University of Trento

Department of Mathematics

Doctoral Programme in Mathematics - Cycle XXXVIII

$\sim \cdot \sim$

Thesis for the Degree of Philosophiae Doctor

Gaussian Stochastic Processes on Metric Graphs

Supervisors
Prof. Claudio Agostinelli
Prof. Emilio Porcu

Candidate
Tobia Filosi
241164

Final examination date: 16th December 2025

To my family.

Contents

Acknowledgments

First, I owe this thesis to my family: the support and care that you have always shown me have been of great encouragement. In addition, I would like to express my deepest gratitude to those who have been contributing to my academic path: the professors, colleagues and staff of the Department of Mathematics in Trento, the Minerva Research Foundation, and, last but not least, both my supervisors: Claudio, thank you for your gentle presence and constant backing; Emilio, thank you for your sharpness, passion and dedicated spurs.

Abstract

Linear networks have been recently generalised by metric graphs: these are flexible graph topologies where each edge is in bijection with a segment of the real line, allowing points on edges to have a *geometric* meaning. Anderes et al. (2020) extended the classical notion of effective resistance distance to these new topologies and provided an ingenious method to define valid isotropic covariance functions (therefore, stochastic processes) on them. In this thesis, we collect some recent developments of this field: first, characterise the role this new framework can play in climate science, together with two other types of graphs. Then, we extend metric graphs to time-evolving dynamics, where both nodes and edges may appear or disappear in subsequent time instants, and where edge may also change length. We considered (for static graphs) a class of multivariate processes, based on a notion of multi-dimensional distance, whose marginal distribution coincide with the one defined in Anderes et al. (2020). Lastly, we explored efficient algorithms to simulate univariate isotropic Gaussian random fields on metric graphs. For each of these analyses, we mainly delved into theoretical properties of the defined objects, yet we motivated the statistical impact that they could have.

Mathematical Notation

The following list summarises all the main symbols used in this thesis. In addition:

- bold lowercase letters denote (deterministic) vectors (*e.g.* $\boldsymbol{c}$),
- bold uppercase letters denote (deterministic) matrices (*e.g.* $\mathbf{A}$),
- bold uppercase italic letters denote random vectors (*e.g.* $\boldsymbol{Z}$), an exception to this rule appears in Chapter 3, where $\boldsymbol{G}$ denotes a time-evolving graph,
- non-bold letters represent scalar (real-valued) quantities or non-numerical objects (such as sets, graphs, vertices, edges, points).

Graph symbols

G	Graph.
V	Set of vertices.
v	Vertex.
E	Set of edges.
e	Edge.
$\ell(e)$	Length of e.
$w(e)$	Weight of e.
$\sim$	Adjacency between vertices.
$\mathbf{W}$	Weight matrix.
$\mathbf{L}$	Laplacian matrix.
$\underline{u}$, $\overline{u}$	Endpoints of the edge containing u.
$\delta(u)$	Relative distance of u from $\underline{u}$.
$\boldsymbol{\delta}(u)$	Vector of zeroes, but two elements being $\delta(u)$ and $1-\delta(u)$, as defined in Equation (3.8).

Mathematical Notation

$\boldsymbol{G}$	Time-evolving graph.
$\widetilde{\boldsymbol{G}}$	Equivalent simple time-evolving graph.
E_S	For time-evolving graphs, set of spatial edges.
E_T	For time-evolving graphs, set of temporal edges.
$\widetilde{E}$	For time-evolving graphs, the union of E_S and E_T.
s	For time-evolving graphs, vertices labelling function.
S	For time-evolving graphs, label set.
$\mathrm{lf}(e)$	For time-evolving graphs, life of e.
$\mathrm{ls}(e)$	For time-evolving graphs, lifespan of e.

Matrix and vector symbols

diag	Diagonal operator.
$\otimes$	Kronecker product.
$\mathbf{X}^\top$	The transpose of $\mathbf{X}$.
$\mathbf{X}^-$	The Moore-Penrose inverse of $\mathbf{X}$.
$\mathbf{I}_n$	The $n \times n$ identitiy matrix.
$\mathbf{J}_n$	The $n \times n$ matrix defined as $\mathrm{diag}\left[0, 0, \ldots, 0, 1\right]^\top$.
$\mathbf{1}_n$	The n-dimensional column vector of all ones.
$\mathbf{1}_{n\times n}$	The $n \times n$ matrix of all ones.
$\mathcal{M}_p\left(\mathbf{A}, \mathbf{B}\right)$, $\mathcal{M}_p\begin{pmatrix}\mathbf{A}\\ \mathbf{B}\end{pmatrix}$	For two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, $np \times np$-dimensional matrix $\mathbf{I}_p \otimes \mathbf{A} + (\mathbf{1}_{p\times p} - \mathbf{I}_p) \otimes \mathbf{B}$.
$\mathbf{A}/\mathbf{B}$	Schur complement.

Other symbols

e	Napier's constant.
$\mathbb{N}$	The set of natural numbers.
$\mathbb{N}^+$	The set of positive integers.
$\mathbb{R}$	The set of real numbers.

$\mathbb{R}^+$ The set of strictly positive real numbers.

$\mathbb{R}_0^+$ The set of non-negative real numbers.

$\mathbb{1}$ Indicator function.

d Distance.

$\mathbf{D}$ Matrix-valued distance.

$\otimes$ In Chapter 3, Hilbert space tensor product.

$\mathcal{H}$ Hilbert space.

$\langle \cdot, \cdot \rangle$ Inner product.

$\|\cdot\|$ Norm.

$\overset{g}{\hookrightarrow}$ g-embedding in Hilbert space.

ψ Real function, usually completely monotonic.

$\mathcal{O}$ In Chapter 5, big-O notation for computational complexity.

T Set of times.

Probability and Statistics symbols

$\mathbb{E}(Z)$ Expected value of Z.

$\mathbb{V}\mathrm{ar}(Z)$ Variance of Z.

$\mathbb{C}\mathrm{ov}(Z_1, Z_2)$ Covariance between Z_1 and Z_2.

k_Z Covariance function (or kernel) of the random field Z.

γ_Z Variogram of the random field Z.

$\mathbf{\Gamma}_{\boldsymbol{Z}}$ Pseudo-variogram of $\boldsymbol{Z}$.

W The Wiener process.

$\sim$ Distribution of a random object.

$\mathcal{N}(\mu, \sigma^2)$ Univariate normal distribution with mean μ and variance σ^2.

$\mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ p-variate normal distribution with mean $\boldsymbol{\mu}$ and variance $\mathbf{\Sigma}$.

Chapter 1

Introduction

Life is full of uncertainties: from people we bump into walking down the street to tomorrow's weather, from the location where a seed is dropped by a bird to the small molecule we inhale that will affect our health, and this list could continue for dozens of pages. Even though we may have ideas about these events and even, sometimes, predict some of them with sufficient accuracy, never will we be able to outdo our world to determinism. This is why Statistics plays a pivotal role in our lives. Statistics is the "Science of Uncertainty": it aims to handle uncertainty, to *formalise* and *measure* it, and, when possible, to "*predict*" future or unobserved events. However, although Statistics relies on grounded mathematical results, such results rest on precise mathematical assumptions, which are often unrealistic (should we say uncertain?) when applied on *real* phenomena. Nevertheless, Statistics is the best tool we have to deal with uncertainty, at least from a rigorous perspective.

Spatial Processes

In this thesis, we mainly deal with *Spatial Processes*, which are precise mathematical objects that describe random phenomena showing spatial dependence. Next, we present the definition of random field and three major specialisations that are often used in Spatial Statistics, to give the context from which we branched in our research. Here, we follow Cressie and Wikle (2011, Chapter 4), to which the reader is referred for a more detailed introduction.

Definition 1.1 (Random Field or Stochastic Process)**.** Let X a non-empty set, Q a topological space and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. A Q-valued *random field* (or *stochastic process*) Z over X is a collection of random variables

$$\{Z_x \,:\, x \in X\},$$

where each $Z_x : (\Omega, \mathcal{F}, \mathbb{P}) \to Q$ is a random variable, namely a measurable function. Equivalently, Z can be seen as a function

$$Z : X \times \Omega \to Q,$$

where, for each $x \in X$, $\omega \mapsto Z(x, \omega)$ is a random variable.

The concept of random field will be recalled through the thesis and an intuitive example can be found, for instance, in Subsection 2.3.2.

Geostatistical Processes

Although the prefix *geo-* was born to indicate Statistics pertaining to the Earth, in the last decades, its methods saw applications in a wider range of fields, from agriculture to zoology. Geostatistics is applicable under mild hypotheses, regarding the mean and covariance functions of the phenomena of interest. Let us state the definition, following Chilès and Delfiner (2012, p. 12).

Definition 1.2 (Geostatistical Process)**.** A Q-valued random field Z over X is called *geostatistical process* if there exist $n, p \in \mathbb{N}^+$ such that:

- $X \subseteq \mathbb{R}^n$ is measurable and has strictly positive Lebesgue measure,
- $Q \subseteq \mathbb{R}^p$.

If $p = 1$, Z is called *univariate* process, whilst if $p \geq 2$, Z is called *multivariate*.

That stated, the intuitive idea underlying geostatistical processes is that, at each spatial location x, $Z(x)$ describes a random phenomenon that may depend on other values of Z, usually for points "close" to x. A natural application may be the temperature in a given geographical region: clearly, if two points of this region are "close", their temperature will *likely* be similar, while "distant" points may show less related temperatures. To give a real-world example of this pattern, Figure 1.1 shows global precipitation anomalies in May 2025 on a 2.5°×2.5° resolution: notice how, albeit casual, close points tend to show similar precipitation anomalies.

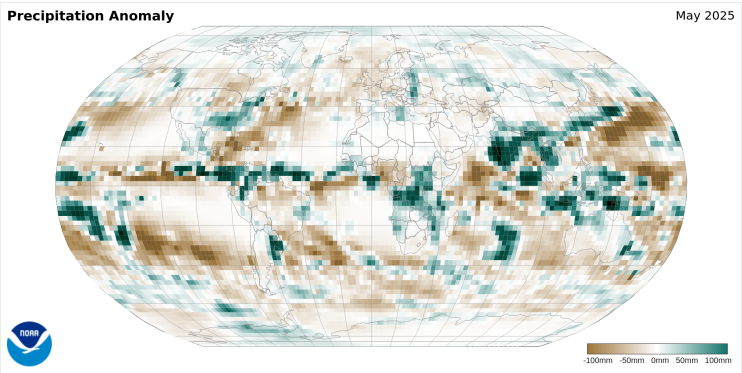


Figure 1.1: An example realisation of a geostatistical process: global precipitation anomalies in May 2025 on a 2.5°×2.5° resolution. Figure taken from NOAA National Centers for Environmental information, Climate at a Glance: Global Mapping.

Usually, the notions of "close" and "distant" rely on some *distance* defined between any couple of points in X, such as the Euclidean distance. However, other distances may be more useful in some specific contexts: for instance, in Earth science, distance is usually the *geodesic* or *great-circle* distance, that is the length of the shortest path on Earth surface. In this thesis, we will often use the effective-resistance distance on graphs with Euclidean edges, whose definition is given below.

Lattice Processes

Lattice processes are a particular case of random fields. Indeed, they require X to be a discrete set, finite or countable, often with some regularity structure. For instance, think about a uniform grid covering the whole Earth surface, where only the intersections between meridians and parallels belong to it. Another example may be the provinces in a given region, where we aggregate data on each province. Think about the life expectancy at birth in African countries (see Figure 1.2). Here, the concept of distance is usually defined via adjacency of different points or regions. In the examples given above, two points on the Earth grid may be adjacent if they have the same parallel and close meridians or vice versa. Similarly, two African Countries may be considered adjacent if their share boundaries.

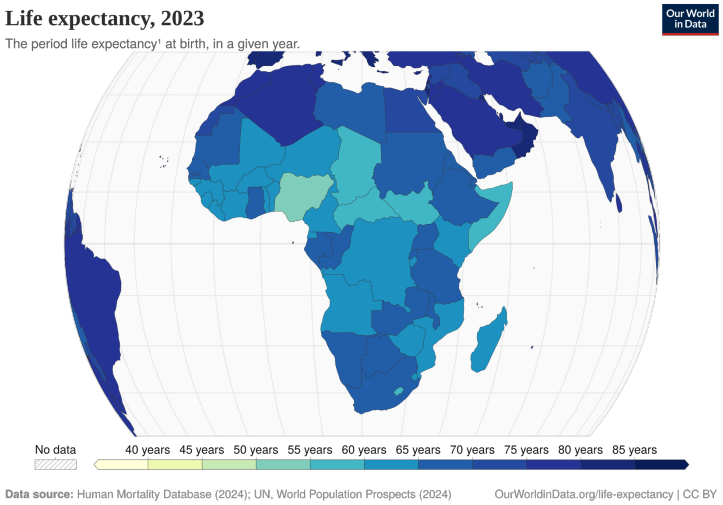


Figure 1.2: An example of lattice process: life expectancy at birth across African countries in 2023. Figure taken from Dattani et al. (2023).

Spatial Point Processes and Random Sets

While the aforementioned geostatistical and lattice processes describe random objects defined on each point of a given spatial domain X, spatial point processes and random sets describe the *locations* (and, sometimes, other properties) of random

events. Spatial point processes assume random events to be points in the spatial domain X, whilst random sets model events as regions $R \subseteq X$. To have an intuitive idea, think about the positions of trees in a forest (see Figure 1.3) or the locations of car accidents on a road network during a given time period: these may be modelled as spatial point processes. Similarly, the area interested by a wildfire or the region of

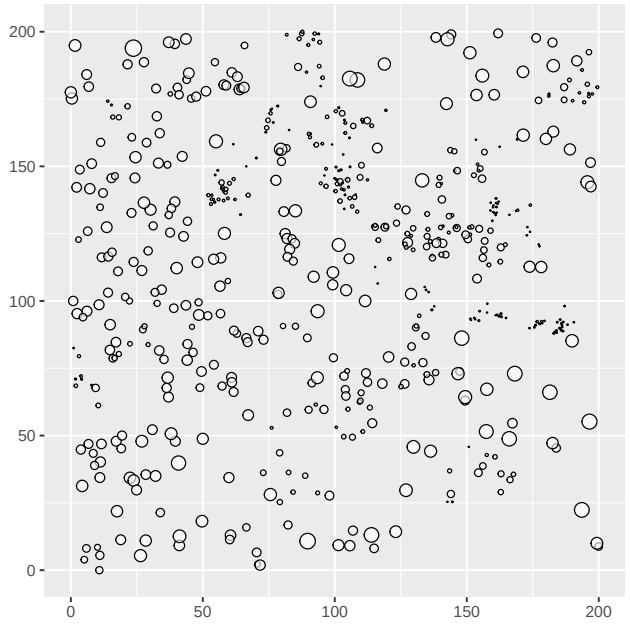


Figure 1.3: Locations and diameters of 584 Longleaf pine (*Pinus palustris*) trees in a 200 x 200 metre region in southern Georgia (USA) (Platt et al., 1988).

cancer cells in a human internal organ may be modelled as random sets. The formal definition of these spatial processes requires some mathematical preliminaries and goes beyond the scope of this thesis: we are happy with these intuitive ideas.

Geostatistical Processes and their Properties

Here, we give some preliminaries about geostatistical processes. As we will consider only *Gaussian* stochastic processes, a brief introduction follows. However, this concept will be often recalled in the remaining chapters of the thesis.

Gaussian Processes

Here, we assume that the reader is already familiar with the notion of (univariate) normal distribution. If this was not the case, we refer to virtually any Probability or Statistics textbook.

Definition 1.3 (Gaussian Random Field)**.** A real-valued random field Z defined on a set X is *Gaussian* if, for any $n \in \mathbb{N}^+$ and for any set of points $x_1, \ldots, x_n \in X$, the n-dimensional vector $(Z(x_1), \ldots, Z(x_n))$ has a multivariate normal distribution. Equivalently, Z is Gaussian if, for all $x_1, \ldots, x_n \in X$ and $c_1, \ldots, c_n \in \mathbb{R}$, the variable

$$Y := \sum_{i=1}^{n} c_i Z(x_i)$$

has a (univariate) normal distribution.

Gaussian processes enjoy several useful properties. For instance, the class of Gaussian processes is closed under sum, a Gaussian process times a deterministic function is still Gaussian, they are the limit distribution of the sum (or mean) of independent arbitrary processes. However, one of the properties that make Gaussian processes so popular in Statistics is that they are completely characterised by their first two moments, videlicet the mean and covariance functions. Although in this thesis we will always consider *zero-mean* Gaussian processes, *i.e.* processes whose mean at every point in the domain is null, it is easily possible to set a custom mean function, say $\mu : X \to \mathbb{R}$, and consider the new process $Z'(x) := Z(x) + \mu(x)$. The second moment is of much more interest, as is it that determines the *shape* of the process.

Covariance Functions

For the sake of completeness, we provide the definitions of mean, covariance and correlation for random variables. However, these concepts are recalled in the following chapters.

Definition 1.4 (Expected Value)**.** Let $Z : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ a random variable. The *mean* (or *expected value*) $\mathbb{E}(Z)$ of Z is the real number (if it exists) defined as:

$$\mathbb{E}(Z) := \int_{\Omega} Z(\omega) \, \mathrm{d}\mathbb{P}(\omega).$$

Definition 1.5 (Covariance and Variance)**.** Let Z_1, Z_2 two real-valued random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the *covariance* between Z_1 and Z_2 is the real number (if it exists) defined as:

$$\mathbb{C}\mathrm{ov}(Z_1, Z_2) := \mathbb{E}\left((Z_1 - \mathbb{E}(Z_1))(Z_2 - \mathbb{E}(Z_2))\right).$$

The *variance* $\mathbb{V}\mathrm{ar}(Z)$ of a random variable Z is the covariance of Z with itself:

$$\mathbb{V}\mathrm{ar}(Z) := \mathbb{C}\mathrm{ov}(Z, Z) = \mathbb{E}\left((Z_1 - \mathbb{E}(Z_1))^2\right).$$

The covariance describes linear concordance or discordance. That is, if the two variables tend to increase or decrease together (concordance), then they will show

positive covariances, whilst if one will usually decrease when the other grows and vice versa, they will show a negative covariance (discordance). Finally, if two random variables are completely independent, that is: the behaviour of one gives absolutely no information about the other, then they will have covariance identically equal to zero. The converse is true if the random field is Gaussian. Think about the mean temperature during one day in two close geographical points: then it is reasonable to assume that, but for some specific local meteorological phenomenon, an increase in the temperature in the first point corresponds to an increase in the second point as well. Conversely, if we think at precipitation and temperature in the same point, it is likely that an increase in one corresponds to a decrease for the other. Correlation (defined next) is a linear transformation of the covariance, so that it assumes only values between -1 and 1.

