results show strong spatio-temporal patterns in the data.

6 Conclusions

We have provided flexible classes of space-time covariance functions that can be used over linear or nonlinear networks. Our exposition strategy has been devoted to simple illustrations that avoid mathematical details (provided in the Online Supplementary Material). This allows the practitioner to understand how to use the new models. We have also focused on how to practically calculate distances over networks. Our simulation study highlighted that if the network geodesic is used to generate the data, then model fit, predictive performance, and parameter estimation are better using the network geodesic distance than Euclidean distance. In addition, we found that we can effectively recover the parameters of the true covariance function; however, we found increased uncertainty in the covariance scale parameter as the spatial range parameter increased. From our analysis of traffic accident patterns on a simple road network, we found that our best model used the network geodesic distance and outperformed all models that failed to account for network structure. Using this model, we explored posterior summaries, the correlation structure of space-time random effects, and the model's posterior mean and standard deviation over space and time. This work lays the foundation for many challenges from both theoretical and applied standpoints, among which are:

1. The problem of multivariate covariance functions over networks has, to our knowledge, not been addressed so far by earlier literature. Modern datasets are often characterised by several georeferenced variables that are observed over time. For them, addressing the cross-correlation is of fundamental importance for modelling, estimation, and prediction (Genton and Kleiber, 2015).
2. Datasets over linear networks often exhibit nonstationarities over space and time, so that using a covariance function that solely depends on distances might result in unrealistic assumptions. The literature on this subject is elusive so far, and it is unclear how to adapt existing approaches to nonstationarity (Paciorek and Schervish, 2006; Porcu et al., 2010) that have been proposed in Euclidean spaces. In turn, nonstationary models would be the key to a fertile literature on reducibility approaches that allow to interpret a nonstationary random field as a

stationary one if commuted into some suitable manifold (Porcu, Senoussi, et al., 2020). An attempt to provide nonstationary models on networks was done by Song and Zimmerman (2021), who have recently challenged the problem through two alternative approaches. The first, called the *elastic* approach, is based on a parameter that allows for either deformation or condensation of the local coordinates (see also Zimmerman and Núñez-Antón, 1997). The second is based on spatially varying moving average models, which represent the nonstationary counterpart of the approach provided by Ver Hoef and Peterson (2010).

3. Graphs with Euclidean edges might be a good alternative to parameterised curves representing special domains in spatial statistics. Abdalla et al. (2018) consider the problem of geostatistical modelling over a coastline. Their strategy is to parameterise the coastline coordinates through parametric curves that can be piecewise approximated through lines for segments of the real line. The approach includes the exponential covariance function on the real line, where the *coordinates* of the real line depend on the parametric curve indexing any point located over the coastline. An alternative approach might be to *split* the coastline into a collection of sub-coastlines with endpoints given by the sampling point. This would allow us to retrieve the machinery provided by this paper for the case of graphs with Euclidean edges. Another alternative is achieved by noting that, given the topology considered by Abdalla et al. (2018), nonstationarities of the second order are plausible. Hence, one might adopt reduction approaches as much as Porcu, Senoussi, et al. (2020): for a given nonstationary covariance K on the coastline, find a pair (Φ, ρ), with ρ a correlation function and Φ a bijection, such that $K(t, t') = \rho(\Phi(t) - \Phi(t'))$. Similarly, a reduction approach might be taken on the basis of dynamical systems driven by partial differential equations as in Senoussi and Porcu (2022).
4. The proposed methodology might be of substantial help to research activity in public health pertaining to data collected from wearable devices as subjects move around (walk, run or carry out their daily physical activities). The use of wearable devices, such as wrist-worn sensors that monitor gross motor activity (actigraphy), have become ubiquitous. Through actigraphs, every individual is recorded continuously over time. Hence, one has a huge amount of data at hand. Alaimo Di Loro et al. (2021) adopt a hierarchical Bayesian approach to this problem. Given the frameworks adopted here and the related solutions, it would be tempting to consider the data as functions (given that they are almost continuous over time) for every spatial location. Hence, one would need to extend our approach to functional covariances for Gaussian processes that are defined over some function spaces.
5. Understanding the smoothness of processes defined over graphs with Euclidean edges (possibly evolving over time) is still elusive. For Gaussian fields, this might be challenged from the perspective of smoothness of covariance functions. In this direction, one promising avenue might be to adapt the approaches based on Bayesian wombling (Banerjee and Gelfand, 2006; Gelfand and Banerjee, 2015). The effort by Anderes et al. (2020) to build Brownian bridges over these topologies suggest that such an extension is not incremental, but requires substantial work.
6. Some comments are in order regarding computational scalability. When working on a linear network, the examples provided in our paper show that substantial computational gains can be achieved through the covariance functions with dynamical support, which allow for sparse spatial matrices for every fixed temporal lag. Such a feature is not present over more general graphs with Euclidean edges, where our examples, as well as those provided by Tang and Zimmerman (2020), refer to covariance functions that are globally supported. At the same time, the approach based on covariance functions allows likelihood approximation approaches that have been introduced in the last 20 years in spatial statistics, and the reader is referred to Sun et al. (2012) for a comprehensive overview about likelihood approximations and covariance tapering. An alternative is represented by the approach discussed by Bolin et al. (2022) over graphs (no time is considered there). Extending such an approach to space-time is a major open challenge.
7. Covariance functions are highly in demand in several branches of statistics and machine learning, where they are called kernels. Reproducing kernel Hilbert space (RKHS) methods have been extremely popular. Our work has a direct impact on RKHS methods on graphs. To mention, Muandet et al. (2017) invoke several connections between kernel mean

