Space-time covariance models on networks

Jun Tang[*] **and Dale Zimmerman**[†]

*Department of Statistics and Actuarial Science,
University of Iowa
e-mail:* jun-tang-1@uiowa.edu*;* dale-zimmerman@uiowa.edu

Abstract: The second-order, small-scale dependence structure of a stochastic process defined in the space-time domain is key to prediction (or kriging). While great efforts have been dedicated to developing models for cases in which the spatial domain is either a finite-dimensional Euclidean space or a sphere, counterpart developments on a generalized linear network are practically non-existent. To fill this gap, we develop a broad range of parametric, non-separable space-time covariance models on generalized linear networks. For the important subgroup of Euclidean trees, we develop models by the space embedding technique, in concert with the generalized Gneiting class of models and 1-symmetric characteristic functions, and by the convex cone and scale mixture approaches. We give examples from each class of models and investigate the geometric features of these covariance functions near the origin and at infinity. We also reveal connections between different classes of space-time covariance models on Euclidean trees. We conclude the paper by investigating the performance of maximum likelihood estimators of certain proposed models in a simulation study.

Keywords and phrases: Convex cone, generalized linear network, Euclidean tree, space embedding, scale mixture.

Received April 2023.

1. Introduction

1.1. Background

Despite its wide variety of applications in different scientific disciplines, including environmental (see for example, [34, 9, 33, 25]), neurological [19, 3], ecological [2] and social sciences [2, 4], the study of a random process observed on a network is a relatively new area in spatial statistics. Observations that are closer together in space tend to be more alike than observations far apart [32]. Thus, the small-scale (covariance) structure of a geostatistical process is usually assumed to be a function of distance between spatial locations. On a network, it is possible

---

[*]Jun Tang is Ph.D. graduate from the Department of Statistics and Actuarial Science, University of Iowa, who now works at Novartis. Please address all correspondence to this author.

[†]Dale L. Zimmerman is Professor, Department of Statistics and Actuarial Science and Department of Biostatistics, University of Iowa.

that two sampling sites are close together in the sense of Euclidean distance, but are far apart within the network. Under such a circumstance, it is more reasonable to use the alternative metric when modeling the dependence structure. However, merely replacing the Euclidean distance in a standard geostatistical model with the shortest path within the network may lead to an invalid (not positive definite) covariance function on the network, and thus result in negative prediction variances [34].

In a finite-dimensional Euclidean space, the well-known Bochner's theorem [6] fully characterizes the class of stationary continuous covariance functions as Fourier transforms of finite, nonnegative measures. Though this powerful theorem provides a sufficient and necessary condition for positive definiteness, closed-form Fourier inversions do not always exist. Schoenberg's result [29], on the other hand, is Fourier transform free. It reveals the one-to-one relationship between isotropic covariance functions and completely monotone functions in a Hilbert space. Quite a few non-separable parametric spatio-temporal covariance functions have been developed based on Bochner's and Schoenberg's theorems; see, for example, [10, 16]. [35]'s kernel convolution-based approach can also be generalized into the space-time domain [28].

Covariance functions receive special attention due to the fact that a Gaussian random process is completely determined by its first- and second-order moments [26]. Although a broad range of classes of space-time covariance models are available in Euclidean space [13] and a thorough review has recently been given by [26], corresponding results for pure spatial linear networks are few and far between – a recent exception being [1] – and space-time results on networks are non-existent apart from the recent work of [27]. To fill this gap, we provide an abundance of non-separable, parametric, and Fourier-free space-time covariance models on generalized linear networks. These models are based on the techniques of space embedding, convex cone property of covariance functions, and the scale mixture approach. Our proposed models have appealing features, e.g. some incorporate directionality into modeling (on Euclidean trees), and some have computational efficiencies that may be exploited for large-scale data. We also identify a connection between two distinct types of models, which is of theoretical interest, and compare and contrast our models with those of [27]. Mathematical and technical details are placed into Supplement A.

1.2. Parametric space-time covariance models

Let $\{Z(\boldsymbol{s};t) : (\boldsymbol{s},t) \in \mathcal{D} \times \mathbb{R}\}$ denote a univariate, real-valued, continuously-indexed stochastic process on the product space of a spatial domain $\mathcal{D}$ and a temporal domain $\mathbb{R}$. In the literature, the spatial domain is usually taken to be either a finite-dimensional Euclidean space ($\mathcal{D} = \mathbb{R}^n$) or a unit sphere ($\mathcal{D} = \mathbb{S}^n$) [26]. In contrast, we consider a random process on a generalized linear network ($\mathcal{D} = \mathcal{G}$) whose definition will be given in the subsequent section. Assume that the first two moments of the random process exist and that the mean structure $\mu(\boldsymbol{s};t)$, which measures the global scale space-time variability, can be fixed as

a constant, i.e. $\mu(\boldsymbol{s};t) \equiv \mu$, or modeled as a linear combination of covariates of interest, i.e. $\mu(\boldsymbol{s};t) = \boldsymbol{x}(\boldsymbol{s};t)'\boldsymbol{\beta}$. The second-order, small-scale dependence structure, commonly described by a parametric covariance function, is key to space-time prediction (or kriging) and regression-type parameter estimation and is the main focus of this paper. We do not distinguish between Gaussian and non-Gaussian processes unless necessary since the covariance function plays an important role in either situation.

To be consistent with the recent literature, we let C denote the covariance function where $C(\boldsymbol{s}_1,\boldsymbol{s}_2;t_1,t_2) := \mathrm{Cov}(Z(\boldsymbol{s}_1;t_1), Z(\boldsymbol{s}_2;t_2))$, $(\boldsymbol{s}_i, t_i) \in \mathcal{D} \times \mathbb{R}$, $i = 1, 2$. From the definition, C is symmetric, i.e., $C(\boldsymbol{s}_1,\boldsymbol{s}_2;t_1,t_2) = C(\boldsymbol{s}_2,\boldsymbol{s}_1;t_2,t_1)$. Moreover, a covariance function must be positive definite, meaning that

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j C(\boldsymbol{s}_i, \boldsymbol{s}_j; t_i, t_j) \geq 0 \qquad (1.1)$$

for any finite collections of $\{a_i\}_{i=1}^{N} \subset \mathbb{R}$ and $\{(\boldsymbol{s}_i, t_i)\}_{i=1}^{N} \subset \mathcal{D} \times \mathbb{R}$. Functions which fail to satisfy this condition are likely to lead to negative prediction variances and undefined probability densities. Whenever a function C satisfies the symmetry and positive definiteness conditions, we call it a valid covariance function.

In geostatistics, a common assumption made by practitioners is second-order stationarity, which requires that the overall mean is constant and that the covariance function depends on the spatial locations only through their relative positions. Moreover, in Euclidean space, a covariance function is called isotropic if it is a function of the Euclidean norm of the difference between locations. Unlike its counterpart in Euclidean space, the definition of stationarity is less clear on networks [1]. Nevertheless, a space-time covariance function is said to be isotropic within components if $C(\boldsymbol{s}_1,\boldsymbol{s}_2;t_1,t_2) = f(d(\boldsymbol{s}_1,\boldsymbol{s}_2); |t_1 - t_2|)$ for some function $f : [0,\infty) \times [0,\infty) \to \mathbb{R}$, where $d : \mathcal{D} \times \mathcal{D} \to [0,\infty)$ is a distance metric [1] which satisfies (1) $d(\boldsymbol{s}_1,\boldsymbol{s}_2) = d(\boldsymbol{s}_2,\boldsymbol{s}_1)$, for any $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{D}$; (2) $d(\boldsymbol{s}_1,\boldsymbol{s}_2) = 0$ if and only if $\boldsymbol{s}_1 = \boldsymbol{s}_2$, and $|t_1 - t_2|$ is the absolute difference between times. We call such an f a radial profile function. We work with either the covariance function or the radial profile function, denoting both by C, when the context causes no confusion. By assuming isotropy, the model guarantees that the covariance function is fully symmetric [16] since $C(\boldsymbol{s}_1,\boldsymbol{s}_2;t_1,t_2) = C(\boldsymbol{s}_1,\boldsymbol{s}_2;t_2,t_1) = C(\boldsymbol{s}_2,\boldsymbol{s}_1;t_1,t_2) = C(\boldsymbol{s}_2,\boldsymbol{s}_1;t_2,t_1) = f(d(\boldsymbol{s}_1,\boldsymbol{s}_2); |t_1 - t_2|)$, for any $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{D}$ and $t_1, t_2 \in \mathbb{R}$.

When it comes to spatio-temporal covariance models, assuming separability is a convenient starting point [28]. Specifically, a space-time model is separable if it can be written as a product or a sum of pure spatial and pure temporal models, i.e., $C(\boldsymbol{s}_1,\boldsymbol{s}_2;t_1,t_2) = C_S(\boldsymbol{s}_1,\boldsymbol{s}_2) \times C_T(t_1,t_2)$ or $C(\boldsymbol{s}_1,\boldsymbol{s}_2;t_1,t_2) = C_S(\boldsymbol{s}_1,\boldsymbol{s}_2) + C_T(t_1,t_2)$, for all space-time coordinates $(\boldsymbol{s}_1, t_1)$, $(\boldsymbol{s}_2, t_2) \in \mathcal{D} \times \mathbb{R}$. Given that C_S and C_T are valid covariance functions on $\mathcal{D}$ and $\mathbb{R}$, the product or the sum is valid on $\mathcal{D} \times \mathbb{R}$. It has been argued that the class of separable covariance models is severely limited due to the lack of space-time interaction [13], and that in

many cases it implies "unphysical dependence among process variables" [26]. We therefore consider non-separable covariance functions only in this paper.

1.3. Overview and contributions

In this paper, we adopt both the space embedding technique and the scale mixture approach to construct a broad range of valid space-time covariance functions on generalized linear networks and/or Euclidean trees. The rest of the paper is organized as follows. Section 2 reviews preliminaries about generalized linear networks equipped with two distance metrics: resistance distance and geodesic distance. Section 3 gives sufficient conditions for constructing isotropic space-time covariance models by space embedding on arbitrary generalized linear networks and then on an important subgroup – Euclidean trees. Besides deriving space-time covariance functions on directed Euclidean trees based on the scale mixture approach and the convex cone property in Section 4, we also show that the exponential tail-down model [33] is the one and only that is directionless (i.e. isotropic) and is thereby a bridge between models in Section 3 and Section 4. In Section 5, we investigate the performance of maximum likelihood estimators of proposed covariance functions and compare different models based on likelihood, in a simulation study. Section 6 concludes the paper with discussion.