Definition 1.6 (Correlation)**.** Let Z_1, Z_2 two real-valued random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the *correlation* between Z_1 and Z_2 is the real number (if it exists) defined as:

$$\mathbb{C}\mathrm{or}\left(Z_1, Z_2\right) := \frac{\mathbb{C}\mathrm{ov}\left(Z_1, Z_2\right)}{\sqrt{\mathbb{V}\mathrm{ar}\, Z_1 \cdot \mathbb{V}\mathrm{ar}\, Z_2}}.$$

As introduced above, covariance plays a pivotal role in the description of random phenomena: it *quantifies* the linear relations between random variables: if two random variables show a positive covariance, they will be more likely to be both high or both low. Conversely, for negative covariances, random variables will likely behave oppositely: one is high, the other is low. Covariance functions are extensions of covariance matrices: whilst the latter model the covariance structure of (finite-dimensional) random vectors, the former are defined on possibly uncountable sets. As such, they are fundamental in the *estimation* of processes and *error quantification*. The method of estimating the unobserved value given related observed ones in geostatistics is called *kriging*.

A formal definition of covariance function follows.

Definition 1.7. Let $Z : X \to \mathbb{R}$ a square-integrable process (namely, for all $x \in X$, $\mathbb{E}\left(Z^2(x)\right) < +\infty$). The *covariance function* (or *kernel*) $k_Z : X \times X \to \mathbb{R}$ of Z is defined as:

$$k_Z(x_1, x_2) := \mathbb{C}\mathrm{ov}\left(Z(x_1), Z(x_2)\right),$$

where $x_1, x_2 \in X$. A covariance function that satisfies $\tilde{k}_Z(x, x) = 1$ for all $x \in X$ is called *correlation function*: it measures the *correlation* of the process Z at two given points. If k_Z is a covariance function, then

$$\tilde{k}_Z(x_1, x_2) := \frac{k_Z(x_1, x_2)}{\sqrt{k_Z(x_1, x_1) k_Z(x_2, x_2)}}$$

is the relative correlation function.

Conversely to mean functions (which can be anything), a covariance function k satisfies two fundamental properties.

- Symmetry: for all $x_1, x_2 \in X$, $k(x_1, x_2) = k(x_2, x_1)$.
- Positive definiteness: for all $n \in \mathbb{N}^+$ and for all $c_1, \dots, c_n \in \mathbb{R}$, it holds:

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0.$$

Perhaps surprisingly, the converse is also true: if a function k is symmetric and positive definite, then there exist a Gaussian process whose kernel is k.

Positive definiteness turns usually out to be challenging to prove for a generic function. However, there are some properties of closure for positive definite functions that may help. For example, if k_1, k_2 are positive definite functions on a domain X, then both $k_1 + k_2$ and $k_1 \cdot k_2$ are positive definite. Another strategy often used to show the positive-definiteness of a function is to build a stochastic process that have such a covariance.

Some of covariance functions defined on metric spaces depend only on the *distance* between the points: in formulae

$$k_Z(x_1, x_2) = C_k(d(x_1, x_2)),$$

for $x_1, x_2 \in X$. These covariance functions are called *isotropic*. Thinking about the Euclidean space $\mathbb{R}^n$ with the Euclidean distance d, this means that the covariance structure is the same in all the directions we move from a given point: the only thing that matters is *how much* we move away.

Next, to introduce the reader to some properties that covariance functions have and how these influence sample paths, we provide some classical examples of (isotropic) covariance functions defined on Euclidean spaces and their relative simulations.

Definitely the simplest covariance function is the indicator function:

$$k_N(x_1, x_2) := \sigma^2 \mathbb{1}(x_1 = x_2) = \begin{cases} \sigma^2 & \text{if } d(x_1, x_2) = 0 \\ 0 & \text{if } d(x_1, x_2) > 0. \end{cases}$$

In the framework of Gaussian processes, a process Z with covariance function k_N is often called a *white noise*. That is, such a covariance function implies a uniform variance $\sigma^2 > 0$ and the independence of the process Z at different points. Sometimes, this covariance function is used as a building block of more complex structures, where it is desired to model, for instance, some (uniform) measurement error. In such a case, this component of the resulting covariance function is called *nugget effect*.

Another simple isotropic covariance function is the *exponential* one:

$$k_E(x_1, x_2) := \sigma^2 \exp\left(-\alpha d(x_1, x_2)\right),$$

with $\sigma^2 > 0$ and $\alpha > 0$. In this case, the correlation tends to 1 for $x_1 \to x_2$, meaning that two very close points will have a very close values, ensuring the continuity of the sample paths.

A little modification of the exponential covariance function is the *Gaussian* covariance:

$$k_G(x_1, x_2) := \sigma^2 \exp\left(-\alpha^2 d^2(x_1, x_2)\right),$$

still with $\sigma^2 > 0$ and $\alpha > 0$. This covariance ensures not only the continuity of the sample paths, but also their differentiability (since the first derivative of k_G for $x_1 \to x_2$ is null). Other notable covariance functions are the *Matérn* (which generalises both the exponential and the Gaussian covariance functions), the *wave* (which also produces negative correlations), the *spherical* and many others. Figure 1.4 plots different types of isotropic covariance functions on $\mathbb{R}^1, \mathbb{R}^2$ and the relative sample paths.

Covariance functions on Euclidean spaces have been deeply explored in literature. Another well-studied domain for the definition of covariance functions are spheres and high-dimensional spheres $\mathbb{S}^n$. Furthermore, there is usually a lot of interest in generalising purely-spatial covariances to space-time: given a spatial domain X and a temporal domain $T \subseteq \mathbb{R}$, one seeks covariance functions defined on $X \times T$, which allow to model spatio-temporal processes.

In this thesis, we focus on defining and studying valid covariance functions on metric graphs, as explained below.

Variograms

Variograms provide a slightly more general point of view of what covariance functions describe. Instead of measuring the covariance between two given points, variograms quantify the variance of the difference. In the following definition, we require for simplicity the process Z to be square-integrable. Actually, this condition is not necessary, but it simplifies both the exposition and the link between variograms and covariance functions. In addition, it is sufficient to understand the research presented in this thesis, as we will always consider square-integrable processes.

Definition 1.8 (Variogram)**.** Let $Z : X \to \mathbb{R}$ a square-integrable process. Its *variogram* is defined as:

$$\gamma_Z(x_1, x_2) := \mathbb{V}\mathrm{ar}\left(Z(x_1) - Z(x_2)\right),$$

where $x_1, x_2 \in X$. The quantity $\gamma_Z/2$ is called *semi-variogram*.

Clearly, the variogram can be expressed as a function of the covariance function:

$$\begin{aligned}\gamma_Z(x_1, x_2) &= \mathbb{V}\mathrm{ar}\, Z(x_1) + \mathbb{V}\mathrm{ar}\, Z(x_2) - 2\mathbb{C}\mathrm{ov}\left(Z(x_1), Z(x_2)\right) \\ &= k_Z(x_1, x_1) + k_Z(x_2, x_2) - 2k_Z(x_1, x_2).\end{aligned}$$

The converse is not true: there may be more covariance functions that share the same variogram. For instance, if k_Z is a valid covariance, then $k_Z + c$ is valid as well

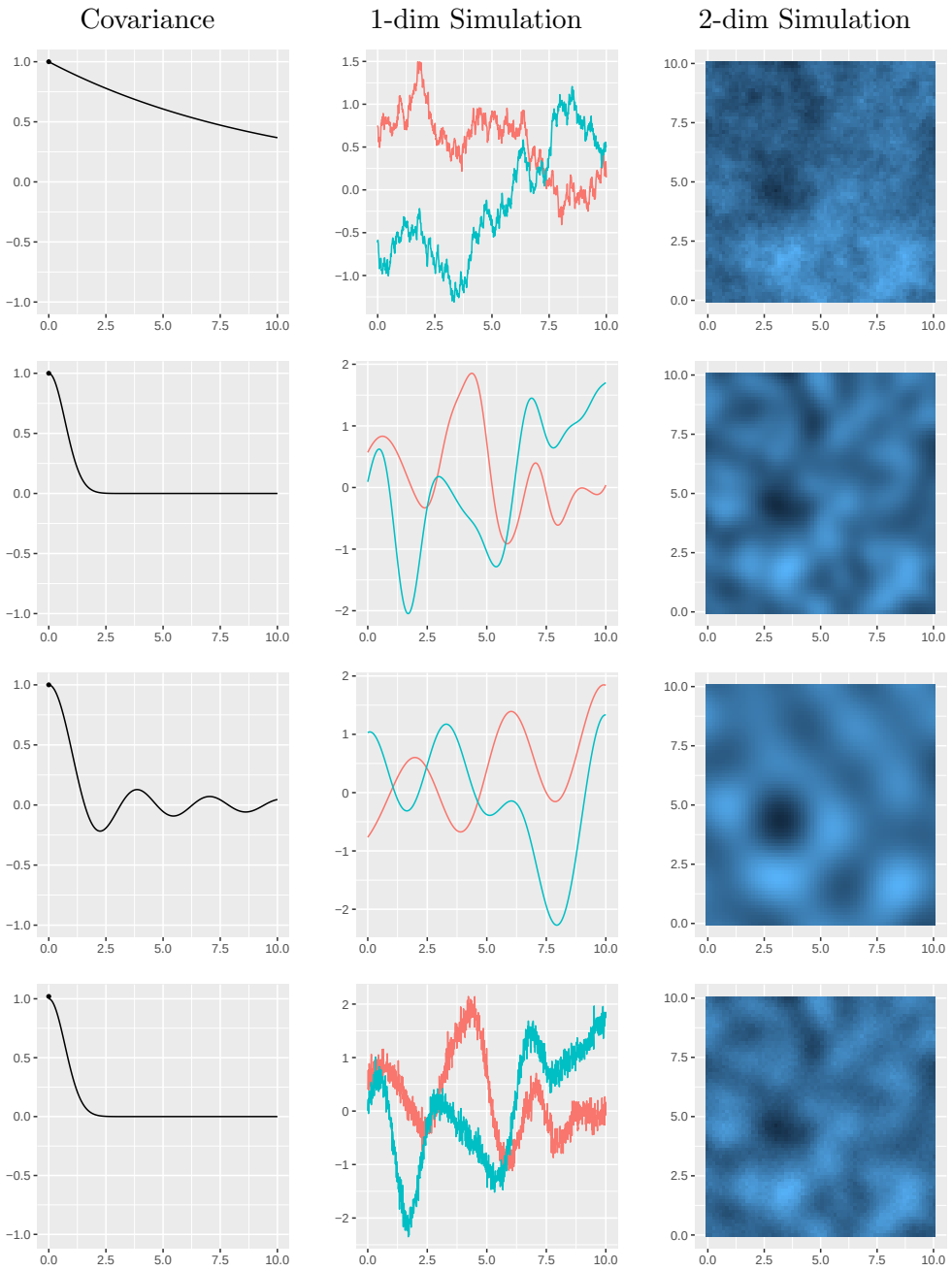


Figure 1.4: Different isotropic covariance functions and relative sample processes. First column shows covariance as a function of (Euclidean) distance, second column shows two independent simulations on $[0, 10]$, whilst third column shows a single realisation on the square $[0, 10]^2$. The four rows show, respectively: exponential (k_E), Gaussian (k_G), wave (k_W), and Gaussian plus nugget effect ($k_G + k_N$) covariances.

as long as $c \geq 0$ (since it is the sum of positive definite functions). However, it is immediate to see that both functions share the same variogram.

Variograms enjoy some properties: they are symmetric, non-negative and conditionally negative definite, that is: for each $n \in \mathbb{N}^+$ and for each $c_1, \ldots, c_n \in \mathbb{R}$ such that $\sum_{i=1}^n c_i = 0$, it holds:

$$\sum_{i,j=1}^{n} c_i c_j \gamma_Z(x_i, x_j) \leq 0.$$

Similarly to covariance functions, this is a characterisation: if a function γ_Z satisfies the three properties stated above, there exists a process Z whose variogram is γ_Z.

Also variograms can be *isotropic*: this happens whenever they depend only on the distance of the points, that is $\gamma_Z(x_1, x_2) = C(d(x_1, x_2))$. It is easy to show that a process with isotropic covariance function also has an isotropic variogram:

$$\begin{aligned}\gamma_Z(x_1, x_2) &= k_Z(x_1, x_1) + k_Z(x_2, x_2) - 2k_Z(x_1, x_2) \\ &= C_k(d(x_1, x_1)) + C_k(d(x_2, x_2)) - 2C_k(d(x_1, x_2)) \\ &= 2C_k(0) - 2C_k(d(x_1, x_2)) = 2C_\gamma\left(d(x_1, x_2)\right),\end{aligned}$$

where $C_\gamma(d) := C_k(0) - C_k(d)$. Again, the converse is not true: think about the standard Wiener process W_t defined on $\mathbb{R}_0^+$. Its covariance function is $k_W(t_1, t_2) = \min(t_1, t_2)$, while its variogram is given by

$$\begin{aligned}\gamma_W(t_1, t_2) &= k_W(t_1, t_1) + k_W(t_2, t_2) - 2k_W(t_1, t_2) \\ &= t_1 + t_2 - 2\min(t_1, t_2) = |t_1 - t_2|.\end{aligned}$$

Isotropic semi-variograms are usually graphically represented, as their plot provides some standard information.

- **Nugget.** It is the limit the semi-variogram has for $d \to 0^+$: represents the nugget effect introduced above: the higher it is, the higher is the influence of a white noise error on the process. If the nugget is null, then the process has continuous sample paths.

- **Range.** It is the distance at which the semi-variogram reaches the variance of the process: it provides information about *how far* there is influence. Low ranges mean that close spatial points will show low correlation, while high ranges mean the opposite. For several models, the variogram never reaches its highest value (sill): in such cases we usually consider the *practical range*, that is the distance d at which the semi-variogram reaches, say, 95% of the total variance.

- **Sill.** It is the limit of the semi-variogram for the distance going to $+\infty$: it coincides with the total variance of the process.

Figure 1.5 shows the semi-variogram of an isotropic process with exponential covariance function, and a nugget effect accounting for $\nu = 20\%$ of the total variance ($\sigma^2 = 1$). The full covariance of the process is given by

$$k(x_1, x_2) = \sigma^2 \left(\nu \mathbb{1}(x_1 = x_2) + (1 - \nu) e^{-d(x_1, x_2)} \right).$$

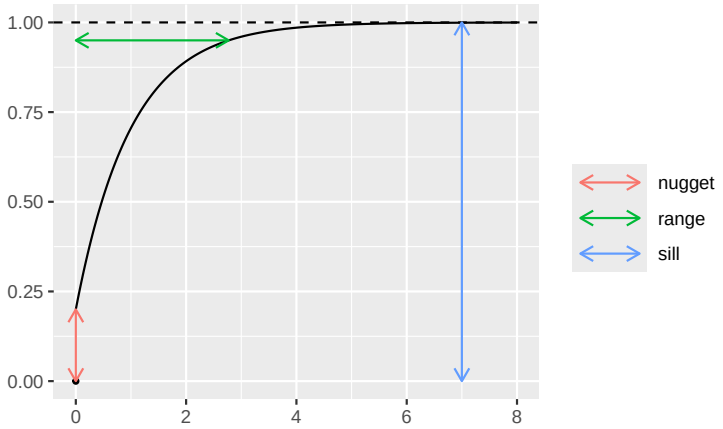


Figure 1.5: Semi-variogram as a function of distance of an isotropic process with exponential covariance function and a nugget effect accounting for 20% of the total variance ($\sigma^2 = 1$).

Metric Graphs

Standard Graph Theory and Notation

Before delving into the ideas underlying metric graphs and their definition, it is in order to provide a brief introduction of standard graph theory and notation.

Definition 1.9 (Graph)**.** Let V a finite set (its elements are called *vertices* or *nodes*) and let $E \subseteq V \times V$ a set of connections (called *edges*). Then the pair $G := (V, E)$ is called *graph* or *network*.

The *rationale* behind Definition 1.9 is simply to consider relations between elements of the vertices set V. To make an example, think about mathematical proofs: usually, a proof relies on other (already shown or taken for granted) results. It could be possible to model this by choosing V as the set of all shown mathematical results in a given field and E as the set of couples (v_1, v_2), where result v_1 has been directly used in the proof of v_2. In this case, (V, E) is a *directed* graph, as the edge (v_1, v_2) also indicates a *direction*. Let us formalise this concept.

Definition 1.10 (Undirected Graph, Directed Graph)**.** A graph $G = (V, E)$ is called *undirected* if, whenever $(v_1, v_2) \in E$, then necessarily $(v_2, v_1) \in E$. Otherwise G is called *directed*.

A simple example of undirected graph is the friendship graph: V is the set of people and two people are connected by an edge if they are friends. It is reasonable to assume that, if v_1 is a friend of v_2, then v_2 is a friend of v_1 as well.

For an undirected graph G, two nodes $v_1, v_2 \in V$ are *adjacent* if $(v_1, v_2) \in E$. We write $v_1 \sim v_2$ if v_1 and v_2 are adjacent, $v_1 \not\sim v_2$ if they are not. We said that a graph G is *simple* if it is *undirected* and if it has no self-loops (namely, for all $v \in V$, $(v, v) \notin E$).

Definition 1.11 (Path)**.** Given a graph G, a *path* between two nodes $v \neq w \in V$ is a finite sequence of vertices $(v = v_1, v_2, \dots, v_p = w)$ such that, $\forall i \in \{1, \dots, p-1\}$, $(v_i, v_{i+1}) \in E$.