embeddings and kernels on graphs; applications in Artificial Intelligence for reducing dimensionality on graphs rely on kernels (Riesen and Bunke, 2009b); kernels on graphs allow for generative models for return probabilities on graphs (Z. Zhang et al., 2018); kernels on graphs are crucial to graph classification through manifold embeddings (Riesen and Bunke, 2009a).

Acknowledgments

We acknowledge Jun Tang for sharing code and data, and we thank the review team for comments that improved the paper.

Conflict of interest: None declared.

Funding

This work was supported by the Khalifa University of Science and Technology Award No. FSU-2021-016 (E. Porcu), NSF-DMS CDS&E grant 2053188 (P. White), and the King Abdullah University of Science and Technology (M. Genton).

Data availability

The traffic accident data can be requested at https://data-uplan.opendata.arcgis.com/.

Supplementary material

Supplementary material are available at *Journal of the Royal Statistical Society: Series B* online.

References

Abdalla N., Banerjee S., Ramachandran G., Stenzel M., & Stewart P. A. (2018). Coastline kriging: A Bayesian approach. *Annals of Work Exposures and Health*, *62*(7), 818–827. https://doi.org/10.1093/annweh/wxy058

Alaimo Di Loro P., Mingione M., Lipsitt J., Batteate C. M., Jerrett M., & Banerjee S. (2021). 'Bayesian hierarchical modeling and analysis for physical activity trajectories using actigraph data', arXiv, arXiv-2101, preprint: not peer reviewed.

Alegría A., Porcu E., Furrer R., & Mateu J. (2019). Covariance functions for multivariate Gaussian fields evolving temporally over planet earth. *Stochastic Environmental Research and Risk Assessment*, *33*, 1593–1608. https://doi.org/10.1007/s00477-019-01707-w

Alsheikh M. A., Lin S., Niyato D., & Tan H.-P. (2014). Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys & Tutorials*, *16*(4), 1996–2018. https://doi.org/10.1109/COMST.2014.2320099

Anderes E., Møller J., & Rasmussen J. G. (2020). Isotropic covariance functions on graphs and their edges. *The Annals of Statistics*, *48*(4), 2478–2503. https://doi.org/10.1214/19-AOS1896

Apanasovich T. V., & Genton M. G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika*, *97*(1), 15–30. https://doi.org/10.1093/biomet/asp078

Baddeley A., Nair G., Rakshit S., & McSwiggan G. (2017). Stationary point processes are uncommon on linear networks. *Stat*, *6*(1), 68–78. https://doi.org/10.1002/sta4.v6.1