2. Preliminaries

2.1. Generalized linear networks

A network, also called a graph, is a collection of vertices (nodes) joined by edges [23] and is denoted by the pair $(\mathcal{V}, \mathcal{E})$. A linear network is the union of finitely many line segments in the plane where different edges only possibly intersect with each other at one of their vertices (see Fig. 1). It is useful to associate each edge with a positive real number, which is called the weight. Weights can be physical edge lengths, strengths, etc. The space-time covariance functions in our paper are defined on generalized linear network, i.e. a triple $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \{\xi_e\}_{e \in \mathcal{E}})$, whose definition was introduced by [1] and is revisited below.

Definition 2.1. A triple $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \{\xi_e\}_{e \in \mathcal{E}})$ which satisfies conditions (I)–(IV) is called a graph with Euclidean edges.

(I) Graph structure: $(\mathcal{V}, \mathcal{E})$ is a finite simple connected graph, meaning that the vertex set $\mathcal{V}$ is finite, the graph has no self-edges or multi-edges, and every pair of vertices is connected by a path.

(II) Edge sets: Each edge $e \in \mathcal{E}$ is associated with a unique abstract set, also denoted by e. The vertex set $\mathcal{V}$ and all the edge sets are mutually disjoint.

(III) Edge coordinates: For each edge $e \in \mathcal{E}$ and vertices $u, v \in \mathcal{V}$ joined by e, there is a bijective mapping ξ_e defined on the union of the edge set e and vertices $\{u, v\}$, i.e. $e \cup \{u, v\}$, such that ξ_e maps e onto an open interval $(\underline{e}, \overline{e}) \subset \mathbb{R}$ and $\{u, v\}$ onto endpoints $\{\underline{e}, \overline{e}\}$, respectively.
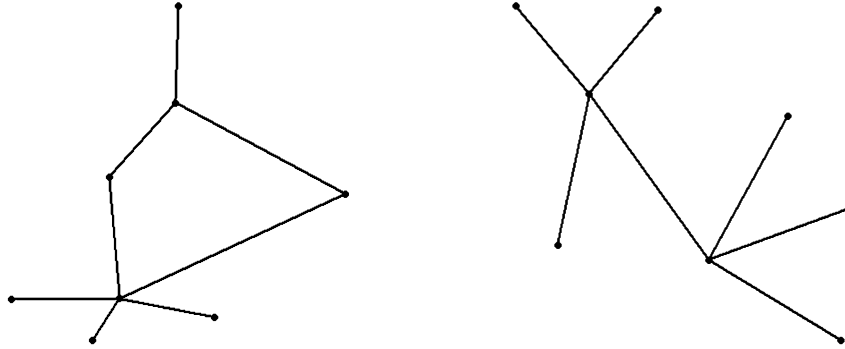
FIG 1. *Examples of networks. The right panel is a directionless tree.*

(IV) Distance consistency: Define $d_{\mathcal{G}}(u,v) : \mathcal{V} \times \mathcal{V} \to [0,\infty)$ as the length of the shortest path on vertices of a weighted graph where the weight associated with each edge $e \in \mathcal{E}$ is defined as $\overline{e} - \underline{e}$. Then, the following equality holds:

$$d_{\mathcal{G}}(u,v) = \overline{e} - \underline{e}$$

for every $e \in \mathcal{E}$ connecting two vertices $u, v \in \mathcal{V}$.

In our work, we assume that the topological structure of $\mathcal{G}$ does not evolve over time (e.g. when considering applications on stream networks, it takes a relatively long time for a stream to change its course. Thus, it is reasonable to assume $\mathcal{G}$ is stable during a short time window.). Any arbitrary site $\boldsymbol{s}$ on such a network $\mathcal{G}$ is denoted as $\boldsymbol{s} \in \mathcal{G} = \boldsymbol{s} \in \mathcal{V} \cup \bigcup_{e \in \mathcal{E}}$. Graphs with Euclidean edges extend the notion of traditional linear networks by including graphs which do not have a planar representation in $\mathbb{R}^2$ (see, e.g., Fig. 3 of [1]). We use the terms generalized linear networks and graphs with Euclidean edges interchangeably in this paper. Any tree-like graph $(\mathcal{V}, \mathcal{E})$ is planar and can be constructed as a graph with Euclidean edges easily. We call such a graph a Euclidean tree and denote it by $\mathcal{T}$. Vertices of a Euclidean tree connected with only one edge are called leaves.

2.2. Distance metrics

Let $\Phi(\mathcal{D}, d)$ denote the class of radial profile functions such that for any $f \in \Phi(\mathcal{D}, d)$,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j f\big(d(\boldsymbol{s}_i, \boldsymbol{s}_j)\big) \geq 0, \tag{2.1}$$

for any finite collection $\{a_i\}_{i=1}^{N} \subset \mathbb{R}$ and $\{\boldsymbol{s}_i\}_{i=1}^{N} \subset \mathcal{D}$. For any $f \in \Phi(\mathcal{D}, d)$, we say that f is positive definite on $\mathcal{D}$ with respect to distance d. When the space domain is a finite-dimensional Euclidean space $\mathcal{D} = \mathbb{R}^n$, for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

where $\boldsymbol{x} = (x_1, \ldots, x_n)'$ and $\boldsymbol{y} = (y_1, \ldots, y_n)'$, let d_p be the standard l_p norm with $d_p(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_p = (\sum_{i=1}^n |x_i - y_i|^p)^{1/p}$, $1 \le p < \infty$. When the space domain is a real Hilbert space $\mathcal{H}$, the norm, denoted by $\|\cdot\|_{\mathcal{H}}$ is induced by the inner product $< \cdot, \cdot >$, such that $\|\boldsymbol{x}\|_{\mathcal{H}} := \sqrt{< \boldsymbol{x}, \boldsymbol{x} >}$, $\boldsymbol{x} \in \mathcal{H}$.

A generalized linear network $\mathcal{G}$ comes along with two distance metrics: one is the standard length of shortest path, a.k.a. geodesic distance or stream distance, denoted by $d_{G,\mathcal{G}}$; the other is resistance distance $d_{R,\mathcal{G}}$. $d_{R,\mathcal{G}}$ is defined as the variogram of an auxiliary random field $Y_{\mathcal{G}}$: $d_{R,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_2) := \mathrm{Var}(Y_{\mathcal{G}}(\boldsymbol{s}_1) - Y_{\mathcal{G}}(\boldsymbol{s}_2))$, $\forall \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{G}$, whose formal construction is given by [1] and we provide a brief introduction in Supplement A. The resistance metric is an extension of the one in electrical network theory from pairs of vertices to any points on the graph. Both metrics satisfy the two conditions mentioned in Section 1.2 (i.e. symmetry and identifiability) and have the relationship given in Proposition 2.1, which is a portion of Proposition 4 of [1]:

Proposition 2.1. *For a graph with Euclidean edges $\mathcal{G}$, $d_{R,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_2) \le d_{G,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_2)$, $\forall \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{G}$. Equality holds if and only if $\mathcal{G}$ is a Euclidean tree.*

Since, by Proposition 2.1, $d_{R,\mathcal{T}} = d_{G,\mathcal{T}}$, henceforth we let $d_{\cdot,\mathcal{T}}$ denote either one.

3. Isotropic space-time models by space embedding

In this section, we adopt the space embedding technique to transform the abstract, less familiar spatial domain, $\mathcal{G}$, to simpler, well-studied spaces, e.g. Hilbert and Euclidean spaces, and build isotropic space-time models from there.

We follow the definition of isometric spaces given by [1] and [20]. Define a distance space as a pair $(\mathcal{D}, d)$, where $\mathcal{D}$ is a non-empty set and the function d is specified in Section 1.2.

Definition 3.1. A distance space $(\mathcal{D}, d)$ is said to be g-embeddable in a Hilbert space $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ if $g : [0, \infty) \to [0, \infty)$, and there exists a mapping $i : \mathcal{D} \to \mathcal{H}$ such that

$$g\big(d(\boldsymbol{s}_1, \boldsymbol{s}_2)\big) = \|i(\boldsymbol{s}_1) - i(\boldsymbol{s}_2)\|_{\mathcal{H}}, \qquad \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{D}.$$

If function g is the identity map, we say $(\mathcal{D}, d)$ is isometrically embeddable in $\mathcal{H}$.

3.1. Hilbert space embedding of generalized linear network

Our first main contribution is based on the square-root embedding result of a graph with Euclidean edges into a Hilbert space proved by [1] and restated below.

Theorem 3.1 (Square-root embedding, Anderes et al.)**.** *Given $\mathcal{G}$ a graph with Euclidean edges, there exists a Hilbert space $\mathcal{H}$ and a mapping $i : \mathcal{G} \to \mathcal{H}$ such that*

$$\sqrt{d_{R,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_2)} = \|i(\boldsymbol{s}_1) - i(\boldsymbol{s}_2)\|_{\mathcal{H}}$$

for all $s_1, s_2 \in \mathcal{G}$. In the special case in which $\mathcal{G}$ is a Euclidean tree, the above result also holds for the geodesic distance.

The so-called Gneiting class of covariance functions has been especially popular in space-time geostatistical modeling [26] and will be the building block of the isotropic covariance functions given in this section. Despite some discrepancy in the literature, here a function $\varphi : [0, \infty) \to \mathbb{R}$ is said to be completely monotone on $[0, \infty)$ if φ is continuous on $[0, \infty)$, infinitely differentiable on $(0, \infty)$ and $(-1)^j \varphi^{(j)}(t) \geq 0$ over $(0, \infty)$ for every integer $j \geq 0$, where $\varphi^{(j)}$ denotes the jth derivative of φ and $\varphi^{(0)} = \varphi$. A nonnegative continuous function $\psi(t) : [0, \infty) \to \mathbb{R}$ with completely monotone derivative is called a Bernstein function. In analogy to the definition of positive definite functions, we recall that for a distance space $(\mathcal{D}, d)$, a continuous function $f : \mathcal{D} \to \mathbb{R}$ is called *conditionally negative definite* (see for example in [20]) on $\mathcal{D}$ with respect to d if

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j f\big(d(s_i, s_j)\big) \leq 0, \tag{3.1}$$

for any finite collections of $\{a_i\}_{i=1}^{N} \subset \mathbb{R}$ and $\{s_i\}_{i=1}^{N} \subset \mathcal{D}$, given $\sum_{i=1}^{N} a_i = 0$. For such a function we write $f \in CND(\mathcal{D}, d)$.