Definition 1.12 (Connected Component, Connected Graph)**.** A *connected component* of a graph G is a maximal subset of vertices $V' \subseteq V$ such that, for each pair of nodes $v, w \in V'$, there is a path between v and w. A graph G is *connected* if there is only one connected component, namely if there is a path between each pair of nodes.

From Linear Networks to Metric Graphs

In this thesis, we focus on geostatistical processes whose underlying spatial domain is a (continuous) network, namely a network whose edges are not sheer relations, but *physical* links between nodes locations. We will call them *graphs with Euclidean edges* or *metric graphs*. There are plenty of real-world examples of phenomena that live on such topologies. To mention just a few, think about water temperature or fish concentration on river networks, car speed or density on road networks, information flow on neuron networks, ...

Data indexed over networks have been attracting more and more interest in the last decades among researchers. Indeed, they allow to represent complex systems whose interactions between components play an important role. Networks are abstract, yet very flexible, frameworks to model such interactions. Traditionally, the mathematical objects taken as domain for the definition of geostatistical processes over networks have been *linear networks*. Linear networks are subsets of an Euclidean space E (usually $\mathbb{R}^2$), composed of nodes (*i.e.* points of E) and of straight segments connecting such nodes. Although this is a first step towards the definition of general continuous networks, it inherits some limitations and constraints given by the Euclidean nature of the space E. Firstly, the *length* of each edge is determined by the Euclidean distance of the nodes it connects. Although this should not be a limitation if we aim to model continuous networks that live on a plane, it could turn out to be an unwanted constraint in more general contexts, such as the modelling of a path network in a mountainous area: indeed, the distance between two points might also consider the difference in their altitudes. Secondly, in linear networks, edges cannot

cross without intersect: if two edges share a point in E, such point is necessarily a node of the network. This may be a strong limitation, for instance, in road networks, where bridges and tunnels allow roads to "cross" without intersect.

In order to relax these limitations, metric graphs have been recently proposed. The core idea is to free linear networks from the rigid Euclidean topology they are constrained into, and therefore consider proper abstract sets. The first definition of graphs with Euclidean edges has been given in Anderes et al. (2020), though similar definitions have been proposed by Bolin et al. (2024) and Filosi et al. (2025a). Here, we report the very first definition given by Anderes et al., yet in the following chapters we may consider slightly modified versions.

Definition 1.13 (Graph with Euclidean edges (Anderes et al., 2020))**.** A triple $G = (V, E, \{\varphi_e\}_{e \in E})$ satisfying the following conditions is termed *graph with Euclidean edges*.

a) **Graph structure.** (V, E) is a simple connected graph, meaning that the vertex set V is finite, the graphs has no repeated edges or edge which joins a vertex to itself and every pair of vertices is connected by a path.

b) **Edge sets.** Each edge $e \in E$ is associated with a unique abstract set, also denoted e, where the vertex set V and all the edge sets $e \in E$ are mutually disjoint.

c) **Edge coordinates.** For each edge $e \in E$, if $u, v \in V$ are the vertices connected by e, then φ_e is a bijection defined on $e \cup \{u, v\}$ (the union of the edge e set and the vertices $\{u, v\}$) such that φ_e maps e onto an open interval $(\underline{e}, \overline{e}) \subset \mathbb{R}$ and $\{u, v\}$ onto the endpoints $\{\underline{e}, \overline{e}\}$.

d) **Distance consistency.** Let $d : V \times V \to [0, +\infty)$ denote the standard shortest-path weighted graph metric on the vertices of (V, E) with edge weights given by $\overline{e} - \underline{e}$ for every $e \in E$. Then, for each $e \in E$ connecting two vertices $u, v \in V$, the following equality holds:

$$d(u, v) = \overline{e} - \underline{e}.$$

Remark 1.1. Definition 1.13 will be adapted to the context in the following Chapters. More specifically, in Chapters 3 and 4 we will not require any "distance consistency" property.

Clearly, given a linear network, it is immediate to build a graph with Euclidean edges with the same structure, as Anderes et al. remark immediately after the definition; while the converse is not always possible. It is in order to notice that the topology of graphs with Euclidean edges is completely abstract: it is built on the notion of *graph*, yet any edge "becomes" a segment of the real line, completely unrelated to others (at least *a priori*).

One may ask what is the meaning of the *distance consistency* property required in the above definition, and why it is necessary. Anderes et al. exploit it in the definition of *geodesic* or *shortest-path* distance: it ensures that the shortest-path between two adjacent vertices is the length of the arc connecting them. However, as we are not going to make use of such a distance (as explained below), we will neglect such a condition. Anyway, notice that, if $\mathcal{G}$ is a graph satisfying the conditions a)-c) but not condition d), it is straightforward to build an *equivalent* graph that satisfies the distance consistency. Indeed, if an edge (u, v) does not satisfy d), then it is possible to "break" it in two or more parts (having equal lengths and whose sum of lengths equals the original one of (u, v)) so that the condition is satisfied.

Effective Resistance Distance on Metric Graphs

Once the formalisation of a metric graph has been provided, since we are interested in seeing it as a spatial domain for a geophysical process, it is somewhat customary to define a *distance* as well. Although the notion of geodesic distance is both intuitive and vastly adopted, it turns out that it is not tailored for the definition of isotropic stochastic processes (Anderes et al., 2020, Theorem 2). Therefore, another notion of distance is needed. One option, which enjoys several good properties, is the *effective resistance* distance (Devriendt, 2022). Effective resistance distance is a physics-based concept: it sees a weighted graph as a resistor network, whose nodes are junction points. Effective resistance distance between two nodes is then defined as the voltage drop when injecting 1 Ampere of current in one and withdrawing 1 Ampere from the other. This ensures that nodes connected by several paths have smaller distance, as current is able to split among these paths (in particular, effective resistance distance is never greater than shortest-path distance). To have a more precise introduction, see Subsection 2.3.1 in Chapter 2.

A remarkable contribution of Anderes et al. is that they *defined* effective resistance on graphs with Euclidean edges: not only is it possible to easily compute the effective resistance between two vertices, but also between any two given points on the graph. In addition, such a computation is straightforward (and fast) once the graph has been set. The way they achieve such a result is via the definition of an *ad-hoc* process on the whole graph, whose variogram is the resistance distance. Such a process is a proper adaptation of a standard Brownian bridge, classically defined on a segment of the real line, to the new topology. We do not report its full construction here, as it is quite involved and it is reported in the next chapters.

Stochastic Processes on Metric Graphs

We are now able to finally focus on how to define stochastic processes on graphs with Euclidean edges. There are two different schools of thought: one, stemming directly from Anderes et al. (2020) approach, aims to define stochastic processes via the direct definition of their covariance functions. The other considers solutions to particular classes of stochastic differential equations. Both approaches have pros and

cons. On one side, the SDEs allow for flexible definitions of processes, attaining, in particular, processes whose realisations are differentiable. On the other hand, the kernel approach is — at the moment and to our knowledge — the only one that has been able to deal with non-trivial spatio-temporal and multivariate settings. Next, we briefly expose how these two schools define stochastic processes.

The kernel approach rests on the fundamental result of Schoenberg (1938), which proved the equivalence between the conditionally negativeness of a distance space and the positive definiteness of the function

$$d \mapsto e^{-\alpha d}.$$

More specifically, we have the following theorem.

Theorem 1.1 (Schoenberg (1938))**.** *Let (X, d) a distance space, then the following are equivalent.*

1. *There exist a Hilbert space $\mathcal{H}$ and a bijection $\varphi : X \to \mathcal{H}$, such that, for all $x, y \in X$,*
$$d(x, y) = \|\varphi(x) - \varphi(y)\|_{\mathcal{H}}.$$

2. *The function*
$$(x, y) \mapsto e^{-\alpha d^2(x,y)}$$
is positive definite on X, for any $\alpha > 0$.

3. *d^2 is conditionally negative definite, namely for each $n \in \mathbb{N}^+$, $x_1, \ldots, x_n \in X$ and $c_1, \ldots, c_n \in \mathbb{R}$, if $\sum_i c_i = 0$, then*
$$\sum_{i,j=1}^{n} c_i c_j d^2(x_i, x_j) \leq 0.$$

From this theorem, and given the fact that a variogram is necessarily conditionally negative definite, we get an easy recipe to construct valid covariance functions by composing a variogram with a negative exponential. Furthermore, since negative exponential functions "generate" the set of completely monotonic functions, we get Theorem 3.1, which states that completely monotonic functions composed with variograms provide valid covariance functions. This construction method is very flexible (notice that no particular assumptions have been made on the space X), however, it has some drawbacks. First, all the covariance functions built in this way are, by definition, isotropic. In addition, they provide covariance functions that are not differentiable in the origin, resulting in realisation that are not differentiable, albeit continuous, almost everywhere. Last, whenever inference has to be made on such processes, a covariance matrix has to be inverted, which is an expensive operation for large datasets.

The SDEs approach, on the other hand, considers known stochastic differential equations operating in $\mathbb{R}^n$ and adapts them to fit on metric graphs. For instance, Bolin et al. (2024) consider the Whittle equation

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau u) = W,$$

previously studied on $\mathbb{R}^n$, to metric graphs. Here, Δ is the Laplacian operator (properly adapted to cope with vertices) and W is a Gaussian white noise. The reason underlying this choice is that its solution on Euclidean spaces enjoys the Matérn covariance function, isotropic with respect to the Euclidean distance. With this construction, it is possible to obtain Gaussian processes on metric graphs whose sample paths are differentiable. However, we believe that the covariance function approach is, albeit perhaps less powerful, more clear and interpretable.

Structure of the Thesis

The reminder of this thesis is organised as follows. Each Chapter is a self-contained scientific article that delves into a specific theme, linked to graphs with Euclidean edges as spatial domains.

Chapter 2 (Porcu et al., 2025) provides a sound analysis of the role of different types of networks in climate science. It defines and explores networks *of* data (namely networks that describe interdependencies between climate phenomena), climate data *over* networks (that is networks as a spatial domain for climate processes) and networks *for* data (the huge machinery based on networks within the realm of machine learning and statistics). Furthermore, it provides an analysis of the possible *bridges* between these networks and a sound motivation for them, which sees the fusion of the three network classes as a necessary condition for the implementation of the Shaw-Steven agenda about climate modelling and analysis.

Chapter 3 (Filosi et al., 2025a) is a pioneering exploration of the time evolution of graphs with Euclidean edges. As literature has mostly considered *static* graphs, even when time is considered, first it rigorously defines time-evolving and periodic time-evolving graphs, where time is always considered discrete. Then it defines a spatio-temporal semi-distance on time-evolving metric graphs and finally analyses a class of isotropic univariate processes on them.

Chapter 4 (Filosi et al., 2025b) aims to extend the work by Anderes et al. (2020) to multivariate setting. More precisely, it finds and analyses the (unique) extension to the distance defined therein satisfying a particular conditional independence structure. In addition, it proves that the element-wise composition of this multivariate distance with completely monotonic functions provides valid multivariate covariance functions. Finally, it provides some generalisations and a more specific analysis on Euclidean trees.

Chapter 5 (Alegría et al., 2025) defines some algorithms to efficiently simulate (one-dimensional) isotropic Gaussian processes over (static) metric graphs. It shows that the main computational burden is given by the number of nodes, whilst the

number of sampling points show a much smaller impact. Finally, it compares the algorithms in terms of both statistical accuracy and running time. We remark that Chapter 5 is extracted from a paper with Alfredo Alegrìa and co-authors, where Alfredo has been the main contributor.

Finally, we draw some conclusions and make considerations on future developments.

Chapter 2

A Triad of Networks and a Triad of Fusions for the *Other Climate Crisis*[1]

Preludium: the *Other Climate Crisis*

The recent *tour de force* by Shaw and Stevens (2025) has caused an intense debate about what climate science can predict and how those predictions are to be justified. While Shaw and Stevens do not question the law of physics to explain global warming, they warn that the paradigm of Large Scale Determinism (LSD) – being the dominant culture in climate prediction – is under strain when analysed within the realm of regional phenomena having direct societal relevance.

According to the LSD paradigm, small–scale processes (deep convection, boundary layer turbulence, ocean eddies) are parametrised as functions of large–scale state variables. The LSD has been a major breakthrough in climate science, and has allowed for a better understanding of certain dynamics, from the attribution of anthropogenic warming to a broad set of large–scale signals. However, it is being questioned because of the increasing amount of discrepancies at regional scales where local phenomena pop up. Rephrasing, Shaw and Stevens note that, as observations lengthen and as regional signals sharpen, the LSD hypothesis encounters accumulating anomalies, which translate into large discrepancies between what is simulated on the basis of LSD and what is locally attained. The most challenging regimes are those where scale coupling is strongest, such as, for instance, the deep tropics, the Southern Ocean, and high–impact extremes. Additionally, there might be uncertainties in boundary conditions and structural choices that observations cannot yet tightly constrain. In conclusion, Shaw and Stevens claim that this is the "Other Climate Crisis" of methodology and predictability.

Shaw and Stevens do not deny the established agenda. However, they construc-

[1]This chapter is a version of the paper Porcu et al. (2025), available on arXiv:2510.09728.

tively criticise the existing paradigm and integrate it with (a) embracing discrepancies as tests, (b) resort to hierarchies, and (c) make experiments with disruptive tools (Shaw and Stevens, 2025). As for (a), embracing discrepancies allows to treat mismatches as opportunities to falsify mechanistic claims, and to design observations that actually adjudicate between competing hypotheses. Regarding (b), hierarchies are a natural instrument to provide fusions in the sense of components (atmosphere, ocean, land) and scales (from convection to planetary waves), so that complexity does not obfuscate interpretability. Finally, disruption (c) can be achieved in three directions: very large ensembles for an improved uncertainty accuracy; global models run at kilometre resolution; and machine learning, both as a tool as well as a new paradigm for building predictive emulators. According to the authors, weather predictability has a mature theory, and climate predictability does not. Understanding how the future looks like under these new paradigms is a major scientific task for the coming decade (Shaw and Stevens, 2025).

The perspective provided in the present paper (a triad of networks for climate data and a triad of fusions) stems from Shaw-Stevens motivation while deviating ontologically, and while taking a complementary route. Our point is that our organization of networks and consequent paradigm provides a concrete support for the implementation of Shaw-Stevens agenda. The way this will happen is described throughout, and we avoid anticipating too much in this direction. And there is more: we show how what we term fusion *within* the proposed network triad, and then *between* the triad and the Shaw-Stevens agenda, becomes then the natural environment of what we term a *meta* fusion that opens the door to the upcoming Shaw-Stevens networks ecosystem.

The remainder of the paper develops these claims.

2.1 Introduction

2.1.1 Context

The word *network* is everywhere in the entire landscape of contemporary data science and artificial intelligence. At the methodological level, the word *network* is reminiscent to the architectures that are the basis of the transformation of modern learning – starting with the densely connected layers of neural networks, until the graph based operators of geometric deep learning. At the applied level, *network* is a natural way of thinking when the goal is to describe the structures through which complex systems are represented: social networks for interaction, transportation networks for flow, and climate networks for interdependence, to mention a few.

Hence, the word *networks* transcended their canonical meaning to become a universal language for expressing relation, propagation, and emergence. So much so, that we find difficult to talk about data science or AI without speaking about *networks*.

Climate (data) science is certainly an area where networks are ubiquitous. Net-

work-based methods integrate both data-driven and model-based tools to uncover underlying structures in climate dynamics, particularly in complex space-time systems. This paper introduces a triad of networks that serves as a unifying framework for interpreting climate data. The paper is not a descriptive and aesthetics exercise, as will be shown throughout the subsequent sections.

(a) Networks *of* Climate Data: here, we start from climate data (typically, climate model outputs) and construct a network where the nodes are the spatial locations, and where the edges are a graphical representation of certain statistical dependencies. For example: *if the temporal correlation between a time series at location x_1 and another at location x_2 is bigger than* 0.5*, then we draw an edge connecting them.*

(b) Climate Data *over* Networks: in this case, the network structure is not inferred through some statistical procedures, but given. It is a representation of a spatial, or a physical, or a topological structure (*e.g.*, a river basin, transport grid), and climate data (e.g., temperature, precipitation) are modeled or predicted as processes that are continuously indexed over the nodes and edges of this network.

(c) Climate Data *for Networks*: modeling approaches based on networks, that are typically coming from machine learning and AI, and are used to understand the structure of climate data (unsupervised learning) or predict their patterns (supervised).

These classes are not mutually exclusive: there are obvious intersections, especially the pairs (a,c) and (b,c). Less obvious is the intersection (a,b) – however, this inspires considerable amount of research, new methods, and climate applications for the future.

The juxtaposition of these paradigms serves as a platform for providing unifying perspectives, highlight methodological advances, challenges, and provide opportunities to integrate different paradigms and philosophies in climate data science.

Table 2.1 opens for some preliminary considerations about the difference between the first two paradigms. Clearly, the first two paradigms are coming from similar

Feature	**Networks of Climate Data**	**Climate Data over Networks**
Network Source	Inferred from Data	Pre-Determined
Node	Spatial Location	Locations in a Structured Network
Edge means	Statistically Similar (Synchronization)	Physical connection
Covariance	Through similarity	Direct constructions
Typical Applications	Telecommunication Analysis, Flipping Points	Kriging

Table 2.1: A comparative description of networks *of* and *for* through some features.

frameworks, as the reference set is, for both, space-time. Hence, the space-time correlation plays an important role for both. Robustness from both perspective

of noisy data and models covers a central aspect for both paradigms. A recent contribution (Haas et al., 2023) proves how important can be sampling errors for the first paradigm.

The three paradigms are compatible, and their fusion and bridging will cover an important part of the subsequent sections.

There are clear intersections (pairwise and triplewise): *of* and *for* have an obvious intersection as the networks extracted from data are then used as analysis tools – for instance, feedback into models, or guiding assimilation strategies. As for *of* and *over*, the topologies extracted from data can be then used as "spatial" domains where some processes are continuously defined. For instance, propagation of certain phenomena in climate dynamics. *For* and *over* have an obvious intersection - for instance, the topology created through the first might be embedded in the second to build an hybrid structure. Finally, the three of them intersect through the sequence *of-over-for*. For instance, dynamic climate networks used in spatiotemporal modeling over changing graphs.