Baddeley A., Nair G., Rakshit S., McSwiggan G., & Davies T. M. (2021). Analysing point patterns on networks—a review. *Spatial Statistics*, *42*, 100435. https://www.sciencedirect.com/science/article/pii/S2211675320300294. Towards Spatial Data Science. https://doi.org/10.1016/j.spasta.2020.100435

Banerjee S., & Gelfand A. E. (2006). Bayesian wombling: Curvilinear gradient assessment under spatial process models. *Journal of the American Statistical Association*, *101*(476), 1487–1501. https://doi.org/10.1198/016214506000000041

Bevilacqua M., Caamaño-Carrillo C., & Porcu E. (2022). Unifying compactly supported and Matérn covariance functions in spatial statistics. *Journal of Multivariate Analysis*, *189*, 104949. https://doi.org/10.1016/j.jmva.2022.104949

Bevilacqua M., Faouzi T., Furrer R., & Porcu E. (2019). Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. *The Annals of Statistics*, *47*(2), 828–856. https://doi.org/10.1214/17-AOS1652

Bolin D., Simas A. B., & Wallin J. (2022). 'Gaussian Whittle-Matérn fields on metric graphs', arXiv, arXiv:2205.06163, preprint: not peer reviewed.

Borovitskiy V., Karimi M. R., Somnath V. R., & Krause A. (2022). 'Isotropic gaussian processes on finite spaces of graphs', arXiv, arXiv:2211.01689, preprint: not peer reviewed.
Chilès J., & Delfiner P. (2012). *Geostatistics: Modeling spatial uncertainty*. Wiley.
Cressie N., Frey J., Harch B., & Smith M. (2006). Spatial prediction on a river network. *Journal of Agricultural, Biological, and Environmental Statistics*, *11*(2), 127.150 https://doi.org/10.1198/108571106X110649
Czado C., Gneiting T., & Held L. (2009). Predictive model assessment for count data. *Biometrics*, *65*(4), 1254–1261. https://doi.org/10.1111/j.1541-0420.2009.01191.x
Dawid A. P., & Sebastiani P. (1999). Coherent dispersion criteria for optimal experimental design. *The Annals of Statistics*, *27*(1), 65–81. https://doi.org/10.1214/aos/1018031101
De Cesare L., Myers D., & Posa D. (2001). Product-sum covariance for space-time modeling: An environmental application. *Environmetrics*, *12*, 11–23. https://doi.org/10.1002/(ISSN)1099-095X
Deng N., Zhou W., & Haenggi M. (2014). The Ginibre point process as a model for wireless networks with repulsion. *IEEE Transactions on Wireless Communications*, *14*(1), 107–121. https://doi.org/10.1109/TWC.2014.2332335
de Valpine P., Turek D., Paciorek C. J., Anderson-Bergman C., Lang D. T., & Bodik R. (2017). Programming with models: Writing statistical algorithms for general model structures with nimble. *Journal of Computational and Graphical Statistics*, *26*(2), 403–413. https://doi.org/10.1080/10618600.2016.1172487
Ehm W., Gneiting T., & Richards D. (2004). Convolution roots of radial positive definite functions with compact support. *Transactions of the American Mathematical Society*, *356*(11), 4655–4685. https://doi.org/10.1090/tran/2004-356-11
Emery X., & Porcu E. (2022). *Extending the Gneiting class for modeling spatially isotropic and temporally symmetric vector random fields* (Technical Report). Submitted for publication.