We are now ready to formulate and prove our first main result. Denote the generalized *Gneiting* class of continuous functions by G_α, where

$$G_\alpha(d, u) = \frac{1}{\psi(d)^\alpha} \varphi\bigg(\frac{u}{\psi(d)}\bigg), \quad d, u \geq 0,$$

with ψ and φ being strictly positive and continuous. Theorem 3.2 in [20] provides sufficient conditions for G_α to be positive definite over the product space of a quasi-metric space and a finite Euclidean space, which extends [16]'s results in Euclidean spaces and will be applied directly on $\mathcal{G} \times \mathbb{R}$.

Theorem 3.2 (Generalized Gneiting class, an application of Menegatto et al.)**.** *Let G_α be the function defined above with $\varphi(\cdot)$ being completely monotone. Assume that $a \in (0, 1]$ and $\alpha \geq 1/2$. Then for any pairs $(s_1, t_1), (s_2, t_2) \in \mathcal{G} \times \mathbb{R}$, where $\mathcal{G}$ is equipped with the resistance distance $d_{R,\mathcal{G}}$, the following statements are true:*

- *(a) the function $C(d_{R,\mathcal{G}}(s_1, s_2); |t_1 - t_2|) := G_\alpha(d_{R,\mathcal{G}}(s_1, s_2), |t_1 - t_2|^{2a})$ is a valid covariance function over $\mathcal{G} \times \mathbb{R}$ provided that $\psi \in CND(\mathcal{G}, d_{R,\mathcal{G}})$;*
- *(b) the function $C(d_{R,\mathcal{G}}(s_1, s_2); |t_1 - t_2|) := G_\alpha(d_{R,\mathcal{G}}(s_1, s_2), |t_1 - t_2|^{2a})$ is a valid covariance function over $\mathcal{G} \times \mathbb{R}$ provided that $\psi := g \circ h$, where g is a positive Bernstein function and h is a nonnegative valued function such that $h \in CND(\mathcal{G}, d_{R,\mathcal{G}})$;*
- *(c) the function $C(d_{R,\mathcal{G}}(s_1, s_2); |t_1 - t_2|) := G_\alpha(d_{R,\mathcal{G}}(s_1, s_2)^b, |t_1 - t_2|^{2a})$ is a valid covariance function over $\mathcal{G} \times \mathbb{R}$ provided that $b \in (0, 1]$ and ψ is a positive Bernstein function.*

Moreover, when $\mathcal{G}$ is a Euclidean tree, the above results hold for $d_{G,\mathcal{G}}$ as well.

Proof. The symmetry of the functions defined in all three parts is obvious. A distance space $(\mathcal{D}, d)$ defined in this paper is also a quasi-metric space in [20], while the converse is not necessarily true. Therefore, parts (a) and (b) of Theorem 3.2 are direct applications of [20]'s work to the case where the dimension of the Euclidean space is 1 and the quasi-metric space is $(\mathcal{G}, d_{R,\mathcal{G}})$. By Theorem 3.1, $(\mathcal{G}, d_{R,\mathcal{G}})$ is square root-embeddable in a Hilbert space $\mathcal{H}$. Notice that $(\mathcal{G}, \sqrt{d_{R,\mathcal{G}}})$ is again a distance space and thus isometrically embeddable into $\mathcal{H}$ by Definition 3.1. Statement (c) follows in concert with Theorem 3.2 (iii) in [20]. When $\mathcal{G}$ is a Euclidean tree, Proposition 2.1 gives that $d_{R,\mathcal{G}} = d_{G,\mathcal{G}}$, which completes the proof. $\qquad\square$

Each part (a)–(c) of Theorem 3.2 provides researchers an easy-to-implement method for constructing valid space-time covariance functions on a generalized linear network. Let us digress for a moment and consider a pure spatial, isotropic random process $Z(\boldsymbol{s})$ defined on $(\mathcal{G}, d_{R,\mathcal{G}})$ with an overall constant mean μ. Similar to the conclusion in geostatistics, the semivariogram, defined as $\gamma\big(d_{R,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_2)\big) := \frac{1}{2}Var(Z(\boldsymbol{s}_1) - Z(\boldsymbol{s}_2))$ is conditionally negative definite, i.e. $\gamma \in CND(\mathcal{G}, d_{R,\mathcal{G}})$. This result holds for $d_{G,\mathcal{G}}$ as well when $\mathcal{G}$ is Euclidean tree. Following the previous notation, let $Cov(Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)) = C(\boldsymbol{s}_1, \boldsymbol{s}_2) = f(d_{R,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_2))$. Then by definition, there exists the following relationship

$$\gamma\big(d_{R,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_2)\big) = f(0) - f\big(d_{R,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_2)\big), \qquad (3.2)$$

for all $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{G}$, between γ and f. Hence, given any radial profile $f \in \Phi(\mathcal{G}, d_{R,\mathcal{G}})$, we can construct γ based on (3.2), which belongs to $CND(\mathcal{D}, d_{R,\mathcal{G}})$. For examples of the class $\Phi(\mathcal{G}, d_{R,\mathcal{G}})$, we refer to [1]. Note that statements (a)–(c) are not exclusive. For instance, $\psi(t) = t^\lambda + \beta$, for $t \geq 0$ with $0 < \lambda \leq 1, \beta > 0$ is a positive Bernstein function (given in Table 1) and also belongs to $CND(\mathcal{T}, d_{.,\mathcal{T}})$ (see Lemma A1 in Supplement A). When $b = 1$ in (c), both (a) and (c) give the same subclass of valid covariance functions on $\mathcal{T} \times \mathbb{R}$. A similar construction based on the functional form by [16] (i.e. G_α in our work) can be found in [27], where the spatial distance is re-scaled by the temporal distance, under certain condition.

In addition to constructing valid covariance functions on $\mathcal{G} \times \mathbb{R}$, we also investigate the geometric features of marginal functions whose definition will be given shortly, near the origin and at infinity. It has been discussed in [11] that spatial and temporal marginals, defined as $f_S(d) := f(d, 0)$ and $f_T(u) := f(0, u)$, respectively (where f denotes the space-time radial profile function), play a significant role in selecting an appropriate and physically meaningful class of covariances in applications. By comparing empirical covariance functions with estimated ones, any obvious disagreement would indicate model misspecification [31].

It is clear that the covariance function, as well as both marginal functions, constructed by Theorem 3.2 are continuous at the origin since ψ and φ are continuous on $[0, \infty)$. Although Lévy-Khinchin's formula [5] characterizes the class of conditionally negative definite functions in $\mathbb{R}^n$, analogous results in the distance space $(\mathcal{D}, d)$, especially $(\mathcal{G}, d_{R,\mathcal{G}})$, have not been obtained, as far as

TABLE 1
Examples of completely monotone functions $\varphi(t)$ and positive Bernstein functions $\psi(t)$, $t \geq 0$.

Function	Parameters	Function	Parameters
$\varphi(t) = \exp(-ct^{\nu})$	$c > 0, 0 < \nu \leq 1$	$\psi(t) = (\kappa t^{\lambda} + 1)^{\beta}$	$\kappa > 0, 0 \leq \beta \leq 1, 0 < \lambda \leq 1$
$\varphi(t) = \exp(ct^{\nu})$	$c > 0, \nu < 0$	$\psi(t) = \frac{\log(\kappa t^{\lambda} + \beta)}{\log(\beta)}$	$\kappa > 0, \beta > 1, 0 < \lambda \leq 1$
$\varphi(t) = (\frac{2}{\exp(ct^{1/2}) + \exp(-ct^{1/2})})^{\nu}$	$c > 0, \nu > 0$	$\psi(t) = t^{\lambda} + \beta$	$0 < \lambda \leq 1, \beta > 0$
$\varphi(t) = (1 + ct^{\gamma})^{-\nu}$	$c > 0, \nu > 0, 0 < \gamma \leq 1$	$\psi(t) = \beta - \exp(-\kappa t)$	$\kappa > 0, \beta > 1$

we know. Hence, we defer marginal results related to $CND(\mathcal{G}, d_{R,\mathcal{G}})$ for future research and investigate properties of marginals pertaining to the covariance functions by Theorem 3.2(c) only.

Proposition 3.3. *Let $C(d_{R,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_1); |t_1 - t_2|) := G_{\alpha}(d_{R,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_2)^b, |t_1 - t_2|^{2a})$, where $\alpha \geq 1/2$, $a \in (0, 1]$, $b \in (0, 1]$, φ is completely monotone and ψ is a positive Bernstein function. Then the spatial f_S and temporal f_T marginal functions, i.e., $f_S(d) = G_{\alpha}(d^b, 0)$ and $f_T(u) = G_{\alpha}(0, u^{2a})$, are non-increasing over $[0, \infty)$.*

Proposition 3.4. *Let $C(d_{R,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_1); |t_1 - t_2|) := G_{\alpha}(d_{R,\mathcal{G}}(\boldsymbol{s}_1, \boldsymbol{s}_2)^b, |t_1 - t_2|^{2a})$, where $\alpha \geq 1/2$, $a \in (0, 1]$, $b \in (0, 1]$, φ is completely monotone and ψ is a positive Bernstein function. Then the spatial marginal function $f_S(d) = G_{\alpha}(d^b, 0)$, is convex on $(0, \infty)$.*

Proofs of Proposition 3.3 and 3.4 are given in Supplement A. Note that the temporal marginal function f_T does not share the convexity property in general. Justified by Proposition 3.3, space-time covariance functions constructed by Theorem 3.2(c) satisfy the physical law [32] which says observations that are closer in space and time have higher correlation. The asymptotic behavior of the model, e.g. $\lim_{d \to \infty} G_{\alpha}(d^b, u^{2a})$ for fixed u and $\lim_{u \to \infty} G_{\alpha}(d^b, u^{2a})$ for fixed d, depends on the asymptotic behavior of ψ and φ, respectively, thus does not present a unified conclusion, in general.

We list a few examples of completely monotone functions and positive Bernstein functions in Table 1, which can be found in [16, 21, 5]. More completely monotone functions can be built by constructive tools, such as the property that the class of functions is closed under addition and multiplication [21]. We show the application of Theorem 3.2 by a couple of examples.