2.1.2 Networks *of*: Some Literature

The interpretation of the climate system as a network emerged after the painstaking work by Prof. A. Tsonis and his group. The work by Tsonis et al. (2006) provides a paradigm where geographic locations or climate indices are nodes, where their statistical affinities are edges.

Tsonis' paradigm provided a simple construction that fixed a common language for teleconnections, allowing correlations, mutual information, or event synchronization to be re-expressed as a topological structure.

Subsequent developments and findings extended the paradigm through the inclusion of questions related with architecture and organization (see the characterization of climate network topology in Tsonis and Swanson, 2008; Tsonis and Roebber, 2004).

While Tsonis paradigm has been the building block, subsequent contributions have provided more sophistication inside a very promising architecture. This new wave has gone under the trendy word of *complex networks theory*: Donges et al. (2009a) introduced the notion of a basis network to reduce redundancy and emphasisethe most informative connections, while Donges et al. (2009b) provided a broader methodological framework built upon linking climate dependence measures with canonical network diagnostics (degree, clustering, centrality, modularity).

Introducing temporal evolution within complex systems is a natural step. Evolving networks allow, for instance, to better understand different types of El Niño episodes as they allow to track changes in connectivity patterns (Radebach et al., 2013). By quantifying structural changes in connectivity, network measures allow an objective separation of Eastern- and Central-Pacific El Niño/La Niña episodes (Wiedermann et al., 2016).

Networks *for* have been proved useful to show that connectivity among rare events (extreme values) can reveal long-range structures that would result undetected

with classical methodologies (for instance, Gaussian processes) as shown for Indian Summer Monsoon precipitation (Stolbova et al., 2014) and for global extreme-rainfall teleconnections (Boers et al., 2019).

Stability, embedding, and prediction have also played a major role regarding sophistication of the basic networks structure. Berezin et al. (2012) studied persistence and stability in climate networks. They proved that much of their apparent robustness is due to the geographical embedding of nodes in concert with the physical coupling among different regions of the climate system. Predictive analysis has drown the attention of several scientists: Ludescher et al. (2013) reported improvements in El Niño forecasting via cooperativity detection, while interaction-network early-warning indicators were proposed for a potential AMOC collapse (Van Der Mheen et al., 2013) and further consolidated through observation-based signals of AMOC weakening (Boers, 2021).

All the previously-mentioned contributions are mostly based on *ad hoc* procedures rather than statistical rigor. This has come into place more recently, with Deza et al. (2013) for interdependencies at multiple time scales, and Runge et al. (2015, 2019a,b) for causal-discovery frameworks in spatio-temporal data.

This natural evolution — from basic to sophisticated models, from *ad hoc* procedures to statistical rigor — finds a natural step in implementing rigorous reassessments of inference practices. Haas et al. (2023) asserts that finite samples, spatial autocorrelation, and seasonal nonstationarity may inflate connectivity. Their findings suggest surrogate-based nulls, spatial block resampling, and uncertainty quantification as gold standards in statistical practices. Taken together, these developments move *networks of climate data* from heuristic discovery toward a statistically principled toolbox: Tsonis-type constructions reveal coherence, teleconnection, and tipping precursors; complex network diagnostics and causal frameworks organisethese signals; modern validation closes the loop between insight and robustness. This provides the empirical vertex of our triad: the descriptive structures of *networks of* will inform the geometric realism of *climate data over networks* and the task-oriented machinery of *networks for climate data*, developed in the subsequent sections.

2.1.3 Climate Data over Networks: a Recent Story

Data science has been a revolution in many theoretical and applied fields, and has provided many challenges within the realm of space-time data (see Anderes et al., 2020; Moradi and Mateu, 2020; Baddeley et al., 2021; Rakshit et al., 2017, for recent contributions). Data *over* networks have preoccupied both statistical and machine learning (ML throughout) communities for a longtime. Within ML community, the amount of literature is huge. However, only a small part of it is climate-focused, and mostly serves as a computational background as per Section 2.1.4. Since a good part of this manuscript is related to network topologies, we mention that topological complexities through ML under the framework of data analytics are discussed in Stanković et al. (2020). Another relevant comment is that, in the great majority of ML related contributions, the process is assumed to be defined exclusively over

the vertices of the graph. The extension to processes that are continuously defined over both graphs and edges requires substantial mathematical work, and has been challenged only recently under the paradigm of Gaussian processes over metric graphs (Porcu et al., 2023; Bolin et al., 2024), which encompasses other sophisticated structures such as linear or nonlinear (generalized) networks, or over Euclidean trees (Anderes et al., 2020).

The main advantage of working with Gaussian processes is that the dependence structured is completely specified through the covariance function (Porcu et al., 2021). What makes things complicated is the covariance function typically depends on the distance between every pair of random variables defined over the graph. This is where the topology and the language of metric graphs (and metric spaces) comes into play.

Additionally, metric graphs might not be enough to describe sophisticated topologies. For instance, climate processes might temporally evolve over these topological structures. Building covariance functions over networks is mathematically challenging as it will be explained in subsequent sections. The task is made even more complex due to the fact that specifying the right metric is extremely challenging. Notable works in this direction have been made by Ver Hoef et al. (2006) and Peterson et al. (2007) to model stream river flows.

Anderes et al. (2020) have proposed graphs with Euclidean edges, which are graphs where each edge is associated with an abstract set being in bijective correspondence with a segment of the real line. In turn, this provides each edge with a Cartesian coordinate system to measure distances between any two points on that edge. Alegría et al. (2025) proposed efficient algorithms to simulate isotropic fields on graphs with Euclidean edges, while Filosi et al. (2025a) (see Chapter 3) defined temporally-evolving networks and processes on them. Finally, Filosi et al. (2025b) extended the processes defined by Anderes et al. (2020) to the multivariate setting. Different, albeit similar, graph structures have been considered by Bolin and Lindgren (2011) and Bolin et al. (2024) through Gaussian Markov random fields.

2.1.4 Networks *for*: State of the Art

A third and increasingly influential perspective concerns networks *for* climate data. Here, networks represent tools, mostly computational instruments rather than objects of inference or observation. As previously specified, within the *for* class of networks the topological structure is not inferred *from* climate data, as in networks *of*, nor prescribed *over* a given geography, as in networks *over*. Instead, the topology is constructed to perform a task, which might be prediction, reconstruction, downscaling, or data assimilation.

We note that under *for*, edges are not necessarily understood as physical teleconnections. More broadly, they are operational devices where any source of information and uncertainty is used to execute a specific model objective or task. This implies a conceptual shift from descriptive to functional use of networks, from mapping interactions to *designing* them.

ML, signal processing, and statistical inference converge within this paradigm. While there is a wealth of tools especially coming from the ML literature, our manuscript will focus on Graph Signal Processing (GSP), Probabilistic Graphical Models (PGMs), and Graph Neural Networks (GNNs). The reason is, a part from their popularity, that these three classes of tools provide a hierarchy of increasing flexibility and abstraction. Further, our manuscript will argue that these tools perfectly fit inside the call by Shaw and Stevens (2025) for "disruptive computation with interpretability".

We anticipate that each of these three families covers a complementary layer in the construction of networks *for*: GSP offers a basic (linear) and interpretable foundation layer; PGMs introduce conditional independence and uncertainty quantification; GNNs extend both to nonlinear, data-driven propagation across space and time. Some short introduction follows for each of them.

In GSP, the graph represents the geometry of computation and their goal is to process or reconstruct signals that live on a known graph. GSP treats a climate field as a signal defined on the nodes of a graph. Within this class of tools, spectral graph filters are typically used to smooth out or denoise, while graph Laplacians provide a natural language for interpolation and uncertainty quantification under missing observations (Shuman et al., 2013; Ortega et al., 2018). These methods have been successful in climate modeling, in particular to design optimal sensor networks, filter satellite data, or reconstruct fields affected by observational gaps.

PGMs are based on conditional dependence structures where a key role is played by the sparsity of the precision matrix, which in turn allows for scalable inference while retaining interpretability (Lauritzen, 1996; Friedman et al., 2008; Bolin et al., 2024). Dynamic Bayesian networks represent the natural generalization attained to incorporate temporal dependence, and have been largely used within the realm of causal discovery and multi-step forecasting (Shimizu et al., 2006). Within climate science, PGMs have been used to explore directional interactions (causal teleconnections), to estimate lagged dependencies across hemispheres, and to quantify uncertainty in ensemble forecasts.

GNNs generaliseboth GSP and PGMs by learning nonlinear propagation mechanisms across the graph. They perform the taks of adapting the connectivity to optimisepredictive skill (Defferrard et al., 2016). GNNs have proved effective in basin-scale flood prediction, temperature downscaling, and urban heat mapping. Physics-informed variants integrate physical constraints such as mass conservation, monotonicity, and boundary conditions, aligning computational scalability with the laws of thermodynamics and hydrology. These methods yield improved forecasts while maintaining interpretability and physical consistency.

Some comments are in order. The objective of networks *for* is thus *functional* rather than representational, in the sense that they represent both tools for computation, but also layers within hierarchical structure, within the broad scope of solving a task as previously described. Hence, under *for* the abstract notion of graph is embedded into computational engine, and the same topology becomes a design

variable.

Networks *for* are a natural tool to execute the Shaw-Stevens agenda as they provide a computational framework that implements the call for disruption as per Shaw and Stevens (2025). While the LSD framework remains generative and equation driven, networks *for* are inductive and relational. The coexistence of these two epistemic cultures (physical closure and relational computation) anticipates the fusion perspective developed in subsequent sections, where the triad of networks becomes a practical instrument to implement the Shaw–Stevens agenda.

2.1.5 Contribution

We guide the reader through the following facts.

- We describe what we term *Tsonis* networks as a synonym of networks *of* climate data. We illustrate how to build these networks using certain classes of (dis)similarity measures. An extensive literature review opens for constructive criticism about Tsonis networks, while, at the same time, recognisingtheir indisputable impact on the scientific literature as well as on measures of governance for climate change.

- We use the synonym *geophysical* networks to refer to the situation where climate data are observed over networks with geophysical or geographical constraints, that is, with a predefined topology. We illustrate three situations, namely (a) static networks (observations that are not repeated over time), (b) product spaces, where observations over a metric graph are repeated over time, but the topology of the draft does not change, and (c) dynamical metric graphs with their inducted metrics. We put special emphasis on Gaussian processes defined over geophysical networks: we describe their covariance functions and their metrics. We discuss the impact of this part of literature on climate sciences.

- We provide a thorough description of three classes of networks *for* while avoiding mathematical obfuscation.

- A triad of networks is then followed by a triad of bridges. Bridges *within* provides a perspective of a *corpus unicum* within the triad. For the bridges *between* perspective, the previous step of a bridge *within* becomes the perfect tool to implement the Shaw-Stevens agenda. This all culminates into a *meta* bridge that allows to build a new ecosystem for the coexistence of Shaw-Stevens agenda with our triad in a new sistem of knowledge that dynamically updates the data universe through data discoveries.

The paper is written for a broad public. Hence, we have avoided mathematical formalism, while focusing on the ideas and the way things are getting done for each type of network. For more mathematically-grounded introductions, the reader can find, for each subject, a detailed reference list.

2.2 Networks *of* Data: The Tsonis Paradigm

2.2.1 Tsonis Networks and Topologies

Given the impact of the Tsonis school in this field of research, the remainder of the manuscript will adopt the nomenclature *Tsonis networks* for the described networks *for*. Accordingly, we shall respectively term Tsonis paradigm and Tsonis topology the framework adopted by the Tsonis school and the type of topologies described therein. As for the second, the term Tsonis topology is actually an umbrella that embraces different topologies, which are described below for the convenience of the reader. Specifically, Tsonis topologies, which are summarized in Table 2.2 and illustrated in Figure 2.1, include the following.

Regular Topologies. The topology is very simple, as nodes are located over a lattice on the two-dimensional sphere. They exhibit deterministic patterns, and the structure of the nearest neighbor is fixed. Examples of these networks can be found in atmosphere or ocean grids that are used for simulations (without considering teleconnections).

Random Topologies. The topology is that of graphs that are purely non-deterministic. Although there is a wealth of available models for random graphs, the literature makes abundant use of *Erdős–Rényi* structures, where edges are equiprobable, which implies homogeneous degree distributions. They can be used as *null* hypothesis within hypothesis testing statistical procedures, for instance to test for clustering and path length (Tsonis et al., 2006; Donges et al., 2009b). Normally these networks are associated with spurious connectivity. See also the more statistical approaches by Hlinka et al. (2013); Runge et al. (2015).

Small World Topologies. These networks exhibit low path-length and high clustering. The last indicates that neighboring nodes (that is, regions with highly correlated time series) are themselves likely to be interconnected. Short-average path length means that the network has efficient connectivity because any two nodes can be connected through a small number of edges. Under such characteristics, there is a high degree of local coherence, yet, at the same time, long-range interdependencies between regions that are very distant from each others. Several works have shown that *climate is a small world Tsonis network*, and we refer to subsequent sections for more clarity and documented references.

Some comments are in order. Rather than being descriptive, Tsonis approach usually serves as inquisitive and diagnostic tool: the first two types of topologies are actually used to empirically test whether a climate system is a small world or not, exhibiting high clustering and short path lengths relative to a random baseline. Hence, Tsonis topologies should be taken as diagnostic archetypes — not design families.

Topology	Description	Climate Relevance	Key Studies
Regular	Nodes on lattice, deterministic. Each node connected to fixed neighbors.	Used in climate models and spatially-regular sensor grids.	Donges et al. (2009a), Berezin et al. (2012)
Random	Usually Erdős–Rényi graphs: homogeneous edge probability.	Null-model for statistical testing. Identification of random connections.	Tsonis et al. (2006), Hlinka et al. (2013), Donges et al. (2009b)
Small world	High clustering. Short (average) path length.	Captures local coherence and persistent teleconnections.	Tsonis et al. (2006), Donges et al. (2009b), Berezin et al. (2012), Feldhoff et al. (2012), Ludescher et al. (2013), Tsonis et al. (2011), Boers et al. (2019)

Table 2.2: A summary of the different types of Tsonis Networks and their uses.

Another relevant point is that, within the literature related with Tsonis networks, scale-free networks have covered central importance. These are topologies that exhibit power law distributions that describe the distribution of the degree associated with their nodes. In other words, most of the nodes have a small number of links, while some few have a incoherent high number of connections (Barabási and Albert, 1999). Hence, these networks are quite different from Tsonis networks as they exhibit a high level of heterogeneity. In particular, Tsonis topologies are not cleanly power-law; further, they exhibit exponential or truncated power-law tails rather than genuine scale-free behavior; finally, Tsonis networks are the corollary of thresholding continuous correlations, which is not a preferential attachment process as per Barabási and Albert (1999).

Scale-free networks suggest dominant regions that have a predominant role within the climate system they belong to. For this kind of network, the probability for a given node to have k links is of the order to $k^{-\gamma}$, with typically $2 < \gamma < 3$ (Barabási, 2009). Table 2.1 provides a schematic comparison within Tsonis topologies.

2.2.2 Tsonis Paradigm

To build a Tsonis network, correlations are computed between each pair of nodes based on a chosen climate variable (*e.g.*, surface temperature or atmospheric pressure). Then, edges are built through a simple rule: an edge between two nodes exists if and only if the correlation between them overpasses a given threshold. The resulting adjacency matrix defines a static or time-evolving climate network.

To prove evidence of a small world network, Tsonis et al. (2006) provide statistical evidence of a big discrepancy with purely random graphs in terms of cluster coefficient, as well as in terms of path length. All these evidences support the idea of a system that is not completely random, nor completely regular.

The structure of the network can change over time. Some nodes can keep what

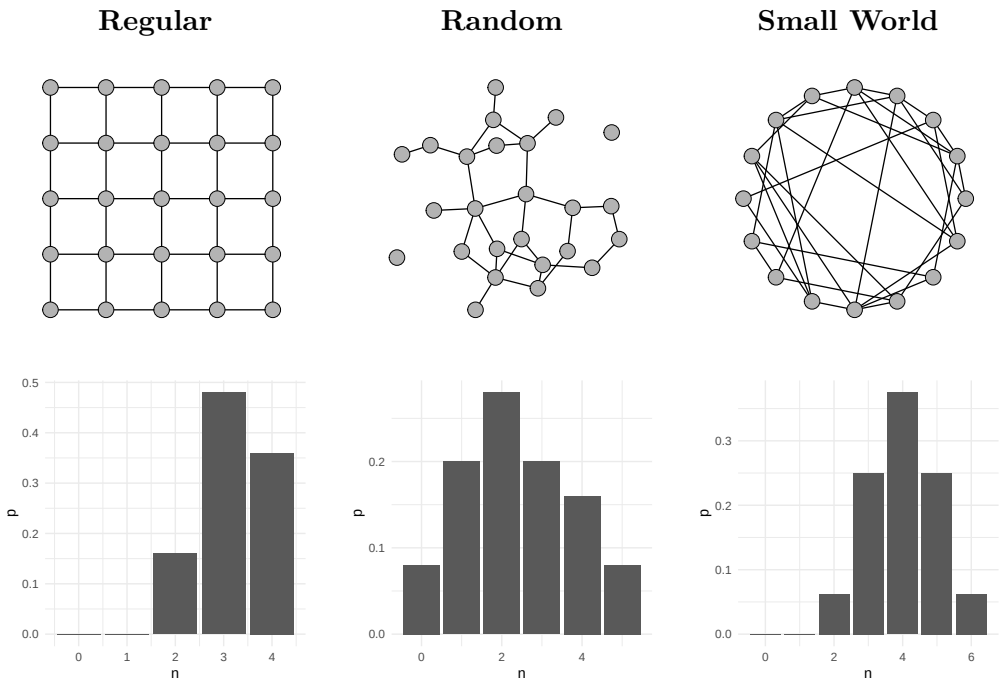


Figure 2.1: A representation of different Tsonis networks with their degree distributions.

the authors call *high degree centrality* - that is, they maintain strong correlations with many other locations being even far apart across the globe. A notable example is ENSO (El Niño Southern Oscillation) in the Equatorial area. The geographical region corresponding to ENSO has a high correlation with other regions, even being far apart. This can be translated in terms of Tsonis network as the ENSO nodes have a high degree. Hence, networks help strengthen some already-established theories of teleconnections in the climate system.