Epstein E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology (1962–1982)*, *8*, 985–987. https://doi.org/10.1175/1520-0450(1969)008¡0985:ASSFPF¿2.0.CO;2
Fonseca T. C. O., & Steel M. F. J. (2011). A general class of nonseparable space-time covariance models. *Environmetrics*, *22*(2), 224–242. https://doi.org/10.1002/env.v22.2
Gardner B., Sullivan P. J., & Lembo Jr A. J. (2003). Predicting stream temperatures: Geostatistical model comparison using alternative distance metrics. *Canadian Journal of Fisheries and Aquatic Sciences*, *60*(3), 344–351. https://doi.org/10.1139/f03-025
Gelfand A. E., & Banerjee S. (2015). Bayesian wombling: Finding rapid change in spatial maps. *Wiley Interdisciplinary Reviews: Computational Statistics*, *7*(5), 307–315. https://doi.org/10.1002/wics.1360
Gelman A., Hwang J., & Vehtari A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. https://doi.org/10.1007/s11222-013-9416-2
Genton M. G., & Kleiber W. (2015). Cross-covariance functions for multivariate geostatistics (with discussion). *Statistical Science*, *30*, 147–163. https://doi.org/10.1214/14-STS487
Georgopoulos L., & Hasler M. (2014). Distributed machine learning in networks by consensus. *Neurocomputing*, *124*, 2–12. https://doi.org/10.1016/j.neucom.2012.12.055
Gneiting T. (2002a). Compactly supported correlation functions. *Journal of Multivariate Analysis*, *83*(2), 493–508. https://doi.org/10.1006/jmva.2001.2056
Gneiting T. (2002b). Stationary covariance functions for space-time data. *Journal of the American Statistical Association*, *97*, 590–600. https://doi.org/10.1198/016214502760047113
Gneiting T., & Raftery A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437
Hamilton W. L., Ying R., & Leskovec J. (2017). 'Representation learning on graphs: Methods and applications', arXiv, arXiv:1709.05584, preprint: not peer reviewed.
Hauer E. (2001). Overdispersion in modelling accidents on road sections and in empirical Bayes estimation. *Accident Analysis & Prevention*, *33*(6), 799–808. https://doi.org/10.1016/S0001-4575(00)00094-4
Isaak D. J., Luce C. H., Chandler G. L., Horan D. L., & Wollrab S. P. (2018). Principal components of thermal regimes in mountain river networks. *Hydrology and Earth System Sciences*, *22*(12), 6225–6240. https://doi.org/10.5194/hess-22-6225-2018
Jones B., Janssen L., & Mannering F. (1991). Analysis of the frequency and duration of freeway accidents in Seattle. *Accident Analysis & Prevention*, *23*(4), 239–255. https://doi.org/10.1016/0001-4575(91)90003-N
Krüger F., Lerch S., Thorarinsdottir T., & Gneiting T. (2021). Predictive inference based on Markov Chain Monte Carlo output. *International Statistical Review*, *89*(2), 274–301. doi:10.1111/insr.12405
Lindgren F., Rue H., & Lindstroem 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*, *73*(4), 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x
Mastrantonio G., Jona Lasinio G., Pollice A., Capotorti G., Teodonio L., Genova G., & Blasi C. (2019). A hierarchical multivariate spatio-temporal model for clustered climate data with annual cycles. *The Annals of Applied Statistics*, *13*(2), 797–823. https://dx.doi.org/10.1214/18-AOAS1212