Example 1. Consider the completely monotone function and the Bernstein function from the first row of Table 1. To avoid the model being overly complicated, assume $\alpha = \frac{1}{2}$, $a = 1$ and write the geodesic distance between sites as d and the time lag as u throughout. The space-time covariance function C_0 given by Theorem 3.2(c) may be written as follows:

$$C_0(d; u) = \frac{1}{(\kappa d^{b\lambda} + 1)^{\beta/2}} \exp\left(-c\left(\frac{u^2}{(\kappa d^{b\lambda} + 1)^{\beta}}\right)^{\nu}\right),$$

where $c > 0, 0 < \nu \leq 1, \kappa > 0, 0 \leq \beta \leq 1, 0 < \lambda \leq 1$ and $0 < b \leq 1$. Since b and λ appear only as the product $b\lambda$, and both have the same support, i.e. $(0, 1]$, henceforth to avoid an identification issue we use b instead of $b\lambda$.

Imitating Example 1 from [16], consider the pure spatial covariance function C_S [1] on $(\mathcal{G}, d_{R,\mathcal{G}})$: $C_S(d) = (\kappa d^b + 1)^{-\delta}$, where $\kappa > 0, 0 < b \leq 1$ and $\delta \geq 0$. By the Schur product theorem [30], it follows that the product of C_0 and C_S also defines a valid space-time covariance function on $\mathcal{G} \times \mathbb{R}$. After reparameterization (i.e. let $\tau = \frac{\beta}{2} + \delta$), we have

$$C(d; u) = \frac{1}{(\kappa d^b + 1)^\tau} \exp\left(-c\left(\frac{u^2}{(\kappa d^b + 1)^\beta}\right)^\nu\right), \qquad (3.3)$$

where $c > 0, 0 < \nu \leq 1, \kappa > 0, 0 \leq \beta \leq 1, \tau \geq \frac{\beta}{2}$ and $0 < b \leq 1$. We call β the space-time interaction parameter since when $\beta = 0$, the covariance function becomes separable. The spatial and temporal marginals, as well as the covariance function itself, all decay to zero as $d \to \infty$ and/or $u \to \infty$, which indicates the same variability in the spatial and temporal components in the sense discussed by [13] (visualizations of marginal functions, along with the covariance surface can be found in Suppplement B). We will come back to the model given by (3.3) later.

Example 2. Let $\varphi(t) = (\frac{2}{\exp(ct^{1/2}) + \exp(-ct^{1/2})})^\nu$ and $\psi(t) = t^\lambda + 1$, where $t \geq 0$ with parameters $c > 0, \nu > 0$, and $0 < \lambda \leq 1$. Then Theorem 3.2(c) gives another valid space-time covariance function:

$$C(d; u) = \frac{2^\nu}{(d^{b\lambda} + 1)^\alpha} \left\{ \exp\left(c\frac{u^a}{(d^{b\lambda} + 1)^{1/2}}\right) + \exp\left(-c\frac{u^a}{(d^{b\lambda} + 1)^{1/2}}\right) \right\}^{-\nu}, \quad (3.4)$$

where $0 < a \leq 1, \alpha \geq 1/2, 0 < b \leq 1, c > 0, \nu > 0$, and $0 < \lambda \leq 1$. Again, λ only appears as a multiplier of b and both have the same support, i.e. $(0, 1]$, so we will drop λ. The asymptotic behaviors of (3.4), and its corresponding marginals, are the same as in Example 1.

3.2. l_1 *embedding of Euclidean trees*

Now, we focus on a subclass of generalized linear networks $\mathcal{G}$: the Euclidean trees $\mathcal{T}$. In addition to the square-root embedding result which holds for any $\mathcal{G}$, [1] provides another space embedding result for $\mathcal{T}$ only, which is restated below.

Theorem 3.5 (l_1 embedding, Anderes et al.)**.** *Let $\mathcal{T}$ be a Euclidean tree with m leaves, where $m \geq 3$. Then $(\mathcal{T}, d_{\cdot, \mathcal{T}})$ is isometrically embeddable into $(\mathbb{R}^n, \rho_1)$, where $n = \lceil \frac{m}{2} \rceil$ and ρ_1 is the l_1 norm, such that there exists a mapping $i : \mathcal{T} \to \mathbb{R}^n$ satisfying: $d_{\cdot, \mathcal{T}}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \rho_1(i(\boldsymbol{s}_1) - i(\boldsymbol{s}_2))$, for any $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{T}$.*

The l_1 embedding result comes directly from the proof pertaining to Theorem 4 in [1]. Meanwhile, the positive definite functions on $(\mathbb{R}^n, \rho_1)$ are essentially the same as α-symmetric (here $\alpha = 1$) characteristic functions in $\mathbb{R}^n$ by Bochner's theorem, which have been extensively studied by [7, 15, 36] and will play a fundamental role in constructing space-time covariance functions on

$\mathcal{T} \times \mathbb{R}$. Before we dive into our second main contribution, we give the definition of *linear* isometric embedding due to [36]. The term *linear* is added to distinguish from Definition 3.1.

Definition 3.2. Suppose that $\mathcal{L}_i$ is a linear space, and a function $\rho_i : \mathcal{L}_i \to [0, \infty)$ exists, that satisfies $\rho_i(c\boldsymbol{x}) = |c|\rho_i(\boldsymbol{x})$, for any scalar c and $\boldsymbol{x} \in \mathcal{L}_i$, $i = 1, 2$. The pair $(\mathcal{L}_1, \rho_1)$ is said to be *linearly* isometrically embeddable in $(\mathcal{L}_2, \rho_2)$ if there is a linear operator $A : \mathcal{L}_1 \to \mathcal{L}_2$ such that $\rho_1(\boldsymbol{x}) = \rho_2(A\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{L}_1$. If either of $(\mathcal{L}_1, \rho_1)$ and $(\mathcal{L}_2, \rho_2)$ is linearly isometrically embeddable in the other, we call these spaces *linearly* isometric.

Being slightly different from (2.1), let $\Phi(\mathbb{R}^n, \rho)$ denote the class of functions such that for any $f \in \Phi(\mathbb{R}^n, \rho)$, $\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j f(\rho(\boldsymbol{x}_i - \boldsymbol{x}_j)) \geq 0$, for any finite collection $\{a_i\}_{i=1}^{N} \subset \mathbb{R}$ and $\{\boldsymbol{x}_i\}_{i=1}^{N} \subset \mathbb{R}^n$. For any such f and ρ, we say that $f \circ \rho$ is positive definite on $\mathbb{R}^n$.

Lemma 3.6. *Consider* $(\mathbb{R}^n, \rho_1)$ *and* $(\mathbb{R}^n, \rho_2)$, *where* ρ_1 *is the* l_1 *norm and* $\rho_2(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{1}{c_i} |x_i|$ *with* $\{c_i\}_{i=1}^{n}$ *being fixed positive scalars, for* $\boldsymbol{x} \in \mathbb{R}^n$. *Then we have* $\Phi(\mathbb{R}^n, \rho_1) = \Phi(\mathbb{R}^n, \rho_2)$.

Proof. Let A be the diagonal matrix with elements $c_1, \ldots, c_n$ on the main diagonal. Then for all $\boldsymbol{x} \in \mathbb{R}^n$, we have $\rho_1(\boldsymbol{x}) = \sum_{i=1}^{n} |x_i| = \sum_{i=1}^{n} \frac{1}{c_i} |c_i x_i| = \rho_2(A\boldsymbol{x})$. By Definition 3.2, $(\mathbb{R}^n, \rho_1)$ and $(\mathbb{R}^n, \rho_2)$ are linearly isometric. Lemma 3.6 then follows from Lemma 2(2) in [36]. $\qquad \square$

Let $\boldsymbol{\theta}$ denote the vector of covariance parameters and Θ_n the parameter space with subscript n emphasizing that the dependence relates to $\mathbb{R}^n$. The theorem below gives a general framework for constructing valid space-time covariance functions on $\mathcal{T} \times \mathbb{R}$ by l_1 embedding.

Theorem 3.7 (Metric models)**.** *Suppose that* $\mathcal{T}$ *is a Euclidean tree with* m *leaves, where* $m \geq 3$. *Define* $n = \lceil m/2 \rceil$. *If* $f_{\boldsymbol{\theta}} \in \Phi(\mathbb{R}^{n+1}, \rho_1)$, *where* $\boldsymbol{\theta} \in \Theta_{n+1}$, *then* $C(d_{\cdot, \mathcal{T}}; u) := f_{\boldsymbol{\theta}}(\frac{d_{\cdot, \mathcal{T}}}{\alpha} + \frac{u}{\beta})$, *where* $\alpha, \beta > 0$ *and* $\boldsymbol{\theta} \in \Theta_{n+1}$, *is a valid covariance function on* $\mathcal{T} \times \mathbb{R}$.

Proof. Let $\alpha = c_1 = \cdots = c_n$ and $\beta = c_{n+1}$ in ρ_2, from Lemma 3.6. It follows that if $f_{\boldsymbol{\theta}}(\rho_1(\boldsymbol{x})) = f_{\boldsymbol{\theta}}(\sum_{i=1}^{n+1} |x_i|)$ is positive definite on $\mathbb{R}^{n+1}$ for $\boldsymbol{\theta} \in \Theta_{n+1}$, then $f_{\boldsymbol{\theta}} \circ \rho_2 = f_{\boldsymbol{\theta}}(\frac{\sum_{i=1}^{n} |x_i|}{\alpha} + \frac{|x_{n+1}|}{\beta})$ is also positive definite on $\mathbb{R}^{n+1}$ given $\alpha, \beta > 0$, in addition to $\boldsymbol{\theta} \in \Theta_{n+1}$. In concert with Theorem 3.5, there exists a mapping $i : \mathcal{T} \to \mathbb{R}^n$ such that

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j C\big(d_{\cdot, \mathcal{T}}(\boldsymbol{s}_i, \boldsymbol{s}_j); |t_i - t_j|\big)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j f_{\boldsymbol{\theta}} \left(\frac{d_{\cdot, \mathcal{T}}(\boldsymbol{s}_i, \boldsymbol{s}_j)}{\alpha} + \frac{|t_i - t_j|}{\beta} \right)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j f_{\boldsymbol{\theta}} \left(\frac{\rho_1(i(\boldsymbol{s}_i) - i(\boldsymbol{s}_j))}{\alpha} + \frac{|t_i - t_j|}{\beta} \right)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j f_{\boldsymbol{\theta}} \left(\frac{\sum_{k=1}^{n} |i(\boldsymbol{s}_i)_k - i(\boldsymbol{s}_j)_k|}{\alpha} + \frac{|t_i - t_j|}{\beta} \right)$$
$$\geq 0,$$

for any finite collection $\{a_i\}_{i=1}^{N} \subset \mathbb{R}$ and points $\{\boldsymbol{s}_i\}_{i=1}^{N} \subset \mathcal{T}$, where $\alpha, \beta > 0$ and $\boldsymbol{\theta} \in \Theta_{n+1}$. □

The scaling parameters, α and β, make the spatial and temporal distances comparable. In the literature, $\frac{d_{\cdot, \mathcal{T}}}{\alpha} + \frac{u}{\beta}$ has been called the space-time distance and the models given by Theorem 3.7 have been called metric models [13]. Combining Theorem 3.7 with the sufficient conditions of 1-symmetric characteristic functions given by [7, 15], we have Corollary A1 given in Supplement A.