More interesting things are actually stemming from the structure of Tsonis networks. For instance, Tsonis et al. (2006) show that the climate network can be decomposed into clusters of nodes, which in turn are associated with teleconnection zones or climate regimes. Hence, physically meaningful subsets of the nodes can be extracted. In conclusion, Tsonis networks are extremely useful to capture local aspects (coherence) and global teleconnectivity.

Special attention by this group of authors is devoted to *El Niño* and *La Niña*. During strong El Niño events, the global climate network registers notable changes in terms of connectivity patterns.

Tsonis network have been extremely influential for this part of the literature, as reported in the introduction to this paper. More rigor in terms of statistics is offered by Donges et al. (2009b). Tsonis networks are the backbone for climate

network approaches, such as tipping point detection and causal discovery. A tipping point is a threshold beyond which a small perturbation can lead to a qualitative, often irreversible, change in the state of a climate subsystem. Examples of tipping points include the collapse of the Atlantic Meridional Overturning Circulation (AMOC), the disintegration of polar ice sheets, and the transition from rainforest to savanna. Within the framework of Tsonis networks, tipping points can represent sharp transitions in the network topology. Such transitions can provide important indications, such as loss of coherence in the climate system (for which the sign will be a drop in connectivity), or desynchronization (seen through a sudden rise in network fragmentation), or major shifts in climate change, which can be seen through change in betweenness centrality.

2.2.3 Criticism and Impact

Tsonis paradigm is experimental rather than methodological. While its impact has been largely acknowledged in the literature, more recently some constructive criticism has emerged. The main reason is the lack of methodological and statistical rigor. The main pitfalls within Tsonis networks are associated with the following aspects, which in turn call for more rigor into future research.

1. Tsonis networks are threshold-sensitive. The topology of the network is dramatically affected by the chosen threshold. Intuitively, high threshold will allow to isolate neater clusters, while low threshold create geographic macro clusters.

2. Autocorrelation bias happens because spatial autocorrelation can "artificially" increase collocated correlations (*e.g.*, Person correlation).

3. Sampling variability is a big issue and has been examined to a very limited extent.

Haas et al. (2023) is a notable example of methodological innovation and statistical rigor inside the Tsonis architecture. They provide constructive criticism on the choice of similarity measures and their impact. They prove that there is a risk of having spurious links because of finite sample effects, spatial autocorrelation, and seasonal nonstationarity. The alternative proposal by Haas et al. (2023) is that of surrogate-based null models in concert with bootstrapping techniques in order to understand the *significance* of inferred edges. Further, they claim to be able, through this approach, to quantify uncertainty in network metrics.

Haas et al. (2023) point the ways to Tsonis networks improvement, through, for instance:

1. objective criteria to threshold selection;

2. robustness for the selection of links;

3. improvement or extensions such as dynamical and multivariate.

Hence, Haas et al. (2023) contribution can be seen as a rigorous statistical formalism that allows for improvements within the realm of Tsonis networks. The main innovations in Haas et al. (2023) come from spatial block-bootstrap methods as a replacement of threshold selection. Further, Haas et al. (2023) implicitly suggest an integration of data-driven with model-driven choices in climate sciences.

The scientific impact of Tsonis paradigm is unquestionable: Tsonis frameworks created an insatiable thread of contributions, where many of them can be seen as methodological innovation, or as more sophisticated architectures that use Tsonis topology as a baseline. Within the extension of similarity measures, mutual information and nonlinear metrics have been introduced by Donges et al. (2009b), and subsequently by Berezin et al. (2012) and Deza et al. (2013). Rigorous approaches to synchronization have been taken by Boers et al. (2019) and by Stolbova et al. (2014). Early warnings (tipping points) have been challenged in Tsonis and Swanson (2008) and further studied by Berezin et al. (2012). Then, Runge et al. (2015) supported causality-connection findings through graphical models. Wiedermann et al. (2016) worked on ENSO predictability. Similar works can be found in Tsonis et al. (2011). In causal inference, notable works are those of Donges et al. (2015) and Runge et al. (2019a). While the previously-mentioned contributions might be seen as generalizations of a basic structure, it is also true that Tsonis paradigm triggered a diversified portfolio of methodological innovations that range from statistics, dynamical systems and causality, and that intersect with networks *for* in several directions (for instance, GNNs). Applications have clearly benefited of these methodological insights. To mention a few of them, early detection of tipping points (for instance in monsoon systems) can be found in Boers (2021); Stolbova et al. (2014). Circulation regimes in the arctic region can be found in Boers (2021); Feldhoff et al. (2012). Analysis of extreme events (floods) are accounted Taghizadeh et al. (2025). ENSO, NAO, IOD studies can be found in Stolbova et al. (2014); there are also studies of extreme rainfall patterns (Boers et al., 2019) and Antarctic/Arctic dynamics (Boers, 2021).

2.2.4 Tsonis Networks and Climate Governance

Within the past two decades, there has been a clear shift towards data-driven decisions in climate governance. The influence of Tsonis networks extends beyond academia into climate resilience planning, early-warning design, city governance, and multi-actor policy coordination.

Regarding early warning systems, all agencies at both national and international level, such as NOAA, ECMWF, and WMO, have accepted metrics based on networks to integrate geophysical indicators (Runge et al., 2019a). See for instance how Boers (2021) has established important transitions in terms of abrupt ENSO phase shifts, and this enables anticipation of regional drought or flood regimes. The same remark applies for national agencies in India and East Africa with NGOs such as Red Cross Climate Center.

Networks representations are definitely helpful for urban planning. The celebrated projects under the *C40 Cities* and *Cities for Climate Protection* (CCP) initiatives

use graphs and network-based methodologies to identify cities vulnerabilities and cross-regional spillovers (Acuto, 2020; Abel, 2025).

Tsonis networks can be seen as *boundary objects* (Star and Griesemer, 1989): real or abstract entities that can be used across heterogeneous communities while keeping their identity. Applying this concept to climate governance, we see that Tsonis networks are simultaneously a scientific paradigm, an instrument for data science visualization, and an instrument for policies.

The logic of Tsonis networks aligns with the post-Paris logic of polycentric governance. National contributions are part of a network commitment, and agencies are now calling for network interdependence (United Nations Framework Convention on Climate Change, 2024). The *Climate Technology Centre and Network* (CTCN), a UNFCCC entity, supports the use of climate networks in guiding adaptation investment, especially in Least Developed Countries (LDCs) (Climate Technology Centre and Network, 2024). Likewise, World Bank resilience planning now includes network fragility metrics in urban diagnostics.

In conclusion, Tsonis networks go beyond data science and enter the realm of decision science, influencing a wealth of directions where important decisions need to be taken for a better future of our planet.

2.3 Climate Data *over* Networks

2.3.1 What are geophysical Networks?

The previous sections have focused on networks of climate data as topologies that shape up from statistical dependencies, such as correlations, information-theoretic links, or any other form of statistical dependencies between space-time climate variables. These networks are typically inferred from data, with the goal of revealing latent connectivity patterns in the climate system.

The next paradigm considers climate data *over* networks: here, a geophysical processes is modeled on top of pre-specified spatial or structural graphs. These networks may arise from hydrological catchments, transportation grids, sensor layouts, or river basins, and are typically defined *a priori* based on physical, infrastructural, or geometric constraints.

Throughout, the predefined networks described in this section will be termed geophysical networks. Accordingly, the paradigm ruling geophysical networks will be termed a geophysical paradigm.

The language of metric graphs is the best platform to mathematically describe networks of the geophysical type. Recent research has proven how these graphs are extremely flexible in that they allow for space-time processes as well for spatially dynamical processes, where both vertices (nodes) and edges can appear or disappear over time.

A very important innovation with geophysical graphs is that the great majority of earlier contributions considers processes that are exclusively defined over vertices.

The definition of a process that is continuously indexed over both vertices and edges requires a fair amount of mathematical work. Geophysical graphs are ingenious topological structures that allow to generaliselinear networks to nonlinear edges. Further, the process defined over such structures can have realizations over any point over the edges, and not only in the nodes.

More simplified forms of networks have already received attention within the ML community (Alsheikh et al., 2014; Georgopoulos and Hasler, 2014; Hamilton et al., 2017; Pinder et al., 2021; Borovitskiy et al., 2023). We already mentioned the increasing availability of this kind of networks for spatial data (Cressie et al., 2006; Gardner et al., 2003; Ver Hoef et al., 2006; Peterson et al., 2013, 2007; Montembeault et al., 2012) and point processes (Xiao et al., 2017; Perry and Wolfe, 2013; Deng et al., 2014; Baddeley et al., 2017).

Surprisingly, all the above mentioned contributions refer to static networks (see also the more recent approach by Bolin et al., 2024), where temporal dynamics have not been part of the game, till very recently.

Porcu et al. (2023) and Tang and Zimmerman (2024) overcome this problem by considering product metric spaces, where one space is the network and the other is time. However, a limit of this construction is that the topology cannot be adapted to change over time. This drawback has attracted the interest of several applied scientists (Hanneke et al., 2010), but produced very limited results. See fore instance Mankad and Michailidis (2013); Cherifi et al. (2019); Lim et al. (2019); Divakaran and Mohan (2020); Rossi et al. (2013) within the framework of structural changes detection. All these approaches do not allow for processes that are continuously defined over the network.

Filosi et al. (2025a) (see Chapter 3) proposed time-evolving graphs where (discrete) time is either linear o periodic. They argue that such a choice fits several real-life situations, to include for long-term as well as periodic behaviours. They provide the example of temperatures in a given geographical area, where there might be strong correlations between: (i) contiguous spatial points at a given time (represented by means of spatial edges); (ii) the same points considered at contiguous times (represented by means of temporal edges between temporal layers) and (iii) the same points considered at the same periods of the year, which are considered in the model as they are exactly the same point in the temporally evolving graph.

A gentle mathematical description of geophysical networks follows. When defining geophysical networks, we are referring to, either, (a) a static metric graph, (b) a product space involving metric graphs, or again (c) a temporally dynamical metric graph. A brief illustration about metric graphs follows. A closer focus will be given to a broad class of metric graphs that are termed graphs *with Euclidean edges* after Anderes et al. (2020).

A metric graph is a graph whose edges are not only relations between vertices, but also sets of *physical* points. More formally, a graph $G = (V, E)$, where V is the set of vertices and $E \subset V \times V$ is the set of edges, becomes a metric graph when each edge $e \in E$ is associated to a geometric segment, say $s(e)$, of the real line (see Figure

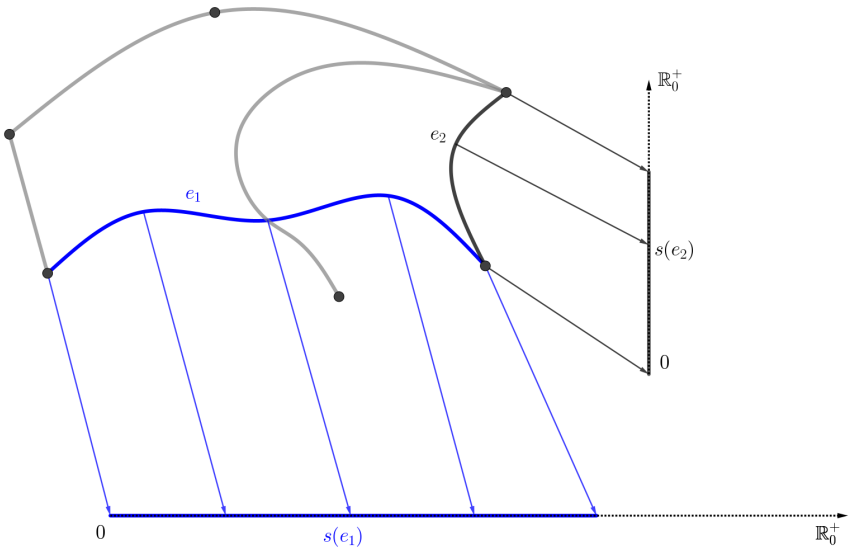


Figure 2.2: A graph with Euclidean edges, where the bijections between the edges e_1 and e_2 and their abstract segments $s(e_1)$ and $s(e_2)$ have been highlighted. Figure adapted from Filosi et al. (2025a) (see Chapter 3).

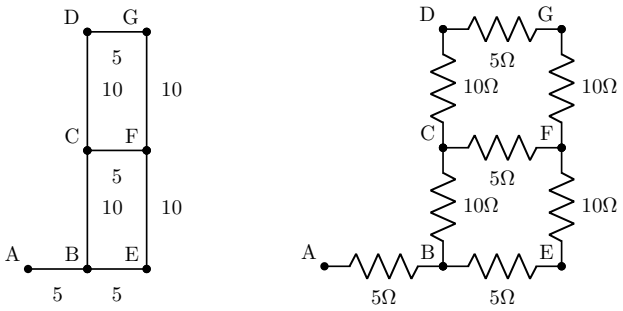
2.2). In such a way, each point on the segment $s(e)$ has its own geometrical meaning. Typically, the *length* of the segment $s(e)$ coincides with the *geometrical length* of the edge e, which is usually inherited from the meaning of the graph G. For example, if G represents a road network, where nodes are crossroads and edges are connections among them, the length of an edge e could be naturally chosen as the physical length of the relative road. However, there are no constraints that force this choice. The advantages of such a topological structure is that it is possible to define distances among any couple of physical points.

Once such a structure has been defined, it is natural to consider some extension: for instance, it is possible to define a product space between a metric graph and an arbitrary domain. Perhaps the most intuitive example is to consider a graph with Euclidean edges cross time (Porcu et al., 2023). This means that it is possible to consider every point in the graph at different time instants. Thinking about road networks, this allows to model a city over an arbitrary time set, such as a day or a year (see Subsection 2.3.3). This approach paves the way to the extremely reasonable question: may graphs change over time? Filosi et al. (2025a) (see Chapter 3) have defined an ad-hoc mathematical structure that represents both graph with an arbitrary dynamics over discrete time (that is: nodes and edges may appear or disappear at subsequent time instants) or with periodic evolution (see Figure 2.7). Although this may seem a strong hypothesis, they argue that it may appear in some specific context, such as river deltas governed by the tide or road networks where traffic lights induce a periodic dynamic of the flow.

A linear network is a special case of a metric graph that is basically the union of

nodes through linear segments on the plane. For such a case, measuring distances over a linear network is straightforward. However, defining distances (called metrics) over a general metric graph is challenging.

One intuitive way to define distance between any two points over the metric graph is by using the shortest-path distance, which measures the total length of the shortest path that connects two nodes. This metric has been proved by Anderes et al. (2020) to have limited applicability (details coming later). As an alternative, they propose the effective resistance distance, which takes into account *all* the paths that connect two given nodes. As a consequence, the resistance metric can be more useful than the shortest-path in several contexts. To give an example, think about a pedestrian network as depicted in Figure 2.3: people that have to go from point B to point G have 3 possible paths, all with the same total physical length of $25m$. Therefore, it may be reasonable to "reduce" the distance between them, as they are more interconnected then the nodes B and A (for which there is just one path). Effective resistance distance overcomes this problem as it models a resistor graph, where edges are electrical resistors (whose resistances are proportional to edges lengths) and nodes are junction points. Thus, for example, two resistors connecting two nodes in parallel result in a lower effective resistance, according to the Ohm's law, whilst two edges connected by a single path will have a resistance distance equal to the shortest-path metric. For a more detailed analysis of the effective resistance distance, the reader is deferred to, for instance, Jorgensen and Pearse (2010). Anderes et al. (2020) found an ingenious method for extending the classical resistance distance to geophysical networks, where distances can be computed not only between nodes, but also between any pair of points on the graph.



N_1	N_2	d_{SP}	d_R
A	B	5	5
B	G	25	12.1
E	G	20	12
C	F	5	3.6

Figure 2.3: A physical network (left), its associated resistor graph (center) and distances between some couple of nodes (right). Distances are always measured in meters. d_{SP} denotes the shortest-path distance, while d_R is the resistance distance.

2.3.2 Climate Processes over Geophysical Networks

A random variable is, roughly speaking, a mathematical object that accounts for some random behaviour. Such objects are extensively used to rigorously model,

predict and asses the variability of real random phenomena, including climate. As an example, it is of obvious interest to know what the temperature will be tomorrow in a given spatial point.

A *random field* is a collection of random variables, that is: a set of random variables indexed by some additional (non-random) variable. This can be formally represented via $\{Z(\boldsymbol{x}) : \boldsymbol{x} \in X\}$, being X the above-mentioned set of indices. Thinking about the temperature example, a random field could be the temperature at a given spatial point through the hours of a day: in such a case the formal representation could be $\{Z(t) : t = 0, 1, \ldots, 23\}$, where t is interpreted as the time instant ($Z(t = 0)$ indicates the temperature at 00:00, whilst $Z(t = 7)$ is the temperature at 07:00 and so on). Another example may be the total precipitation amount (measured in mm) on a large geographical region in a given day. In this case the index set X is the set of spatial locations belonging to the region and $Z(x)$ is the local precipitation amount at the specific location $x \in X$. In many real-world climate applications, we are interested not only in either spatial or temporal dynamics, but in spatio-temporal one. Think about at the precipitation amount in a large region through one year period. In such a case, the set X could be rigorously defined as the Cartesian product of a space domain (e.g. S) and a temporal one (e.g. T). As a consequence, the random field Z will be $\{Z(x, t) : x \in S, t \in T\}$.

To help the reader understanding the above-introduced concepts, Figure 2.4 shows two climate phenomena that may be seen as random fields. More specifically, Figure 2.4 (left) shows the time series of global land temperature anomalies over the period 1950-2024. In this case, $T = \{(y, m), 1950 \leq y \leq 2024, 01 \leq m \leq 12\}$, being y the year and m the month. Figure 2.4 (right) shows precipitation anomalies in May 2025 on a regular grid of $2.5° \times 2.5°$. In this case, the spatial domain is the set $S = \{(lat, long) : lat \in \{88.75°N, \ldots, 88.75°S\}, long \in \{\{88.75°E, \ldots, 88.75°W\}\}\}$. Although, in some sense, both the phenomena are deterministic, the chaotic behaviour of climate and whether, together with our incomplete knowledge of the system, force us to cope with some *uncertainty*: here the just exposed machinery definitely help. Indeed, we *model* climate and weather phenomena as random fields, to have a rigorous framework where we can measure and handle uncertainty.