Matheson J. E., & Winkler R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, *22*(10), 1087–1096. https://doi.org/10.1287/mnsc.22.10.1087
Menegatto V., Olira C., & Porcu E. (2020). Gneiting class, semi-metric spaces and isometric embeddings. *Constructive Mathematical Analysis*, *3*(2), 85–95. https://doi.org/10.33205/cma.712049
Miaou S.-P., & Lum H. (1993). Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis & Prevention*, *25*(6), 689–709. https://doi.org/10.1016/0001-4575(93)90034-T
Montembeault M., Joubert S., Doyon J., Carrier J., Gagnon J.-F., Monchi O., Lungu O., Belleville S., & Brambati S. M. (2012). The impact of aging on gray matter structural covariance networks. *Neuroimage*, *63*(2), 754–759. https://doi.org/10.1016/j.neuroimage.2012.06.052
Moradi M. M., & Mateu J. (2020). First- and second-order characteristics of spatio-temporal point processes on linear networks. *Journal of Computational and Graphical Statistics*, *29*(3), 432–443. https://doi.org/10.1080/10618600.2019.1694524
Muandet K., Fukumizu K., Sriperumbudur B., & Schölkopf B. (2017). Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning*, *10*(1-2), 1–141. https://doi.org/10.1561/2200000060
Murray I., Adams R., & MacKay D. (2010). Elliptical slice sampling. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (pp. 541–548). JMLR Workshop and Conference Proceedings.
Paciorek C. J., & Schervish M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, *17*(5), 483–506. https://doi.org/10.1002/(ISSN)1099-095X
Peron A., Porcu E., & Emery X. (2018). Admissible nested covariance models over spheres cross time. *Stochastic Environmental Research and Risk Assessment*, *32*(11), 3053–3066. https://doi.org/10.1007/s00477-018-1576-3
Perry P. O., & Wolfe P. J. (2013). Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: Series B*, *75*(5), 821–849. https://doi.org/10.1111/rssb.12013
Peterson E. E., Theobald D. M., & ver Hoef J. M. (2007). Geostatistical modelling on stream networks: Developing valid covariance matrices based on hydrologic distance and stream flow. *Freshwater Biology*, *52*(2), 267–279. https://doi.org/10.1111/fwb.2007.52.issue-2
Peterson E. E., Ver Hoef J. M., Isaak D. J., Falke J. A., Fortin M. -J., Jordan C. E., McNyset K., Monestiez P., Ruesch A. S., Sengupta A., Som N., Steel E. A., Theobald D. M., Torgersen C. E., Wenger S. J., & Blasius B. (2013). Modelling dendritic ecological networks in space: An integrated network perspective. *Ecology Letters*, *16*(5), 707–719. https://doi.org/10.1111/ele.2013.16.issue-5
Pew T., Warr R. L., Schultz G. G., & Heaton M. (2020). Justification for considering zero-inflated models in crash frequency analysis. *Transportation Research Interdisciplinary Perspectives*, *8*, 100249. https://doi.org/10.1016/j.trip.2020.100249
Pinder T., Turnbull K., Nemeth C., & Leslie D. (2021). 'Gaussian processes on hypergraphs', arXiv, arXiv:2106.01982, preprint: not peer reviewed.
Porcu E., Alegría A., & Furrer R. (2018). Modeling temporally evolving and spatially globally dependent data. *International Statistical Review*, *86*(2), 344–377. https://doi.org/10.1111/insr.v86.2
Porcu E., Bevilacqua M., & Genton M. G. (2020). Space-time covariance functions with dynamical compact supports. *Statistica Sinica*, *30*, 719–739. https://doi.org/10.5705/ss.202017.0385
Porcu E., Bevilacqua M., Schaback R., & Oates C. J. (2023). 'The Matérn model: A journey through statistics, numerical analysis and machine learning', arXiv, arXiv:2303.02759, preprint: not peer reviewed.
Porcu E., Furrer R., & Nychka D. (2020). 30 years of space–time covariance functions. *Wiley Interdisciplinary Reviews: Computational Statistics*. https://doi.org/10.1002/wics.1512
Porcu E., Gregori P., & Mateu J. (2006). Nonseparable stationary anisotropic space–time covariance functions. *Stochastic Environmental Research and Risk Assessment*, *21*(2), 113–122. https://doi.org/10.1007/s00477-006-0048-3
Porcu E., & Mateu J. (2007). Mixture-based modeling for space-time data. *Environmetrics*, *18*(3), 285–302. https://doi.org/10.1002/(ISSN)1099-095X
Porcu E., Mateu J., & Bevilacqua M. (2007). Covariance functions which are stationary or nonstationary in space and stationary in time. *Statistica Neerlandica*, *61*(3), 358–382. https://doi.org/10.1111/stan.2007.61.issue-3
Porcu E., Mateu J., & Christakos G. (2010). Quasi-arithmetic means of covariance functions with potential applications to space-time data. *Journal of Multivariate Analysis*, *100*(8), 1830–1844. https://doi.org/10.1016/j.jmva.2009.02.013
Porcu E., Senoussi R., Mendoza E., & Bevilacqua M. (2020). Reduction problems and deformation approaches to nonstationary covariance functions over spheres. *Electronic Journal of Statistics*, *14*(1), 890–916. http://dx.doi.org/10.1214/19-EJS1670
Porcu E., & Zastavnyi V. (2011). Characterization theorems for some classes of covariance functions associated to vector valued random fields. *Journal of Multivariate Analysis*, *102*(9), 1293–1301. https://doi.org/10.1016/j.jmva.2011.04.013