In Euclidean space, metric models have received criticism for their equal treatment of the space and time dimensions. Nevertheless, metric models are still considered one of the basic classes of spatio-temporal covariance functions [18], and their flexibility can be extended when coupled with the convex cone property [17]. When such a model would perform well is, as for other models, a data-driven model selection question.

We end this section with an explicit example, by applying Theorem 3.7 to a 1-symmetric characteristic function.

Example 3 (Powered linear with sill models). The parametric model given by the following corollary belongs to the family of metric models and has the powered linear with sill representation.

Corollary 3.8. *Suppose that $\mathcal{T}$ is a Euclidean tree with m leaves, where $m \geq 3$. Then $C(d_{\cdot, \mathcal{T}}; u) := (1 - (\frac{d_{\cdot, \mathcal{T}}}{\alpha} + \frac{u}{\beta})^{\nu})_{+}^{\delta}$, where $\delta \geq 2\lceil \frac{m}{2} \rceil + 1$, $\alpha, \beta > 0$ and $\nu \in (0, 1]$, is a valid covariance function on $\mathcal{T} \times \mathbb{R}$.*

Proof. Let $f(x) := (1 - x^{\nu})_{+}^{\delta}$, where $x \geq 0$. It is clear from [36] that $f \circ \rho_1$ is positive definite on $\mathbb{R}^n$, given $\delta \geq 2n-1$ and $\nu \in (0, 1]$, for any positive integer n. Therefore, $f(\frac{d_{\cdot, \mathcal{T}}}{\alpha} + \frac{u}{\beta}) = (1 - (\frac{d_{\cdot, \mathcal{T}}}{\alpha} + \frac{u}{\beta})^{\nu})_{+}^{\delta}$ with $\delta \geq 2\lceil \frac{m}{2} \rceil + 1, \nu \in (0, 1]$ and $\alpha, \beta > 0$ is positive definite on $\mathcal{T} \times \mathbb{R}$ by Theorem 3.7. □

The powered linear with sill model given by Corollary 3.8 is continuous near the origin. Both marginals, as well as the covariance function itself, monotonically decay to zero. By direct calculation, one can show that both spatial and temporal marginal functions are convex near the origin. One feature of this model is that it has compact support, i.e. it reaches zero when the space-time distance is sufficiently large. This property is appealing for modeling large scale space-time datasets. [27] also included covariance models with compact support for the case of Euclidean trees with a given number of leaves.

4. Space-time models on directed Euclidean trees

Instead of working with heavily mathematically involved functions, the kernel convolution-based (or moving average) models tackle the problem from another

perspective. According to [35], a large class of stationary covariance functions on the real line can be obtained by constructing a random process $\{Z(x) : x \in \mathbb{R}\}$, which convolves a square-integrable kernel function $g(\cdot)$ over a white noise process $Y(x)$ defined on $\mathbb{R}^1$ as: $Z(x) = \int_{-\infty}^{\infty} g(s - x)dY(s)$, $x, s \in \mathbb{R}$. When $Y(x)$ is Brownian motion, the induced covariance function is valid and can be shown to be $Cov(Z(x), Z(x+h)) = C(x, x+h) = \int_{-\infty}^{\infty} g(x)g(x-h)dx$, $x, h \in \mathbb{R}$. The kernel convolution-based approach allows considerable flexibility and can be generalized to nonstationary [14], space-time [28] and tree-like network [33] settings. Details of the latter generalization are given below.

4.1. Tail-up and tail-down models

The space-time covariance functions given in the previous section are isotropic, which might not always be an appropriate assumption due to the fact that some networks are directed in nature. For instance, in streams, flow direction is yet another important factor, in addition to shortest path length (geodesic distance), that researchers should take into consideration when modeling physical processes. Variables that move passively downstream, e.g. chemical particles, and variables that may move upstream, e.g. fish and insects [33] may need to be modeled differently. Especially for the former, we may want to allow the correlation between locations that do not share flow to be small or even zero. Based on the kernel convolution approach, [34, 33] introduce the unilateral tail-up (i.e. C_{TU}) and tail-down (i.e. C_{TD}) covariance models on streams, which manage to handle these two scenarios differently. For detailed discussion of the models, we refer readers to [33]. Here, we only give the most necessary background, which will later become essential components in our space-time covariance functions on tree-like networks. Despite the similarity between [27] and some of our work, models proposed in the former cannot incorporate the important geographical feature, directionality of the network.

The dendritic structure of streams guarantees that condition (I) in Definition 2.1 holds. Since every tree-like network is planar, we follow the prescription of [1] by letting each edge set $e \in \mathcal{E}$ be the interior of the corresponding line segment in $\mathbb{R}^2$ and letting $\mathcal{V}$ be the set of endpoints of the line segments. Moreover, let the bijection ξ_e preserve the path-length parameterization of each line segment. Thus, conditions (I)–(IV) in Definition 2.1 are satisfied and a stream equipped with stream distance, denoted by $(\mathcal{T}, d_{\cdot,\mathcal{T}})$, is a (directed) Euclidean tree. Note that models built in this section can apply to any directed Euclidean tree, which we call a stream for convenience.

Depending on the flow direction, the tail-up and tail-down models also assume there exists a single most downstream location, which is called the outlet (see for example Fig. 2). Let the index set of all stream segments be denoted by A, and let the index set of stream segments that are upstream of site $\boldsymbol{s}_i \in \mathcal{T}$, including the segment where $\boldsymbol{s}_i$ resides, be denoted by $U_i \subseteq A$. Two sites $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ are said to be "flow-connected" if they share water, i.e. if $U_i \cap U_j \neq \emptyset$, and are "flow-unconnected" if the water at one location does not flow to the other,
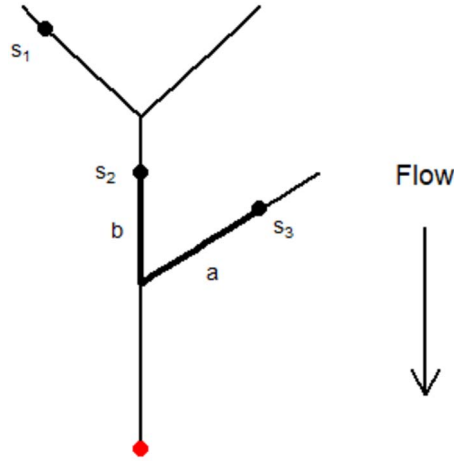
FIG 2. *Directed Euclidean tree with outlet (in red) superimposed. s_1 and s_2 are flow-connected, while s_2 and s_3, and s_1 and s_3 are flow-unconnected.*

i.e. if $U_i \cap U_j = \emptyset$. Equivalently, two sites are called flow-connected if and only if one is on the path of the other downstream to the outlet. The pure spatial tail-up and tail-down models are given below, where the unilateral kernel $g(x)$ is nonzero only when $x > 0$.

- Tail-up models:

$$C_{TU}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \begin{cases} \pi_{1,2} \int_d^\infty g(x)g(x-d)dx & \text{if } \boldsymbol{s}_1, \boldsymbol{s}_2 \text{ are flow-connected} \\ 0 & \text{if } \boldsymbol{s}_1, \boldsymbol{s}_2 \text{ are flow-unconnected,} \end{cases}$$

where d is the stream distance between sites $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ (i.e. $d_{.,\mathcal{T}}(\boldsymbol{s}_1, \boldsymbol{s}_2)$) and $\pi_{1,2}$ is a weight defined as follows. Let $\Omega(x)$ be a positive additive function such that $\Omega(x)$ is constant within a stream segment, but is the sum of each segment's value when two segments join at a junction, following the flow direction. Then the weight $\pi_{1,2} = \sqrt{\frac{\Omega(\boldsymbol{s}_1)}{\Omega(\boldsymbol{s}_2)}} \wedge \sqrt{\frac{\Omega(\boldsymbol{s}_2)}{\Omega(\boldsymbol{s}_1)}}$ ensures a constant variance of the process. In the literature, there exist different weighting schemes, see [9, 34], and they have been proven equivalent [33].
- Tail-down models:

$$C_{TD}(\boldsymbol{s}_1, \boldsymbol{s}_2)$$
$$= \begin{cases} \int_{-\infty}^{-d} g(-x)g(-x-d)dx & \text{if } \boldsymbol{s}_1, \boldsymbol{s}_2 \text{ are flow-connected} \\ \int_{-\infty}^{-a \vee b} g(-x)g(-x-|b-a|)dx & \text{if } \boldsymbol{s}_1, \boldsymbol{s}_2 \text{ are flow-unconnected,} \end{cases}$$

where d has the same as in tail-up models and a, b represent the distances from each site to the nearest junction downstream of which it shares flow with the other site (see for example Fig. 2). When $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{T}$ are flow-unconnected, $d = a + b$.

Commonly used kernels on streams can be found in [33]. Obviously, a non-trivial tail-up covariance function cannot be isotropic as the covariance is always zero when sites are flow-unconnected, while a tail-down model is a function of a and b, in general. It has been shown by [34, 33] that when the kernel is exponential, i.e. $g(x) = \theta_1 \exp(-x/\theta_2)$ for $x \geq 0$, $\theta_1, \theta_2 > 0$, the tail-down model is a function of the geodesic distance $d_{.,\mathcal{T}}$ alone, regardless of flow-connectedness. Before introducing our next main contribution in terms of space-time covariance functions, we prove that the exponential kernel is the one and only which makes the tail-down model depend on $d_{.,\mathcal{T}}$ alone, or in other words, isotropic.