2.3.3 Covariance Functions

This Subsection recalls and extends Subsection Covariance Functions of the Introduction (p. 9). A random field is called Gaussian when, for any arbitrary collection of points in the reference domain, say $x_1, \ldots, x_N$, with N arbitrary, the random vector $(Z(x_1), \ldots, Z(x_N))^\top$ has a multivariate Gaussian distribution. Gaussian random fields are completely described by their covariance function, which is a linear measure of dependence between the random variable Z at some point x_1 (or at the space-time coordinate (x_1, t_1)) and the random variable Z at some point $\boldsymbol{x}_2$ (or similarly, at the space-time point $(\boldsymbol{x}_2, t_2)$).

When we move from random variables into random fields, the covariance can be computed between any two points of the domain and, as a consequence, becomes

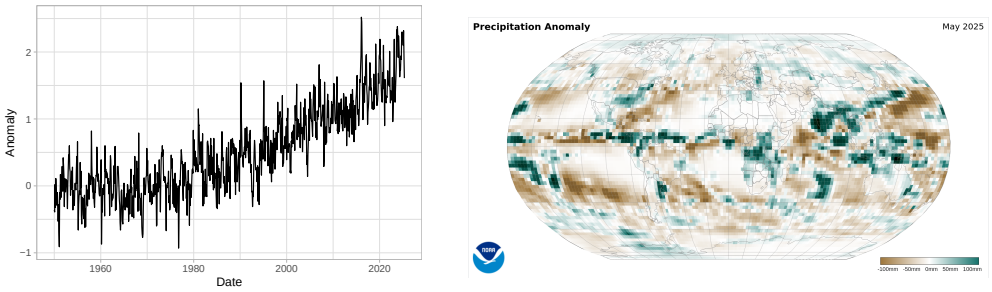


Figure 2.4: Left: monthly global land temperature anomalies since 1950. Right: precipitation anomalies in May 2025 on a $2.5° \times 2.5°$ grid. Data for the first figure and the second figure has been taken from NOAA National Centers for Environmental information, Climate at a Glance: Global Mapping.

a function. More precisely, assuming that our domain is a metric graph G times a temporal set T, the covariance function of a process $Z : G \times T \to \mathbb{R}$ is defined via:

$$C_Z((x,t),(x',t')) := \mathbb{C}\mathrm{ov}\left(Z(x,t), Z(x',t')\right),$$

where (x,t) and (x',t') belong to $G \times T$. It is mandatory to note that covariance functions need to satisfy some conditions: they must be symmetric (that is: the covariance between A and B is the same as between B and A) and positive definite. This last condition is usually particularly challenging (Berg, 2008). However, there are some standard methods to build valid covariance functions. For instance, given a spatial covariance function C_S on a domain S and a temporal one C_T on T, then the function $C_{S\times T}((x,t),(x',t')) := C_S(x,x')\,C_T(t,t')$ is a valid covariance function on $S \times T$ (and is called *separable* covariance function). Although the separability approach is very flexible, as it is possible to combine a wealth of purely spatial and purely temporal covariance functions, several studies (see Porcu et al., 2021, with the references therein) have argued that it cannot cope with the majority of spatio-temporal complex dynamics. Some work in this direction has been recently done, and next we report the main ideas.

River Networks

Perhaps the first and most intuitive example of a geophysical network is a river basin, that is: the network formed by a river, its tributaries, their tributaries and so on up to the springs. This network is, usually, a tree, meaning that given two arbitrary points on the network there exists exactly one path connecting them. This is the case since usually a stream merges with others, and never splits itself in two. Figure 2.5 shows an example of such a river basin (Porcu et al., 2023, Figure 4) and how it is possible to rigorously model it. Ver Hoef et al. (2006) have been the first to properly define valid covariance functions based on the stream distance, as several previous analyses have been performed using spherical or Euclidean distance, leading to possible non-

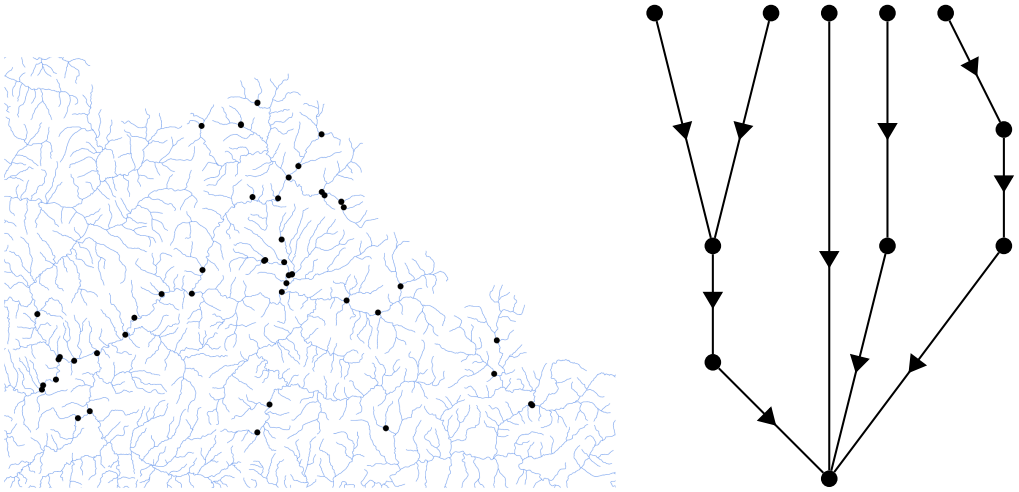


Figure 2.5: Left: Clearwater River Basin in Idaho (USA) and 50 points of interest on it. Right: how a river basin can be modelled through an (oriented) tree.

positive defined covariance functions or unrealistic models. A challenge in modelling river networks is that water *flows*, that is: two close points could share either the same water or some water, or even they could have independent streams. This makes models based on shortest-path distance unrealistic in several cases. Think, for instance, at the estimation of the concentration of some chemical molecule through the basin: the concentration of two streams that merge into another after even a short distance have *a priori* no reason to be correlated. However, the concentration of the stream below the merging point should be the (weighted by volume) mean of the concentrations of the two original streams. For this reason, Ver Hoef et al. (2006) have defined an ingenious framework that result in null covariance for points that are not *flow-connected*, that is that do not share water, and positive covariance for flow-connected points.

Their method relies on the careful integration of proper processes defined on the river network: they *define* a process in a given point x as the integral of a (weighted by water volume and by distance) white noise over the set D_x of all points whose water flows eventually through x:

$$Z(x) := \int_{D_x} g(d(y,x))\, W(y)\, \mathrm{d}y. \tag{2.1}$$

Some examples given by Ver Hoef et al. (2006) for g are: the indicator function over an interval $[0,l]$, the negative exponential $d \mapsto e^{-d}$ and the "spherical" $d \mapsto 1-d$ over $[0,1]$. This allows to achieve the desired result of independence on points that are not flow-connected and ends up in a quite simple, albeit perhaps a bit ugly,

covariance function:

$$C_S(x_1, x_2) = \begin{cases} 0 & x_1,\, x_2 \text{ not flow-connected,} \\ C_1(0) + \nu_{x_1}^2 & x_1 = x_2, \\ \prod_{k \in B_{x_1,x_2}} \sqrt{\omega_k} C_1(d(x_1, x_2)) & \text{otherwise.} \end{cases}$$

Here B_{x_1,x_2} represents the set of stream segments between the two locations x_1 and x_2, including the upstream segments but excluding the downstream locations, d is the stream distance, ν is the nugget effect and ω_k is the weight. In addition, $C_1 : \mathbb{R}_0^+ \to \mathbb{R}$ is defined as

$$C_1(h) := \int_{\mathbb{R}} g(x)\, g(x - h)\, \mathrm{d}x,$$

with $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ the moving average function in (2.1).

Static Metric Graphs

We have already introduced graphs with Euclidean edges, termed geophysical networks, but now the time has come to define processes on them, that is: define random fields continuously indexed over edges of geophysical networks. Ver Hoef et al. (2006) defined proper covariance functions on (oriented) continuously indexed trees, yet the definition of covariance functions on general geophysical networks is more recent. The two main approaches used to achieve such an objective are: isometric embeddings to define isotropic covariances and stochastic partial differential equations (SPDEs).

The first approach (Anderes et al., 2020) exploits a classical argument Schoenberg (1938) and embeddings in Hilbert spaces to show that the function $d \mapsto e^{-ad}$ is a valid covariance function whenever d is the resistance distance. As a corollary, they obtain a whole non-parametric class of functions (termed completely monotonic) that, when composed with the resistance distance, result in valid covariance functions.

Conversely, Bolin et al. (2024) and Bolin et al. (2023a) defined SPDEs whose solutions are random fields over the metric graphs. Remarkably, such an approach is able to produce random fields whose sample paths are differentiable.

Repeating the Same Metric Graph over Time

Porcu et al. (2023) defined a parametric class of non-separable spatio-temporal covariance functions that allow to combine spatial and temporal measures of distance on graphs times time, for both linear and circular time. More specifically, given p and k two positive integers, they considered two parametric classes of functions:

$$\mathcal{D}_\theta := \Big\{ \varphi(x \,|\theta), \quad x \geq 0,\ \theta \in \mathbb{R}^p \Big\} \text{ and } \mathcal{H}_\vartheta := \Big\{ \psi(t \,|\vartheta), \quad t \geq 0,\ \vartheta \in \mathbb{R}^k \Big\}, \tag{2.2}$$

where θ and ϑ are parameter vectors, and then defined

$$G_{\alpha,\beta}(x, t \,|\theta, \vartheta) = \frac{1}{\psi(t|\vartheta)^\alpha} \varphi\left(\frac{x}{\psi(t|\vartheta)^\beta} \,\middle|\, \theta \right), \qquad x, t \geq 0. \tag{2.3}$$

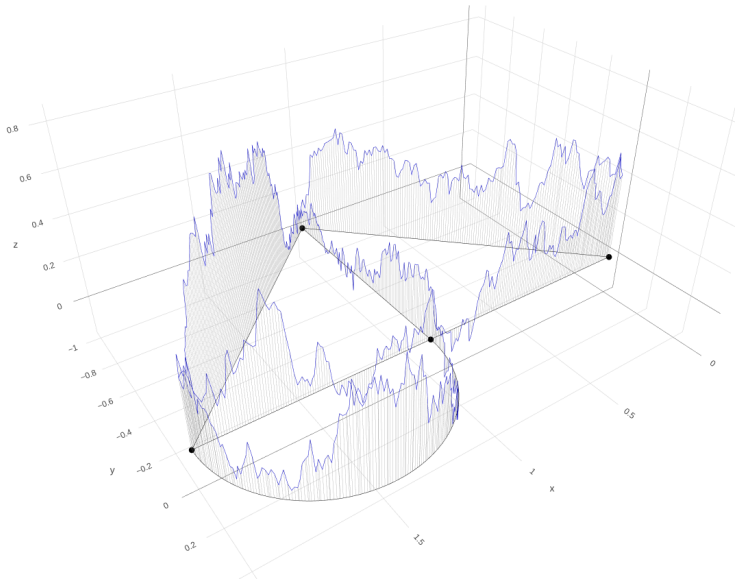


Figure 2.6: An isotropic process on a geophysical network.

Next, they provided sufficient conditions for the parametric families $\mathcal{D}_\theta$ and $\mathcal{H}_\vartheta$ so that

$$G_{\alpha,\beta}\Big(d_\cdot(x,x'), |t\text{-}t'| \,\Big|\, \theta, \vartheta\Big), \qquad x, x' \in G \; t, t' \in \mathbb{R}$$

is a valid covariance function. Here $d_\cdot$ stands for either the resistance distance or the shortest path. The parameters α and β have been left intentionally outside the vectors θ and ϑ because of their physical interpretation: when α and β are both positive, $G_{\alpha,\beta}$ corresponds to a functional form that was originally proposed by Gneiting (2002) for *space* being the d-dimensional Euclidean space. Several generalisations of this class are summarised in Porcu et al. (2021). 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 setting proposed by Porcu et al. (2023) allows for a substantial progress to describe processes that are dependent over both space and time. They work extremely well under the situation of a temporally *invariant* topology (think, for instance, of a traffic road in an old city, which is unlikely to be changed given all the constraints from building and monuments. However, there are situations where the networks are more *dynamical*. For instance, the traffic network in Abu Dhabi changes quite often as many parts of the city are under massive construction. For this situation, different strategies might be needed.

Covariances on Time-Evolving Geophysical Networks

While Porcu et al. (2023) defined a class of covariance functions on a static graph through time, Filosi et al. (2025a) (Chapter 3) formalised a mathematical framework

to deal with time-evolving geophysical networks. They allow both nodes and edges to appear, disappear or change in shape and length over both linear or circular time. To cope with this complex topology, they extend the resistance distance to account for time lag and structural changes and then define a class of covariance functions that are isotropic with respect to this new distance, extending the method provided by Anderes et al. (2020). More in detail, they prove that any function

$$C((x,t),(x',t')) := \psi\big(d((x,t),(x',t'))\big),$$

where (x,t) and (x',t') are spatio-temporal points on a time-evolving geophysical network, and ψ belong to the special class of completely monotonic functions (see, for instance, Miller and Samko, 2001, for more details), is a valid covariance function. Their methods are quite mathematical involved and it is out of the scope of this work to present them here. The main idea, however, is taken from Anderes et al. (2020) and relies on a Hilbert space embedding of their topology. To give an intuitive insight, they consider the different shapes the graph assumes at the different time instants, thus they connect the same vertices at different times, building a new spatio-temporal geophysical network where the machinery of Anderes et al. (2020) can be applied (see Figure 2.7 for a construction example and Figure 2.8 for the case of periodic graphs). Next, they construct a proper process on the resulting spatio-temporal graph, which entails both spatial and temporal dynamics, and finally define a spatio-temporal distance as the variogram of such a process.

Multivariate Covariances on Geophysical Networks

Another extension that has been explored is the definition of multivariate processes on geophysical networks, that is: in each point of the geophysical network, *many* random variables are defined. A natural example from the climate is: pressure, temperature and humidity in a given point, but other topics could have their reasons too: think about the density of different vehicles on a road network. Also this setting lends itself to the use of separable covariance functions, where one (univariate) covariance is defined over the geophysical network, while the other is defined over the set $\{1,\dots,p\}$, being p the number of variables to model.

Filosi et al. (2025b) extended the Hilbert space embedding methods of Anderes et al. (2020) to the multivariate case, providing a covariance (matrix-valued) function with the following properties:

1. the *marginal* covariances functions coincide with the original covariances of Anderes et al. (2020),

2. the multivariate process in the vertices enjoys a sparse and intuitive conditional independence structure.

In addition, they show some results useful to the construction of additional multivariate covariance functions. However, as their arguments are quite mathematical involved, we do not report them here and defer the interested reader to their paper.

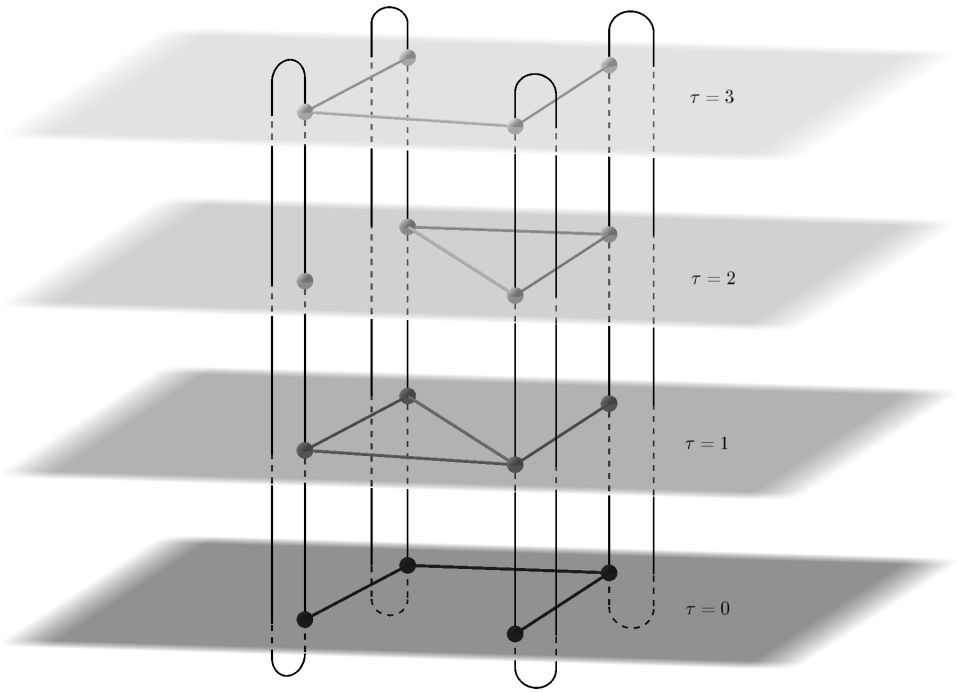


Figure 2.7: Construction of a time-evolving geophysical network that show a periodic evolution in time, with period 4. The different layers ($\tau = 0, \ldots, 3$) represent the 4 different periodic states of the network, with "horizontal" edges being physical (evolving) edges, while "vertical" connections represent additional temporal edges that account for temporal co-dependencies. Figure adapted from Filosi et al. (2025a) (see Chapter 3).

2.3.4 Impact on Climate

Geophysical networks are being very popular in climate science. Here, nodes can be chosen to represent monitoring sites, hydrological junctions, sensors, or infrastructure, while edges can represent physical or functional relations, or again flows. Ver Hoef et al. (2006) provided models for Gaussian processes that are defined over branching stream networks using "tail-up" and "tail-down" kernels. This work was then generalised through SSNbayes and spatial indexing frameworks in Santos-Fernandez et al. (2022); Ver Hoef et al. (2023). Applications to river networks that are repeatedly observed over time are provided by Porcu et al. (2023). Sun et al. (2022) developed physics-driven GNN to mimic certain hydrological processes. Moshe et al. (2020) introduced the so-called *Hydronets* to improve run-off prediction while using the rivers peculiar topology.

In the context of urban flood forecasting, Kazadi et al. (2024) provided *FloodGNN* and extended the previous work by Roudbari et al. (2024). Water quality networks belong to this type of predetermined networks as well, and we mention Sit et al. (2021) who proposed graph-convolutional GRUs to forecast short-term streamflow.