Rakshit S., Nair G., & Baddeley A. (2017). Second-order analysis of point patterns on a network using any distance metric. *Spatial Statistics*, *22*, 129–154. https://doi.org/10.1016/j.spasta.2017.10.002

Riesen K., & Bunke H. (2009a). Graph classification based on vector space embedding. *International Journal of Pattern Recognition and Artificial Intelligence*, *23*(06), 1053–1081. https://doi.org/10.1142/S021800140900748X

Riesen K., & Bunke H. (2009b). Reducing the dimensionality of dissimilarity space embedding graph kernels. *Engineering Applications of Artificial Intelligence*, *22*, 48–56. https://doi.org/10.1016/j.engappai.2008.04.006

Schlather M. (2010). Some covariance models based on normal scale mixtures. *Bernoulli*, *16*(3), 780–797. https://doi.org/10.3150/09-BEJ226

Senoussi R., & Porcu E. (2022). Nonstationary space–time covariance functions induced by dynamical systems. *Scandinavian Journal of Statistics*, *49*(1), 211–235. https://doi.org/10.1111/sjos.v49.1

Shankar V., Milton J., & Mannering F. (1997). Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. *Accident Analysis & Prevention*, *29*(6), 829–837. https://doi.org/10.1016/S0001-4575(97)00052-3

Shirota S., & Gelfand A. E. (2017). Space and circular time log Gaussian Cox processes with application to crime event data. *The Annals of Applied Statistics*, *11*(2), 481–503. https://doi.org/10.1214/16-AOAS960

Song R., & Zimmerman D. L. (2021). Modeling spatial correlation that grows on trees, with a stream network application. *Spatial Statistics*, *45*, 100536. https://doi.org/10.1016/j.spasta.2021.100536

Stein M. L. (1999). *Statistical interpolation of spatial data: Some theory for kriging*. Springer.

Stein M. L. (2005). Space-time covariance functions. *Journal of the American Statistical Association*, *100*, 310–321. https://doi.org/10.1198/016214504000000854

Sun Y., Li B., & Genton M. G. (2012). Geostatistics for large datasets. In *Advances and challenges in space-time modelling of natural events* (pp. 55–77). Springer.

Tang J., & Zimmerman D. (2020). 'Space-time covariance models on networks with an application on streams', arXiv, arXiv:2009.14745, preprint: not peer reviewed.

Ver Hoef J. M., Peterson E., & Theobald D. (2006). Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics*, *13*(4), 449–464. https://doi.org/10.1007/s10651-006-0022-8

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*(489), 6–18. https://doi.org/10.1198/jasa.2009.ap08248

Watanabe S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

White P. A., & Porcu E. (2019). Nonseparable covariance models on circles cross time: A study of Mexico City ozone. *Environmetrics*, *30*(4), e2558. https://dx.doi.org/10.1002/env.2558

Xiao S., Yan J., Yang X., Zha H., & Chu S. (2017). Modeling the intensity function of point process via recurrent neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31).

Zastavnyi V. P. (2000). On positive definiteness of some functions. *Journal of Multivariate Analysis*, *73*(1), 55–81. https://doi.org/10.1006/jmva.1999.1864

Zhang H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, *99*(465), 250–261. https://doi.org/10.1198/016214504000000241

Zhang Z., Wang M., Xiang Y., Huang Y., & Nehorai A. (2018). Retgk: Graph kernels based on return probabilities of random walks. *Advances in Neural Information Processing Systems*, *31*.

Zimmerman D. L., & Núñez-Antón V. (1997). Structured antedependence models for longitudinal data. In *Modelling longitudinal and spatially correlated data* (pp. 63–76). Springer.