Theorem 4.1. *A tail-down model is isotropic, such that there exists a function* f_{TD}, $C_{TD}(\boldsymbol{s}_1, \boldsymbol{s}_2) = f_{TD}(d_{.,\mathcal{T}}(\boldsymbol{s}_1, \boldsymbol{s}_2))$ *for any* $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{T}$, *if and only if the kernel is exponential.*

The proof of Theorem 4.1 is nontrivial and left to Supplement A. When the kernel is exponential, the isotropic tail-down model can be written as

$$C_{TD}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \theta_0 \exp\bigl(-d_{.,\mathcal{T}}(\boldsymbol{s}_1, \boldsymbol{s}_2)/\theta_2\bigr), \qquad \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{T}, \tag{4.1}$$

where $\theta_0, \theta_2 > 0$. (4.1) also appears in [1], where all isotropic covariance functions are developed by space embedding, as a valid covariance function on $(\mathcal{G}, d_{R,\mathcal{G}})$. Therefore, Theorem 4.1 shows that the exponential tail-down covariance function is the only bridge which connects pure spatial covariance functions on Euclidean trees constructed by space embedding and kernel convolution, and will later help us find the linkage of space-time models constructed by different approaches as well.

4.2. Convex cone and scale mixture models

4.2.1. Convex cone

Stemming from the convex cone property of the class of positive definite functions, Theorem 4.2 provides easy to implement, yet practically important, ways to construct space-time covariance functions on directed Euclidean trees.

Theorem 4.2. *The functions given below are valid space-time covariance functions on a directed Euclidean tree:*

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2; t_1, t_2) = C_{TD}(\boldsymbol{s}_1, \boldsymbol{s}_2)C_{T1}(t_1, t_2) + C_{TU}(\boldsymbol{s}_1, \boldsymbol{s}_2) + C_{T2}(t_1, t_2) \tag{4.2}$$

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2; t_1, t_2) = C_{TU}(\boldsymbol{s}_1, \boldsymbol{s}_2)C_{T1}(t_1, t_2) + C_{TD}(\boldsymbol{s}_1, \boldsymbol{s}_2) + C_{T2}(t_1, t_2) \tag{4.3}$$

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2; t_1, t_2) = C_{TU}(\boldsymbol{s}_1, \boldsymbol{s}_2)C_{T1}(t_1, t_2) + C_{TD}(\boldsymbol{s}_1, \boldsymbol{s}_2)C_{T2}(t_1, t_2), \tag{4.4}$$

where the tail-up C_{TU} *and the tail-down* C_{TD} *models are defined in Section 4.1, and* C_{T1} *and* C_{T2} *are valid temporal covariance functions.*

Proof. The symmetry condition holds trivially as each component on the right hand side of (4.2)–(4.4) is symmetric. According the Schur product theorem [30], $C_{TD}(\cdot, \cdot)C_{T1}(\cdot, \cdot)$, $C_{TU}(\cdot, \cdot)C_{T1}(\cdot, \cdot)$, and $C_{TD}(\cdot, \cdot)C_{T2}(\cdot, \cdot)$, are positive definite

on $\mathcal{T} \times \mathbb{R}$. The remaining results then follow easily from the definition of positive definiteness. $\qquad\qquad\square$

In Euclidean space, (4.2) and (4.3) are called product-sum models [12]. Unless $C_{T1} = C_{T2}$ in (4.4), covariance functions given in Theorem 4.2 are non-separable. Similar to the variance components model in [33], these functions allow high autocorrelation among sites that are flow-connected, and small but significant autocorrelation among sites that are flow-unconnected, at fixed temporal components. If we further assume that the number of observations over time on each site is the same, then substantial computational efficiency can be gained by exploiting the covariance matrix structure; details are discussed in Supplement A.

Example 4. Consider the isotropic exponential tail-down model, C_{T1} being a cosine function which captures potential seasonal fluctuations, and C_{T2} exponential as well. For the tail-up spatial component, we adopt a Mariah kernel [33], which specifies $g(x) = \frac{1}{2}\frac{1}{1+x/\theta_1}$, for $x \geq 0$ with $\theta_1 > 0$. After reparameterization, expression (4.4) from Theorem 4.2 gives the valid space-time model as

$$
C(\boldsymbol{s}_1, \boldsymbol{s}_2; t_1, t_2) =
\begin{cases}
\frac{\pi_{1,2}}{2}\frac{\log(d/\theta_1+1)}{d/\theta_1}\cos\left(\frac{u}{\theta_2}\right) + \frac{1}{2}\exp\left(-\left(\frac{d}{\theta_3}+\frac{u}{\theta_4}\right)\right) \\
\quad \boldsymbol{s}_1, \boldsymbol{s}_2 \text{ are flow-connected, } d > 0 \\
\frac{1}{2}\cos\left(\frac{u}{\theta_2}\right) + \frac{1}{2}\exp\left(-\frac{u}{\theta_4}\right) & d = 0 \\
\frac{1}{2}\exp\left(-\left(\frac{d}{\theta_3}+\frac{u}{\theta_4}\right)\right) & \text{otherwise,}
\end{cases}
$$

where $\theta_1, \ldots, \theta_4 > 0$ and the weight $\pi_{1,2}$ is defined in Section 4.1. The model above contains a metric sub-model which is a function of the space-time distance, i.e. $\frac{d}{\theta_3} + \frac{u}{\theta_4}$.

4.2.2. Scale mixture models

We conclude the theoretical development of space-time models with a clever trick [26], which gives the so-called scale mixture model. The trick can trace back to the second stability property of covariance functions given by [8].

Theorem 4.3. *Let* $C_0(\boldsymbol{s}_1, \boldsymbol{s}_2; t_1, t_2; a)$ *be a space-time covariance model on* $\mathcal{G} \times \mathbb{R}$, a *be a parameter where* $a \in \Theta_a \subset \mathbb{R}$, *and* $\mu(\cdot)$ *a positive measure on the set* Θ_a. *Then*

$$
C(\boldsymbol{s}_1, \boldsymbol{s}_2; t_1, t_2) = \int_{\Theta_a} C_0(\boldsymbol{s}_1, \boldsymbol{s}_2; t_1, t_2; a) d\mu(a),
$$

$(\boldsymbol{s}_i, t_i) \in \mathcal{G} \times \mathbb{R}, i = 1, 2$, *is a valid covariance model on* $\mathcal{G} \times \mathbb{R}$ *given that the integral exists for every pair of space-time coordinates.*

Theorem 4.3 can be proved by the definition of positive definiteness directly, which we will skip here. Any valid space-time model that satisfies the condition can be chosen as the integrand, and we emphasize its application on directed Euclidean trees in the following corollary.

Corollary 4.4. *Suppose that $C_S(\cdot,\cdot)$ is a pure spatial covariance function on a directed Euclidean tree $\mathcal{T}$ and $C_T(\cdot,\cdot)$ is a pure temporal covariance function. Parameter a has the support $\Theta_a \subset \mathbb{R}$, and $\mu(\cdot)$ is a positive measure on the set Θ_a. Then*

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2; t_1, t_2) = \int_{\Theta_a} C_S(\boldsymbol{s}_1, \boldsymbol{s_2}; a) C_T(t_1, t_2; a) d\mu(a), \quad (\boldsymbol{s}_i, t_i) \in \mathcal{T} \times \mathbb{R}, i = 1, 2,$$

is a valid space-time covariance function on $\mathcal{T} \times \mathbb{R}$ given that the integral exists.

The proof of Corollary 4.4 follows from Theorem 4.3 in concert with the fact that the separable space-time function $C_S(\boldsymbol{s}_1, \boldsymbol{s_2}; a)C_T(t_1, t_2; a)$ is a valid covariance function on $\mathcal{T} \times \mathbb{R}$. We illustrate the use of Corollary 4.4 by two examples, one of which shows an interesting linkage between space-embedding models and scale mixture models due to Theorem 4.1.

Example 1 (Revisited)*.* Following Example 4 in [12], consider an exponential tail-down model as the spatial component (C_S), a cosine function as the temporal component (C_T) and a half-normal probability density function (μ'), which are parameterized as follows:

$$C_S(d; a) = \exp\left(-\frac{a^2}{\theta_1}d\right), \quad C_T(u; a) = \cos\left[a(2\theta_2 u)\right], \quad \mu'(a) = \frac{2}{\sqrt{\pi}}\exp\left(-a^2\right),$$

where $a > 0, \theta_1 > 0, \theta_2 \in \mathbb{R}$. According to Corollary 4.4,

$$C(d; u) = \int_0^\infty \exp\left(-\frac{a^2}{\theta_1}d\right) \cos\left[a(2\theta_2 u)\right] \frac{2}{\sqrt{\pi}} \exp\left(-a^2\right) da, \qquad (4.5)$$

is a valid space-time covariance model on directed Euclidean tree. Using the result from [24]: $\int_0^\infty \exp(-x^2)\cos(cx)dx = \frac{\sqrt{\pi}}{2}\exp(-\frac{c^2}{4})$, $c \in \mathbb{R}$, (4.5) can be simplified as

$$C(d; u) = \frac{1}{\sqrt{1 + \frac{d}{\theta_1}}} \exp\left(-\frac{\theta_2^2 u^2}{1 + \frac{d}{\theta_1}}\right), \qquad \theta_1 > 0, \theta_2 \in \mathbb{R}. \qquad (4.6)$$

Observe that the model given by (4.6) is a special case of (3.3) in Example 1, where $\kappa = 1/\theta_1$, $c = \theta_2^2$, $\nu = b = \beta = 1$, and $\tau = 1/2$.

Lemma 4.5. *Assume that the scale mixture space-time covariance function is isotropic, that C_S is an exponential tail-down model and C_T depends on the time lag u only. Then $C_T(0; a) \geq 0$ for $a \in \Theta_a$ is a sufficient but not necessary condition for the spatial marginal function $f_S(d) := C(d; 0)$ to be convex on $\mathbb{R}^+$.*

Proof. Let $C(d; u) := \int_{\Theta_a} \exp(-\frac{a^2}{c}d)C_T(u; a)d\mu(a)$. Suppose that the measure $\mu(\cdot)$ and the temporal covariance function C_T are smooth enough to allow interchanging the order of differentiation and integration. Then

$$f_S{}''(d) = \int_0^\infty \frac{a^4}{c^2} \exp\left(-\frac{a^2}{c}d\right) C_T(0; a)d\mu(a) \geq 0,$$

given that $C_T(0; a) \geq 0$ for all $a \in \Theta_a$. However, this condition is not necessary since we have shown in *Example 1 Revisit* that when $C_T(u; a) = \cos[a(2\theta_2 u)]$, the scale mixture covariance model is essentially a special case of Example 1 and by Proposition 3.4, the spatial marginal is always convex on $(0, \infty)$. $\square$

Apart from the sufficient condition provided in Lemma 4.5, convex cone and scale mixture models do not share unified geometric features.