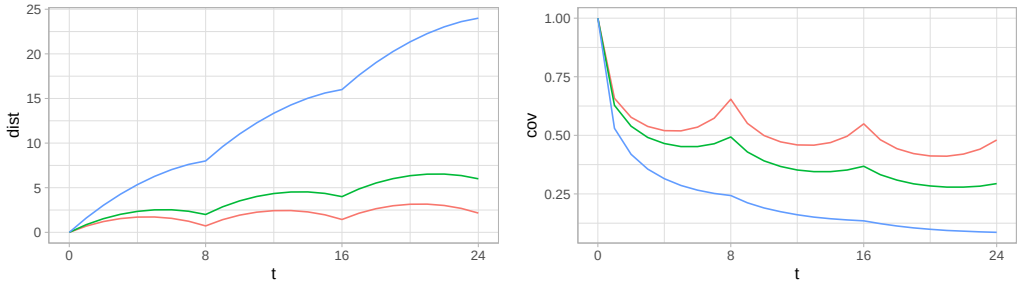


Figure 2.8: Example of distance functions between two vertices of (left) and the induced covariances of a process defined *over* (right) a periodic geophysical network of period 8 look like. The three different lines represent possible choices given by free parameters. See Filosi et al. (2025a) (Chapter 3) for more details.

GNN have also been used within this framework, and we refer to Bai and Tahmasebi (2022) and more recently to Wan et al. (2025).

Urban climate networks are, most of the time, of the geophysical type. Recently, Zanfei et al. (2022) have used these networks to prompt GNN-based temperature prediction in urban networks. All these examples are apparently compatible with the classical geostatistical framework, see for instance Ver Hoef et al. (2023). Thorough works that account for the river topology in concert with the choice of the right metrics can be found in Ver Hoef et al. (2023).

For a comparison between traditional methods and GNN-based approaches in terms of predictive performance show non-uniform results, and the reader is referred to Sun et al. (2022); Kazadi et al. (2024); Moshe et al. (2020). Transformer-based architectures (FloodGTN) use edges weighted by tide and/or dam release, showing considerable computational time gains (Shi et al., 2023).

We see applications to climate within geophysical network as four main groups:

1. Hydrologic Forecasting and River Basin Modelling: Porcu et al. (2023), Sun et al. (2022) and Moshe et al. (2020) are clear applications in this direction, in concert with Taghizadeh et al. (2025).

2. Floods and Urban Hydraulics: Kazadi et al. (2024), Roudbari et al. (2024), and Roudbari et al. (2024).

3. Water quality: Zanfei et al. (2022), and Suri and Mangal (2025).

4. Urban climate: Yu et al. (2024), Chishtie et al. (2024).

2.4 Networks for Climate Data

2.4.1 Why *for?*

This section treats networks *for* climate data as tools, rather than methodological proposals as been introduced in earlier literature. We shall avoid mathematical obfuscation, while specific maths can be retrieved from each of the references that we have compiled hereafter. Networks as tools can be very useful to understand who interacts with whom and at what scale. Further, tools that are typically coming from ML literature allow for qualitative assessment regarding the type of information under process. In this respect, networks *for* substantially differ from Tsonis and geophysical networks, as they are typically used to forecast and nowcast fields in space and time (Lam et al., 2023), downscale coarse fields to local (finer resolutions) (Vandal et al., 2017), *denoise* irregular observations (Shuman et al., 2013), analyse extreme events distribution (Boers et al., 2019) and to assimilate heterogeneous observations into models (Evensen, 2009).

Some comments are in order. The example above showcases the usefulness of graphs *for* as pragmatic solutions to certain operational pipelines (Shuman et al., 2013; Ortega et al., 2018; Kivelä et al., 2014). Moreover, the examples cited hereafter are considered as network *for* for three main reasons:

- The graph is instrumental to solving a task (forecast, downscale, fill gaps), but not to claim that the edges reveal true teleconnections or a unique physical topology (Shuman et al., 2013; Ortega et al., 2018).
- Physics is used as a design prior (e.g., jet streams, drainage direction). However, edges and weights can be rather simplified or adapted for robustness and speed computation (e.g., k-nearest neighbour on the sphere, bipartite coarse→fine) (Defferrard et al., 2016).
- Evaluation is task-centric, as the focus is on whether the graph helps prediction, reconstruction, or assimilation under regime shifts and sparsity, rather than whether its community structure or degree distribution looks the right one (Rasp et al., 2020).

As an opening to the descriptions following below, we stress here that there are notable differences between Tsonis networks, where edges are discovered and describe teleconnections, and networks *for*, where edges are chosen to route information needed to solve a given task; any resemblance to true teleconnections is incidental to the goal of better predictions or reconstructions (Shuman et al., 2013; Ortega et al., 2018). Additionally, we stress that in networks *for* the topologies can be borrowed as a tool to solve a task (e.g., flood nowcasting). However, the topologies can be modified in whatever shape that can help to improve performance (Kivelä et al., 2014). This is a fundamental difference with respect to geophysical networks.

As a result, the graph is the key design choice for this section. The following recipes have been widely used within climate sciences.

1. The simplest construction is that of grid or *mesh graph*, where one node is assigned to each grid cell and edges are chosen from the nearest geographic neighbours. Edge weights are normally assigned through inverse distance, while basic attributes such as a land–sea mask or orography provide essential context. This basic construction has nice features, such as ease of computation, and it serves as a baseline for interpolation, denoising, and forecasting tasks (Shuman et al., 2013; Ortega et al., 2018; Rasp et al., 2020).

2. Another popular architecture is that of k-nearest neighbours on the sphere. (Defferrard et al., 2016; Shuman et al., 2013). For global models (related to the whole surface of the sphere representing our planet), connecting each node to its k closest neighbours by great-circle distance offers a stable (and geometry-aware) alternative to fixed stencil grids. The value of k controls the local properties (neighbour structure), and the graph can be easily adapted to irregular sampling or more complicated geometries (for instance, coastlines).

3. In some cases, the graph structure is enriched through the addition of a small number of edges that reflect known pathways—advection corridors, such as jet streams or prevailing winds (Ver Hoef et al., 2006; Kivelä et al., 2014). These links can be binary or weighted, and serve as informed priors.

4. Special graphs architectures can be used to downscaling. In this case, coarse fields and fine targets (e.g., station or high-resolution grids) are represented as two node sets, and then each coarse node is connected to the fine nodes it influences. This technique allows to improve resolution, allows to mitigate computational burden, and keeps the mapping interpretable and modular (Vandal et al., 2019; Sarkar et al., 2020).

5. When interactions are related to variables or regions, it might be convenient to split the variables into layers, which are connected through inter-layer links (Kivelä et al., 2014). This preserves within-layer geometry while enabling interaction control.

6. The edge set might be allowed to vary over time (for instance, across seasons or ENSO phases). For a recent treatment of this case, the reader is referred to Filosi et al. (2025a) (see Chapter 3) as well as an earlier contribution by Runge et al. (2019b).

2.4.2 Models

This section discusses three families of models that have been especially influential, as networks *for*, within the realm of climate data science.

Graph Signal Processing (GSP). This family of models treats climate fields as a signal defined on the nodes of a graph: values live at grid cells or stations, and edges

explain *who is near whom*, where *near* does not necessarily mean geographically. Within this setup, an important role is played by graph filters, which are linear transformations used to smooth noisy observations, and to interpolate in space-time in the presence of missing data. Regarding interpolation and denoising, applying a low-pass graph filter allows to mitigate small-scale noise while keeping coherent structures (e.g., coastal gradients or orographic bands). See for instance, Hammond et al. (2011) as well as Wang et al. (2016). Another important role of this model is played within the so-called graph-based sampling, which aims (through the same machinery) to quantify uncertainty after filtering. This amounts to evaluate nodes that might be good candidates for new sensors, which makes GSP directly actionable for network design and monitoring expansion (Shuman et al., 2013; Ortega et al., 2018; Anis et al., 2016). Finally, GSP turns to be useful for graph design purposes, rather than taking structure and topology as granted. The structure of the graph can be learned in such a way to smooth the observed signals while keeping the original data structure in terms of plausible connections (Shuman et al., 2013; Ortega et al., 2018). As a result, GSP provides a reliable baseline, and a potential prior (a building block) for more sophisticated models.

GSP are not Tsonis networks. To dissipate potential confusions, we note that in Tsonis networks the graph itself is the empirical result (edges represent statistical teleconnections inferred from data). In GSP, the graph is the computational tool (given or learned as a prior) on which the climate signal is processed. Table 2.3 helps for a better understanding of their differences.

	Tsonis networks	**GSP frameworks**
Graph origin	Inferred from statistical similarity.	Predefined or learned as smoothness prior.
Role of graph	Object of study (teleconnections).	Operator or constraint for signal processing.
Main tasks	Diagnostics, cluster detection.	Filtering, interpolation, UQ.
Output	Network metrics.	Reconstructed fields and design recommendations.

Table 2.3: Tsonis vs. GSP: roles of the graph.

Probabilistic Graphical Models (PGMs). A complementary view of climate network connections can be achieved by looking at conditional probabilities, which allows for alternative interpretation while providing the practical computational advantage of having sparse structures, which in turn allow for scalable inference as well as for uncertainty quantification (Lauritzen, 1996). Within the realm of Gaussian models, sparsity regards the zeros of the precision matrix (the inverse of

the covariance matrix), and the non-zeros can be interpreted as conditional links between the nodes (Friedman et al., 2008; Yuan and Lin, 2007; Lauritzen, 1996).

PGMs provide a principled language for spatio-temporal structure. Dynamic Bayesian networks and related state-space formulations allow to specify lagged correlations and multi-step dependencies. Finally, PGMs represent a natural class of models for causal discovery (de Prado, 2023; Runge et al., 2019b; Shimizu et al., 2006). No procedure yields causal truth from data alone, but PGMs at least separate the questions—*is there a conditional link? which way might it point? what assumptions did we use?*—and provide uncertainty around each step.

As previously explained, PGM can sit as an intermediate layer between GSP and GNN. The smoothness prior inherited from GSP can be used as a candidate graph architecture, which can then be refined (in terms of conditional relations) through PGM, which also provides measures of uncertainty quantification. In turn, these uncertainty measures can be embedded as *weights* in the loss of a GNN (on which we do not anticipate too much here). Standard references are Lauritzen (1996); Friedman et al. (2008); Yuan and Lin (2007); Runge et al. (2019b); Shimizu et al. (2006).

Graph Neural Networks (GNNs). The third class of models learns how information should move along edges to solve a task (forecasting a field, sharpening a coarse map, or estimating a risk index for instance). GNNs are especially tailored to illustrate non-linear relationships. Within the GNNs framework, several subfamilies have been tailored to solve specific tasks: (i) Spectral-GNN learns filters that can be interpreted as flexible versions of the GSP filters. For instance, Chebyshev-based convolutions and the vanilla GCN are strong baselines with few moving parts (Defferrard et al., 2016). (ii) Attention models identify dynamical connection between nodes, and are used within teleconnected regimes where the important neighbour might be far away or change with the season (Veličković et al., 2018). (iii) Space-time models combine graph propagation with recurrence or transformers. Within this family, diffusion-convolutional RNNs and ST-GCNs are reliable choices for sequence-to-sequence forecasting, while graph transformers scale well when histories are long and exogenous drivers are many (Wu et al., 2021).

GNNs benefit from GSP and PGMs at least for three reasons. First, a GSP filter allows for pre-processing and reduces noise before the GNN is applied (ensuring robustness) (Shuman et al., 2013; Ortega et al., 2018). Second, a PGM conditional network can replace a naive kNN (k-nearest neighbour) graph, focusing on edges with statistical support (Friedman et al., 2008; Yuan and Lin, 2007). Finally, uncertainties from the PGM (or from an ensemble of GSP baselines) can be used as loss weights, so that the model allows for putting emphasis on regions having, for instance, sparse data, or extreme events.

Each of these families has practical advantages, but also some caveats. GSP is practical and easy to implement; however, it is limited to linear relationships.

Task family	Graph recipe	Models
Global and regional forecasting	kNN on great-circle distance; add a few physics-aware links; multilayer across variables or levels	Spatio-temporal GNNs (GCN/attention) with temporal GRU/transformer heads; robust to missing data.
Downscaling and super-resolution	Bipartite coarse→fine graph with distance, orography and land-cover attributes	Graph U-Nets and cross-scale attention; preserve sharp fronts and orographic effects (Vandal et al., 2019; Sarkar et al., 2020).
Imputation and gap-filling	Sensor or mesh graph; edges from rolling similarity and distance	Combine a simple graphical model prior with a GNN denoiser; train with masked losses (Friedman et al., 2008).
Extremes and compound events	Event graph built from threshold exceedances; edges capture lag or tail dependence	Detect backbones and hubs during heatwaves and atmospheric rivers (Boers et al., 2019).
Hydrology and floods	Directed river network; weights informed by slope and travel time	Physics-informed GNNs; Graph Transformers with conservation penalties (Ver Hoef et al., 2006).
Urban microclimate	Heterogeneous graph over roads, buildings, and sensors	Heterogeneous GAT; mitigation scenario testing (Chishtie et al., 2024).
Data assimilation	Grid graph with observation–to–state links	Use graph-aware covariances inside EnKF/4D-Var; keep physics in the analysis step (Bocquet and Sakov, 2013).

Table 2.4: Graph design patterns for common networks *for* tasks.

PGM provides sparse and scalable structures, but heavily relies on the modelling choices (especially on the assumption of Gaussianity). Finally, GNNs represent the most flexible choice in terms of portfolio of linear relationships, but require careful regularisation, monitoring for drift, and post-hoc explanations; they are also *hungry* for consistent training data (Wu et al., 2021). Used together, the three families cover the spectrum from simple and stable to rich and adaptive while preserving a clean path for uncertainty and governance. Table 2.4 illustrates the use of this methods, according to the assigned task, within the realm of climate models.

The constraints that are proper of each of these three families should not alter the physics properties. A couple of principles are listed below.

1. When predicting fields over a geophysical network, constraints should preserve certain structures. On rivers, *inflow minus outflow equals storage change*, and a similar principle applies to atmospheric budgets (Bocquet and Sakov, 2013).

2. The physics can be related not only to the geophysical network, but also to the data structure. Some data must be non-negative (precipitation, run-off) or live in sensible ranges (relative humidity, wind speed, see Bocquet and Sakov, 2013). Another relevant situation is related to the presence of coastlines, boundaries and barriers, and the use of weights should reflect smoothness in the interior while allowing for sharp transitions in persistent fronts. This principle has been analysed for both graph modelling and downscaling, and the reader is referred to Wang et al. (2016); Shuman et al. (2013); Ortega et al. (2018) for the former, and to Vandal et al. (2019) for the latter.

Uncertainty quantification plays an important role in networks *for*. Within a huge literature, we mention here (without being exhaustive) deep ensembles (Lakshminarayanan et al., 2017) and scenario-aware checks, for which it is verified when and where uncertainty widens in unfamiliar regimes (e.g., unusual ENSO phase).

2.5 Bridges *within* the Triad

2.5.1 Discrete *versus* Continuum Hypotheses

An intuitive objection against the approaches presented in this Chapter is represented by their potential redundancies in front of the well-established continuous space–time statistical models (Porcu et al., 2024, 2021). The review by Porcu et al. (2021) provides a thorough account of continuous space-time models, with special emphasis on the Gaussian case, starting from models on the sphere into spectral models and models based on SPDEs (Lindgren et al., 2011). This rich literature has provided a wealth of contributions to climate modelling in terms of interpolation, prediction, uncertainty quantification, and physically-informed inference across the planet's surface (Castruccio and Guinness, 2017). An understandable criticism is oriented towards the fact that Tsonis networks lack theoretical foundations, and introduce discretisation techniques that might be easily avoided through the use of continuous models. We do not necessarily agree with this statement. In particular, we claim the following.

Approaches are Complementary. While space-time models are generative, Tsonis networks are inductive. Hence, they do not assume that the data come from a pre-defined stochastic process (for instance, a Gaussian process). From a epistemological viewpoint, both approaches are needed: the first explains, the second detects.

Abrupt Changes. Continuous models are at present struggling to detect abrupt changes, that are structural transitions and non-stationarities. Typical smoothness assumptions can actually mask rapid changes and restructuring in climate dynamics. On the other hand, networks are excellent in detecting transitions. These topological shifts are interpretable as early warning signals and are less visible in classical geostatistical outputs.

Resilience. Space-time models require a sufficiently dense data coverage and are normally based on strong assumptions (*e.g.*, stationarity). However, for most of the planet the available observations are very sparse. Climate networks are more resilient to such limitations, as they include statistically significant links only. Further, they enable inference even in irregularly-sampled domains.

Policy Oriented. This aspect will be expanded in the subsequent section. Governance decisions often require communication between technical experts, decision-makers, and the public. Narratives centred on "central nodes", "broken links", "fragmentation", "resilience" are naturally accepted by policy and decision makers.

We do not advocate abandoning continuous models, but rather integrating them with network approaches. Graphs can inform the design of spatial priors, guide the identification of anisotropies, and contribute to kernel shaping or nonstationary extensions. Likewise, continuous models can be used to simulate or stabilise network estimation under uncertainty. Hybrid frameworks, such as multi-layer graphs or coupled GNN-GP architectures, promise rich avenues for uniting statistical power with topological insight.

2.5.2 Different Worlds

We start with an epistemic discussion that concerns the possible types of networks opened by the bridging *within* philosophy proposed herein. We consider here a combination of some Tsonis networks as well as scale networks, with the geophysical networks. Table 2.5 gives a reliable picture.

Network Type	**Inferred from Data**	**Given by Domain Knowledge**
Random-like	Correlation networks at low thresholds (Tsonis et al., 2006).	Sensor networks or monitoring grids without structure.
Small world	Mutual information networks (Donges et al., 2009b).	River networks with man-made channels (Muhammad, 2016).
Scale free	ENSO-related hubs in dynamic networks (Tsonis and Swanson, 2008).	Electric or road networks in urban heat island studies (Schäfer et al., 2018).

Table 2.5: Framework: Topology × Construction Mode.