Example 5. Again, let the spatial component be an exponential tail-down model with a slightly different parameterization: $C_S(d; a) = \exp(-\frac{a}{\theta_1} d)$, where $a, \theta_1 > 0$. Then consider a non-degenerate temporal covariance function $C_T(u; a) = \exp(-\frac{a}{\theta_2} u^{\theta_3})$, where $a, \theta_2 > 0$ and $\theta_3 \in (0, 2]$ [37]. Let μ be a Gamma distribution whose density function is specified as $f(a) = \frac{\theta_5^{\theta_4}}{\Gamma(\theta_4)} a^{\theta_4 - 1} e^{-\theta_5 a}$ for $a \in (0, \infty)$, with $\theta_4, \theta_5 > 0$. By Corollary 4.4,

$$C(d; u) = \left(\frac{1}{\frac{d}{\theta_1} + \frac{u^{\theta_3}}{\theta_2} + 1} \right)^{\theta_4}, \tag{4.7}$$

where θ_1, $\theta_2, \theta_4 > 0$, and $\theta_3 \in (0, 2]$ is a valid space-time covariance function on Euclidean trees. The model given by (4.7) extends the metric model in Section 3.2 since it is a function of $\frac{d}{\theta_1} + \frac{u^{\theta_3}}{\theta_2}$. The covariance function is continuous at the origin and monotonically decays to 0 as $d \to \infty$ and/or $u \to \infty$. The spatial marginal is convex on $(0, \infty)$, but this is not necessarily so for the temporal marginal.

5. Simulation study

In this section, we examine the performance of models that were introduced in previous sections via a simulation study. Streams and rivers are important to humans and many plants and animals [33], and a tree-like stream network is naturally a Euclidean tree with the geodesic distance. We chose 50 sites from the northeastern region of Clearwater River basin in central Idaho (Fig. 3) for the purpose of our analyses. Stream lines of the study region were downloaded from the National Stream Internet (NSI) dataset (`https://www.fs.fed.us/rm/boise/AWAE/projects/NationalStreamInternet/NSI_network.html`). Observations were integrated with the network object using the STARS (Version 2.0.7) toolset in ArcGIS (Version 10.7.1). A Spatial Stream Network (.ssn) object was then created, from which we extracted the topological structure of the network using the SSN package (Version 1.1.12) in R. We considered observations simulated at those 50 sites, or at subsets of those sites, at either 5 or 10 consecutive, equally-spaced points in time (for simplicity, we refer to these time points as days hereafter). The subsets of sites were chosen in two ways, so that the expansion from 25 to 50 sites conformed to either an increasing-domain or fixed-domain asymptotic regime. For the former, we chose 25 sites in the "interior" of the spatial network domain, so that their spatial extent overlapped with, but was considerably smaller than, the spatial extent of the 50 sites. For the latter, we
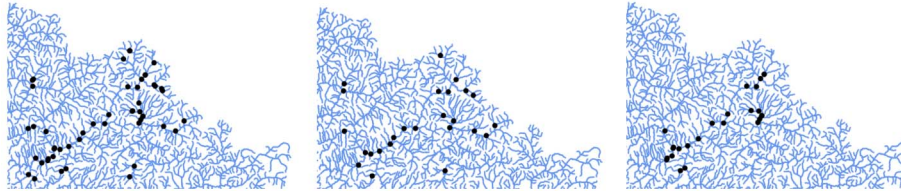
FIG 3. *Delineated stream lines of the Clearwater River basin, with (Left) 50 observation sites; (Middle) 25 observation sites (fixed-domain); (Right) 25 observation sites (increasing-domain), represented by black circles, in the northeastern region superimposed.*

chose the 25 sites systematically by sorting the sites by their upstream distances from the outlet and then alternatively deleting and keeping sites. This results in a sample similar in spatial extent but smaller in size than the original sample. Simulation results from the 50 sites $\times$ 10 days case are given in Sects. 5.1 and 5.2, and the rest can be found in Supplement C.

For each simulation, we generated 1000 realizations of multivariate normal random vectors $\mathbf{z}$ with mean $\mathbf{0}$ (assumed known), no nugget effect, and true covariance functions defined by different space-time models with details given in the following subsections. We investigated our proposed models' performance from two perspectives: (1) comparing the likelihood across different models; and (2) comparing maximum likelihood estimates (MLEs) of unknown covariance parameters to their true values. For the latter, two descriptive summary statistics, mean absolute error (MAE) and root mean squared error (RMSE), were used. We used the function `optim` in R with method "L-BFGS-B" to solve the following non-linear minimization problem to obtain the MLEs:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta}{\arg\min}\ l(\boldsymbol{\theta}) = \log |\Sigma(\boldsymbol{\theta})| + \mathbf{z}^T \Sigma^{-1}(\boldsymbol{\theta})\mathbf{z}. \tag{5.1}$$

5.1. Scenario 1: isotropic true model

In this scenario, we consider the true covariance model as given in Example 1 (i.e., (3.3)) with the following parameter specification: $c = 1$, $\nu = 1$, $\kappa = 1$, $\beta = 0.5, \tau = 1$ and $b = 0.5$. To each realization from the true model, we fit models from Example 1 with κ and ν fixed at their true values, Example 1 assuming $\beta = 0$ (i.e., a separable covariance model) and with κ and ν fixed at their true values, Example 4 (including flow-direction when the underlying true model is directionless) and Example 5. The weight function $\pi_{1,2}$ in Example 4 was computed based on the accumulated drainage area, which is information supplied with the stream lines from the NSI dataset. Estimation results from the fit of the true model are summarized in Table 2.

The MLEs in Table 2 are quite close to their true counterparts, based on median, mean, MAE and RMSE. Furthermore, and not surprisingly, Example 1 has the highest likelihood 98.9% of the time, followed by the separable model, 1.1%, of the simulations. The model given by Example 4, which is the only model

TABLE 2
Summary of maximum likelihood estimates in Example 1.

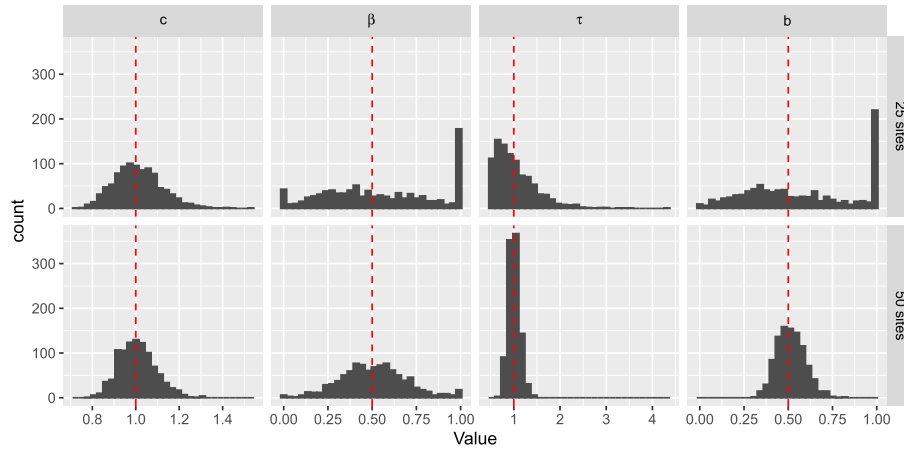|	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	2.5%	97.5%	**True**	MAE	RMSE
c	0.7763	0.9439	0.9993	1.0062	1.0599	1.3200	0.8588	1.1871	**1.0000**	0.0678	0.0856
β	0.0100	0.3857	0.5039	0.5087	0.6254	1.0000	0.1414	0.9257	**0.5000**	0.1485	0.1881
τ	0.6713	0.9166	0.9946	1.0032	1.0770	1.4808	0.7935	1.2753	**1.0000**	0.0980	0.1238
b	0.2915	0.4560	0.5065	0.5142	0.5683	0.8264	0.3682	0.6823	**0.5000**	0.0653	0.0823

FIG 4. *Histograms of estimated model parameters for (top) fixed-domain 25 sites × 10 days and (bottom) 50 sites × 10 days, in Example 1. Vertical dashed lines represent the true values.*

to incorporate flow direction, never had the highest likelihood. The means of the MLE's are consistently slightly larger than their medians, which is due to slight positive skewness in the empirical distribution of the MLE's.

Histograms of estimated parameters in Example 1 under 50 sites × 10 days and fixed-domain 25 sites × 10 days against the true values are provided in Fig. 4. There are no concerns regarding non-consistent parameter estimability (with Example 1).

5.2. Scenario 2: a model that incorporates flow direction

As a counterpart, we considered a second simulation experiment. For this one, the true model is based on Example 4, which incorporates flow direction and has the following parameter specification: $\theta_1 = 1, \theta_2 = 1, \theta_3 = 0.1$ and $\theta_4 = 0.1$. We fit the same models as were fitted in Scenario 1. While Example 4 has the highest likelihood among all 1000 realizations this time, the MLEs do not perform as well and details can be found in Supplement C. We also considered other cases of Example 4 in which the TU model was exponential or spherical; however, there was no substantial improvement in the estimates of θ_1. The heavy tail in the distribution of estimates of θ_1 (spatial component from the TU model) could be due to the fact that the data are "unbalanced," in the sense that only 201 out of the 1225 pairs of sites (about 16%) are flow-connected (see Fig. 5
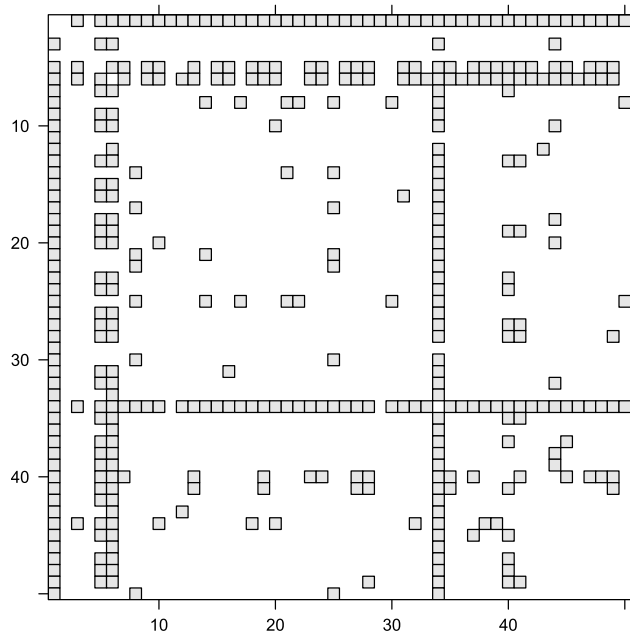
FIG 5. *Flow-connected pairs (in dark) among 50 observation sites.*

for an illustration). Therefore, the spatial component, which distinguishes between flow-connected and flow-unconnected pairs, cannot be reliably estimated. Note that for some stream network datasets the imbalance could tilt the other way, i.e., there could be too few flow-unconnected pairs. Such imbalances would appear to be a potential challenge for practitioners who implement covariance models like Example 4.

5.3. Parameter estimability

Non-consistent estimability of parameters under fixed-domain sampling is a known issue for spatial covariance models in Euclidean and spherical domains [38, 39]. To our knowledge, it has not yet been determined if it is a concern for space-time covariance models that have been proposed, regardless of the spatial domain (Euclidean, spherical, or network). The aforementioned increasing-domain and fixed-domain sampling schemes allow us to investigate this issue for the two models included in this simulation study. Table 3 and Table 4 present, for both numbers of sampling times, the interquartile ranges of the empirical distribution of maximum likelihood estimates of all parameters in both models decrease as the number of sites increases from either a compact 25-site design or the equal-spatial-extent 25-site design to the 50-site design. (We measure dispersion using interquartile ranges rather than variances or ranges because the empirical distributions of some parameter estimates are quite positively skewed.)

TABLE 3

Interquartile ranges of maximum likelihood estimates of Example 1 under different numbers of sites and time points combination.

	50 sites		25 sites (fixed-domain)		25 sites (increasing-domain)	
	5 days	10 days	5 days	10 days	5 days	10 days
c	0.1822	0.1160	0.2336	0.1450	0.2470	0.1557
β	0.3611	0.2397	0.7575	0.5299	0.5548	0.3438
τ	0.2328	0.1604	0.6162	0.5578	0.3083	0.2236
b	0.1593	0.1123	0.6826	0.5924	0.2094	0.1406

TABLE 4

Interquartile ranges of maximum likelihood estimates of Example 4 under different numbers of sites and time points combination.

	50 sites		25 sites (fixed-domain)		25 sites (increasing-domain)	
	5 days	10 days	5 days	10 days	5 days	10 days
θ_1	3.6221	2.7637	16.8395	7.6570	6.3256	5.1076
θ_2	0.0729	0.0232	0.1071	0.0325	0.1012	0.0313
θ_3	0.1007	0.0554	1.7970	1.3519	0.1400	0.0700
θ_4	0.3257	0.2877	0.4118	0.3301	0.4033	0.3179

These results indicate that there is no empirical evidence of non-consistent estimability of parameters under fixed-domain spatial sampling for the two models.

6. Discussion

This article presented a collection of tools to build valid non-separable space-time covariance models on generalized linear networks, and on an important subclass, Euclidean trees. We studied examples obtained by each constructive method and investigated the performance of maximum likelihood estimators of some covariance parameters. We have not yet provided guidance on how to choose the most suitable candidate model for an arbitrary data set, but understanding the underlying physical process [16] and matching geometric features of theoretical covariance functions to the empirical space-time covariance surface [13] would be helpful when we apply the models on real-world data. We also notice that in the simulation study, some maximum likelihood estimators have heavier tails than others. It has been argued that when prediction is the goal, model estimation is just a means to an end [26]. [38] also showed that it is the consistency of certain quantity (i.e. combination of parameters) in Matérn class, instead of individual parameters, that plays an more important role in spatial prediction. Nevertheless, the study of parameter estimability under infill and increasing domain asymptotics on a network is an open question that needs to be addressed and requires special attention. Though we emphasized the decisive role of valid covariance functions in geostatistical models, they also allow direct extension to space-time log Gaussian Cox processes on generalized networks [22, 1], which we leave for future investigation.

Acknowledgments

Supplementary Material

Supplement A
(doi: 10.1214/23-EJS2206SUPPA; .pdf). Description of resistance metric, results pertaining to conditionally negative definite functions and l_1 embedding, technical proofs and computational advantage of convex cone models.

Supplement B
(doi: 10.1214/23-EJS2206SUPPB; .pdf). Visualization of marginal functions and 3-D surfaces of the proposed models.

Supplement C
(doi: 10.1214/23-EJS2206SUPPC; .pdf). Rest of the simulation results in Section 5.

References

[1] ANDERES, E., MØLLER, J. and RASMUSSEN, J. G. (2020). Isotropic covariance functions on graphs and their edges. *Annals of Statistics* **48** 2478–2503. MR4134803

[2] ANG, Q. W., BADDELEY, A. and NAIR, G. (2012). Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scandinavian Journal of Statistics* **39** 591–617. MR3000837

[3] BADDELEY, A., JAMMALAMADAKA, A. and NAIR, G. (2014). Multitype point process analysis of spines on the dendrite network of a neuron. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **63** 673–694. MR3269407

[4] BADDELEY, A., NAIR, G., RAKSHIT, S. and MCSWIGGAN, G. (2017). "Stationary" point processes are uncommon on linear networks. *Stat* **6** 68–78. MR3613182

[5] BERG, C. (2008). Stieltjes-Pick-Bernstein-Schoenberg and their connection to complete monotonicity. In *Positive Definite Functions: From Schoenberg to Space-Time Challenges* (S. Mateu and E. Porcu, eds.) Department of Mathematics, University Jaume I, Castellón de la Plana, Spain.

[6] BOCHNER, S. (1955). *Harmonic Analysis and the Theory of Probability.* University of California Press, Berkeley and Los Angeles. MR0072370

[7] CAMBANIS, S., KEENER, R. and SIMONS, G. (1983). On α-symmetric multivariate distributions. *Journal of Multivariate Analysis* **13** 213–233. MR0705548

[8] CHILÈS, J.-P. and DELFINER, P. (1999). *Geostatistics — Modeling Spatial Uncertainty.* Wiley, New York. MR1679557

[9] CRESSIE, N., FREY, J., HARCH, B. and SMITH, M. (2006). Spatial prediction on a river network. *Journal of Agricultural, Biological, and Environmental Statistics* **11** 127–150.

[10] CRESSIE, N. and HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* **94** 1330–1339. MR1731494

[11] DE IACO, S. (2010). Space–time correlation analysis: a comparative study. *Journal of Applied Statistics* **37** 1027–1041. MR2757111

[12] DE IACO, S., MYERS, D. E. and POSA, D. (2002). Nonseparable space-time covariance models: some parametric families. *Mathematical Geology* **34** 23–42. MR1954124

[13] DE IACO, S., POSA, D. and MYERS, D. (2013). Characteristics of some classes of space–time covariance functions. *Journal of Statistical Planning and Inference* **143** 2002–2015. MR3095089

[14] FUENTES, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika* **89** 197–210. MR1888368

[15] GNEITING, T. (1998). On α-symmetric multivariate characteristic functions. *Journal of Multivariate Analysis* **64** 131–147. MR1621926

[16] GNEITING, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association* **97** 590–600. MR1941475

[17] GRIFFITH, D. A. and HEUVELINK, G. (2012). Deriving space-time variograms from space-time autoregressive (STAR) model specifications. In *Advances in Spatial Data Handling and GIS* 3–12. Springer.

[18] GRÄLER, B., PEBESMA, E. and HEUVELINK, G. (2016). Spatio-temporal interpolation using gstat. *The R Journal* **8** 204–218. https://doi.org/10.32614/RJ-2016-014

[19] JAMMALAMADAKA, A., BANERJEE, S., MANJUNATH, B. S. and KOSIK, K. S. (2013). Statistical analysis of dendritic spine distributions in rat hippocampal cultures. *BMC Bioinformatics* **14** 287.

[20] MENEGATTO, V., OLIVEIRA, C. and PORCU, E. (2020). Gneiting class, semi-metric spaces and isometric embeddings. *Constructive Mathematical Analysis* **3** 85–95. MR4188612

[21] MILLER, K. S. and SAMKO, S. G. (2001). Completely monotonic functions. *Integral Transforms and Special Functions* **12** 389–402. MR1872377

[22] MØLLER, J., SYVERSVEEN, A. R. and WAAGEPETERSEN, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25** 451–482. MR1650019

[23] NEWMAN, M. (2010). *Networks: An Introduction.* Oxford University Press, New York. MR2676073

[24] NG, E. W. and GELLER, M. (1969). A table of integrals of the error functions. *Journal of Research of the National Bureau of Standards B* **73** 1–20. MR0239718

[25] O'DONNELL, D., RUSHWORTH, A., BOWMAN, A. W., SCOTT, E. M. and HALLARD, M. (2014). Flexible regression models over river networks. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **63** 47–63.

MR3148268

[26] Porcu, E., Furrer, R. and Nychka, D. (2020). 30 Years of space–time covariance functions. *Wiley Interdisciplinary Reviews: Computational Statistics* e1512. MR4218945

[27] Porcu, E., White, P. A. and Genton, M. G. (2022). Nonseparable space-time stationary covariance functions on networks cross time. arXiv: 2208.03359. MR4214159

[28] Rodrigues, A. and Diggle, P. J. (2010). A class of convolution-based models for spatio-temporal processes with non-separable covariance structure. *Scandinavian Journal of Statistics* **37** 553–567. MR2779636

[29] Schoenberg, I. J. (1938). Metric spaces and completely monotone functions. *Annals of Mathematics* **39** 811–841. MR1503439

[30] Schur, I. (1911). Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *Journal für die Reine und Angewandte Mathematik* **1911** 1–28. MR1580823

[31] Stein, M. L. (2005). Space–time covariance functions. *Journal of the American Statistical Association* **100** 310–321. MR2156840

[32] Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* **46** 234–240.

[33] Ver Hoef, J. M. and Peterson, E. E. (2010). A moving average approach for spatial statistical models of stream networks. *Journal of the American Statistical Association* **105** 6–18. MR2757185

[34] Ver Hoef, J. M., Peterson, E. E. and Theobald, D. (2006). Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics* **13** 449–464. MR2297373

[35] Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions, Vol. I: Basic Results*. Springer-Verlag, New York. MR0893393

[36] Zastavnyi, V. P. (2000). On positive definiteness of some functions. *Journal of Multivariate Analysis* **73** 55–81. MR1766121

[37] Zastavnyi, V. P. and Porcu, E. (2011). Characterization theorems for the Gneiting class of space–time covariances. *Bernoulli* **17** 456–465. MR2797999

[38] Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association* **99** 250–261. MR2054303

[39] Zhang, H. and Zimmerman, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936. https://doi.org/10.1093/biomet/92.4.921. MR2234195