We provide a description of the *interesting* ones as those networks stemming from the following combinations:

1. Random Networks can have a predetermined structure. For instance, sensor networks for which the nodes have been established without invoking any precise organisation principles. Typical situations are those of climate or environmental monitoring stations (Mearns et al., 2003).

2. Small world networks can also have a predetermined topological structure. Hydrological networks (Ver Hoef et al., 2006), in particular those having engineered channels, can exhibit the characteristics of a small world network.

   Some geophysical networks can be scale-free. For instance, urban infrastructure or road systems often follow scale-free distributions due to cost-efficiency and resilience trade-offs. Their climate relevance lies in vulnerability analysis under extreme weather events (Schäfer et al., 2018).

This framework allows researchers to integrate domain-specific knowledge with statistical inference, facilitating interdisciplinary approaches and hybrid models that respect both empirical data and physical constraints.

2.5.3 Technical Bridges

This section provides some ideas for a transition from separate paradigms into potential fusions. Providing bridges is not a mere intellectual exercise, but opens for a much wider discussion that has been in the core of data science in the last years: should data science be based on *data-first* or *model-first* approaches?

Graph-Signal Reweighting of Tsonis Networks

We have already discussed the intrinsic drawbacks in providing binary large scale connections through Tsonis networks. The recent work by Haas et al. (2023) provides an excellent insight into the statistical limitations of such an approach. It might be tempting to integrate the philosophy of signal processing inside the Tsonis architecture — hence, graph signal processing (Shuman et al., 2013; Perraudin and Vandergheynst, 2017) might represent an interesting tool for integration. Let $G = (V, E, W)$ denote the Tsonis network, where V represents nodes (spatial locations), E the edges (pairs with indicators of statistical similarity being above a given threshold), and where W the adjacency weight matrix. We consider the *climate signal* as the process defined over the nodes at different time instants. The idea is (a) implement a Tsonis network, (b) consider the process defined over the Tsonis network as a graph signal, and (c) update the weighting matrix W with a matrix $\widetilde{W}$ based on graph signal similarity. (As an alternative, one might use graph-based smoothing through a graph Laplacian regularisation approach.) The optimised weights for $\widetilde{W}$ can be iteratively adapted to minimise temporal inconsistency, providing a data-driven refinement of the edge structure (Thanou et al., 2017; Segarra et al., 2017). This operation can be evaluated *locally* by embedding, *e.g.*, a urban network into the pre-existing Tsonis network. Then, the restriction to a subgraph corresponding to the Tsonis graph at the region where the urban network is evaluated, one can create hybrid architectures.

Finally, the re-weighted network can be analysed via the graph Fourier transform to identify coherent modes of variability (Shuman et al., 2013; Mendes et al., 2014). This may reveal latent mesoscale phenomena not observable in the original Tsonis formulation. This integration of dynamic graph signal characteristics with climatological correlation networks provides a promising path to unify the "of" and "over" paradigms.

Local Embedding of Metric Graphs into Climate Teleconnections

Spatially-constrained metric graphs can be embedded into abstract teleconnection networks. We shall avoid mathematical obfuscation and will instead focus on the ideas for such embedding strategies. We start with a global graph G_{Glob} of the Tsonis type,

and with a local metric graph G_{Loc} equipped with some metric d. To embed G_{Loc} into G_{Glob}, we follow the steps. (a) Assign geographical coordinates to the nodes of G_{Loc}. This can be done through remote sensing, for instance. Then, each node is assigned to a single node or a cell, within G_{Glob}. (b) Downscale the climate signals available at G_{Glob} into G_{Loc}, for instance through kriging or through graph based filtering (Perraudin and Vandergheynst, 2017; Dong et al., 2020). (c) Decide on modelling strategies, where the two graphs might be treated as nested, or as multilayers. Cross-scale interactions can be modelled using the available techniques such as graph alignment or multi-resolution techniques (Kivelä et al., 2014; De Domenico et al., 2013). A typical examples is that of urban temperatures (heat mitigation), where the micro-drivers, such as ENSO or Indian monsoon, are located over a global Tsonis network. These approaches are very promising to perform local inference, as well as to propagate information between two different scales (Kipf and Welling, 2017; Bronstein et al., 2021).

Climate-Informed Graph Topology Learning

Previous approaches where based on mutual embedding. Here, we suggest some procedures to *learn* the graph topologies starting from climate data. This idea has been partially pursued by Dong et al. (2020) within the realm of graph topology learning. Adapted to our context, this is about inferring topology of networks in such a way that these reflect the dynamical dependencies while respecting the geographical constraints. This class of approaches is typically based on finding a graph Laplacian, or and adjacency matrix, such that the signal is smooth over the graph, constrained on some functional minimisation problem. One might object that such approaches should be strictly connected to the physics nature of climate data. This might be taken into account by, *e.g.*, penalising connections that contradict known geophysical distances (this can be implicit in the type of kernel that has been used). Bayesian approaches might do a good job through the design of priors that incorporate causality structures. Finally, multivariate based-process techniques might add information on the existence of edges. As a practical example, consider performing inference of a regional graph over the Middle East, using precipitation and temperature data from satellite and reanalysis products.

These bridges motivate a unifying formalism involving multi-layer networks Boccaletti et al. (2014) or more sophisticated form of graph associations (De Domenico et al., 2013; Berlingerio et al., 2011).

2.5.4 Epistemic and Governance-based Bridging *within*

We have analysed the possibility of having hybrid systems that integrate dynamical coherence with georeferenced constraints of the climate system. Such an integration can be viewed from the angle of climate governance, as much as we have done for the case of Tsonis networks.

When Tsonis-style empirical graphs are aligned with predetermined network geometries, they produce hybrid structures that embed observed dynamical dependencies within structural or jurisdictional boundaries. Hence, such an integration will improve capabilities in terms of prediction accuracy, opening the door to multi-level decision making.

In view of networks integration, early warning systems can be improved by enabling the detection of teleconnections under the presence of physical constraints. For example, a Tsonis-derived correlation between upstream precipitation in Region A and downstream flood occurrence in Region B can be embedded onto a river basin network and improve risk predictability. This would be coherent with the purported anticipatory protocols suggested by both national and international agencies.

Hybrid networks can become successful to integrate information within different administrative boundaries (*e.g.*, the provinces in Spain). This fact can activate trans-boundary governance agreements.

Within the realm of urban planning, the identification of climate-derived teleconnections and their embedding on urban networks may allow municipalities to identify the so-called *compound hazards* (for instance, simultaneous blackouts) that may happen within correlated neighbourhoods.

We elaborate a bit more within these directions, wih emphasis on increasingly polycentric and data-driven decision-making frameworks.

From Diagnostic to Design-Oriented Governance. Network models that integrate Tsonis-style empirical structures with jurisdictional and geophysical constraints align naturally with emerging paradigms of anticipatory and risk-informed governance. The integration allows to detect both teleconnections and cascading risks and in turn prompt proactive attitude to enable scenario planning that respects both physical and institutional boundaries (Ostrom, 2017).

Climate Finance, Risk Pools, and Sovereign Instruments. Fusion-based networks can inform climate-indexed financial instruments, including catastrophe bonds and sovereign risk pools. For example, parametric insurance models have been proposed on the basis of metrics (based on networks) of teleconnected droughts or flood propagation patterns (Hellmuth et al., 2009). The African Risk Capacity (Clarke and Hill, 2013) and the Caribbean Catastrophe Risk Insurance Facility (Ghesquiere et al., 2006) have both underscored the need for hazard models that reflect structural linkages and localised vulnerability.

Calibration within International Frameworks. International connections between different policies and paradigms can be favoured by networks integration. The Sendai Framework for Disaster Risk Reduction is centred on improving risk governance through multi-hazard early warning systems and disaggregated risk metrics (Kelman, 2015). Hybrid networks can favour the embedding of risk metrics within physical and administrative systems. Analogously, the Task Force on Climate-

related Financial Disclosures strongly favours explicit metrics of spatial exposure and systemic risk, which can be enriched by network-based reasoning (Board, 2017).

Polycentric Adaptation. Climate adaptation increasingly occurs within decentralised, overlapping jurisdictions. Multilevel coordination and improvement of resource allocation, as well as accountability, can be all improved by embedding empirical dependencies into networks that align with local governance geometries (Jordan et al., 2018; Ostrom, 2017).

We do recommend a development of governance in accordance with the integration of the two networks, especially through multi-layer governance approaches, and through the use of methods that will allow to redefine exposure and vulnerability metrics across multiple governance units.

2.6 The Shaw-Stevens Dialogue and the Fusion *Between*

This section proposes a dialogue between the Shaw-Stevens agenda and the network triad. We start by recognising the different genealogies of the two paradigms - the former being largely inspired by the philosophy of climate science, and the latter substantially coming from data-centric and computational frameworks.

However, they share dissatisfaction for LSD, and this fact is made explicit within the Shaw-Stevens discourse while it reads between the lines of the triad proposal in this Chapter. This section proceeds by outlining the contrasts between the two paradigms, to then open for a reconciliation that will allow for a transition to what we term, throughout, a fusion *between*.

2.6.1 Contrasts and Synchronies *Between*

According to the LSD framework, the Earth system is represented on a latitude-longitude grid. Sub-grid processes are parametrised in terms of grid variables, and local properties are defined with respect to the Euclidean metric. Hence, neighbour structures are determined by those nodes that lie next to one another on this discretised (ad flat) version of the globe. In the network triad, adjacency is not necessarily given or determined through a simplified version of geographical proximity. For Tsonis networks, proximity is relational. Hence, there is a clear ontological shift - from a uniform, rectilinear grid to a flexible graph whose geometry adapts to the processes under study (Donges et al., 2009a; Boers et al., 2019). Distance ceases to be geographical and becomes contextual for Tsonis networks. It is geographical for geophysical network, but the metric is carefully chosen according to the technical developments illustrated in previous sections.

There is an apparent epistemic divergence as well. LSD is generative, while Tsonis networks are inductive. Within geophysical networks, there is certainly space for generative approaches, while networks *for* can be of any type (generative, predictive, inductive).

Another potential discrepancy stands with the fact that Shaw and Stevens (2023) criticiseML as disruptive but opaque. Our approach embraces ML but mitigates it with interpretable layers. These distinctions are summarised in Table 2.6. Beneath

Dimension	**Standard Approach (LSD)**	**Network Triad (of/over/for)**
Ontology of space	Gridded, Locally Euclidean	Relational, geodesic, anisotropic
Epistemology	Generative, dynamical closure	Inductive, detective, relational diagnostics
Role of ML	Disruptive but opaque	Physics-informed, interpretable via graphs
Evidence mode	Hypotheses via canonical models	Hypotheses via network statistics and edge tests

Table 2.6: Contrasting the "Standard Approach" of Shaw–Stevens (LSD) with the Network Triad.

these contrasts lie profound synchronies. Shaw and Stevens call for hierarchies — that is, from simple to complex ladders of models, mechanism-denial experiments, and strategic nudging of dynamics. The network triad provides a natural tool for such hierarchies. In fact, Tsonis networks can be used as a baseline, the geophysical as an intermediate layer (through domain knowledge embedding), and networks *for* climate data represent the upper layer, where algorithms exploit graph structure to perform forecasting, assimilation, or downscaling.

There is an apparent convergence between the two paradigm in the way discrepancies are embraced. Shaw and Stevens (2023) urge climate scientists to embrace mismatches (Southern Ocean cooling paradox, Walker-circulation slowdown, or terrestrial humidity declines) as an indispensable source of data discovery rather than *noise* to be eliminated from the trend. Within the network triad, discrepancies come as measurable shifts in graph statistics: changes in community structure, degree distribution, or causal orientation.

As a final remark, Shaw and Stevens call for disruptive computation (very large ensembles, kilometre/scale models, ML) which is in apparent synergy with our network agenda. Each direction proposed in this Chapter implies a computational disruption, but we also insist on keeping interpretable structures. In this sense, both programs seek acceleration without opacity.

2.6.2 The Fusion *Between*

Our narrative is pointing to the fusion *between* as a natural step forward. While Shaw-Stevens provides the vision and mission of climate science for the years to come, the networks triad provides the concrete artillery to realise that vision and mission.

The natural fusion discussed here is illustrated below through some examples.

Hierarchies are indispensable for identifying the limits of LSD, and the triad provides very good tools for such a task. In particular, through this lens, LSD becomes a hypothesis that can be explicitly tested within a relational framework.

Shaw and Stevens recommend that discrepancies should drive observational design rather than being smoothed away. Network formulations make this actionable: uncertainty fields can be defined on nodes and edges, informing adaptive sampling strategies that target regions of maximal epistemic tension. Table 2.7 illustrates these connections through practical examples.

Discrepancy	**Standard Approach (LSD view)**	**Network Remedy**
Walker-circulation slowdown	Attributed to large-scale ocean–atmosphere coupling; parameter uncertainty	Causal graphs orient edges; PCMCI and LiNGAM identify mediators and exclude passengers
Land-humidity decline	Expected closure with temperature; models diverge regionally	Graphs over basins capture soil–atmosphere coupling; sensor placement reduces uncertainty
Southern Ocean cooling	Models struggle with mixed-layer depth and fluxes	Graph geodesics along isopycnals refine assimilation; targeted observations via community detection
Storm-track biases	Parametrised small-scale eddies in GCMs	Cross-scale bipartite graphs conserve fluxes while preserving coastal/orographic bands

Table 2.7: Examples of regional discrepancies and corresponding network remedies.

The fusion extends beyond science into governance. Networks provide the metrics that can be used to compare different ensembles and then to implement validation, communication and decision making.

In conclusion, the two agendas are not antagonistic but complementary.

2.7 The Meta Fusion: the Shaw-Steven Network Ecosystem

The intellectual trajectory of this paper culminates in what may be called a *fusion within the fusion*. We have explored the inner fusion, termed *within*, and built a unique framework for empirical connectivity, algorithmic complexity, and spatial

geometry & topology. We have also bridged this triad with the epistemological agenda proposed by Shaw and Stevens (2025), whose essay on the "Other Climate Crisis" questions the dominating paradigm of LSD. There is way more than the illustrated parallelism between the two frameworks, and finding suitable connections requires resorting to philosophical arguments as a baseline for scientific thinking. We argue below that the network triad is not just the tool for the implementation of the Shaw-Stevens agenda, but a *sine qua non* element for what we term a Shaw-Stevens network ecosystem, that is, this section argues that a meta fusion between the two frameworks creates a system where epistemology and methodology co-produce one another. Our argument is supported by the following reasoning.

2.7.1 Reflexive Closure

We start by arguing that the network triad and the Shaw-Stevens agenda, while different in their ontological and epistemic baselines, converge through their interaction. Specifically, we argue that their interaction reveals a deeper principle: *the architecture of knowledge in climate science is itself a network*. By this principle, we mean that the architecture of knowledge in climate science behaves under the same logic of self-organisation and feedback that governs the climate system itself, where the data universe, information, models and discoveries become a network having feedback loops that allow for the network to dynamically re-organise itself. Let us carefully justify our assertion.

A first alignment comes from the fact that at the methodological level, the triad provides a grammar of relations. Tsonis networks capture empirical coherence; networks *for* translate this coherence in tools for inference, and geophysical networks position the previous one in a geometric continuum, connecting statistics and computation to the topologies of our planet. This first alignment provides duality for each component of the triad, where each of them is boosted by the others while being at the same time a condition for their coherence.

A second remark regards alignments within the Shaw-Stevens agenda. Reparametrising the *small through the large* is not necessarily a good idea anymore. It is instead time for a reflexive science, where learning comes from interaction of scales rather than from their separation. The network triad is a natural grammatical integration within this narrative, as the triad network is the perfect candidate for multiscale reasoning. Hence, our second alignment illustrates how the network triad becomes a tool for verifying the Shaw–Stevens hypothesis in practice.

We culminate this by invoking what we term a *reflexive closure*, which is substantially the meta fusion emerging from the interaction of the two previous alignments. Hence, the network triad becomes a philosophical baseline that goes beyond providing tools for the Shaw-Stevens program. At this stage, the construction of networks becomes self-validating in the sense that it provides its own epistemic tests by revealing when the statistical relations among data confirm or contradict the modelling assumptions that motivated them. Conversely, the Shaw-Stevens framework, once an abstract critique of scientific culture, becomes operational. In particular, it informs

choices of threshold, regularisation, and hierarchical design in network modelling. We clarify that reflexivity is achieved when a system has categories, classes and variables that are inseparable but keep their distinct identities.

2.7.2 Morphology of the Meta Fusion

By morphology we mean the architecture stemming from the integration of methods and epistemology within the meta fusion. While reflexive closure invokes hierarchies to then build a circular system, this section insists on the circulation part to argue in favour of an ecosystem that dynamically reorganises itself.

A simple way to look into this is through the following layers:

1. Internal coherence within the triad, which is the basis for the invoked fusion *within*. Internal coherence is the condition for internal closure as previously advocated;

2. External alignment with epistemology, which allows the triad to transcend instrumentalism;

3. Reflexive closure to determine autopoiesis, that is the system becomes capable of conditions of its own validity.

A simplified version of the morphology above is provided through Table 2.8. Horizontal reading of the table illustrates the integration of knowledge layers: from technical synergy, to epistemic alignment, to reflexive autonomy. The vertical axis instead describes the intensity of engagement between the three levels. Taken together, the three columns and three rows define a simple scheme that describes the morphology of the meta fusion we have in mind. Figure 2.9 provides a geometric simplification for this closure.

2.7.3 Autopoiesis of Climate Knowledge

The culmination of the invoked fusion is an image of climate science that is autopoietic — that is, a system capable of producing and renewing its own conditions of possibility. We borrow this term from the exemplary theory from Maturana and Varela's (Maturana and Varela, 1980) of living systems. For the authors, an autopoietic system is capable of maintaining itself through the continuous reproduction of its components.

Our claim of autopoiesis of this fusion is not exceptional by itself, as it is consistent with climate system, being also autopoietic. In facts, climate systems reproduce its statistical regularities through the continual circulation of matter and energy. Further, climate systems achieve stability not by eliminating disequilibrium but by continually converting gradients and perturbations into patterns of renewal-circulations that restore balance through motion.

Accordingly, our approach claims for a construction of an autopoietic climate science that transforms uncertainty into insight, and insight back into uncertainty.

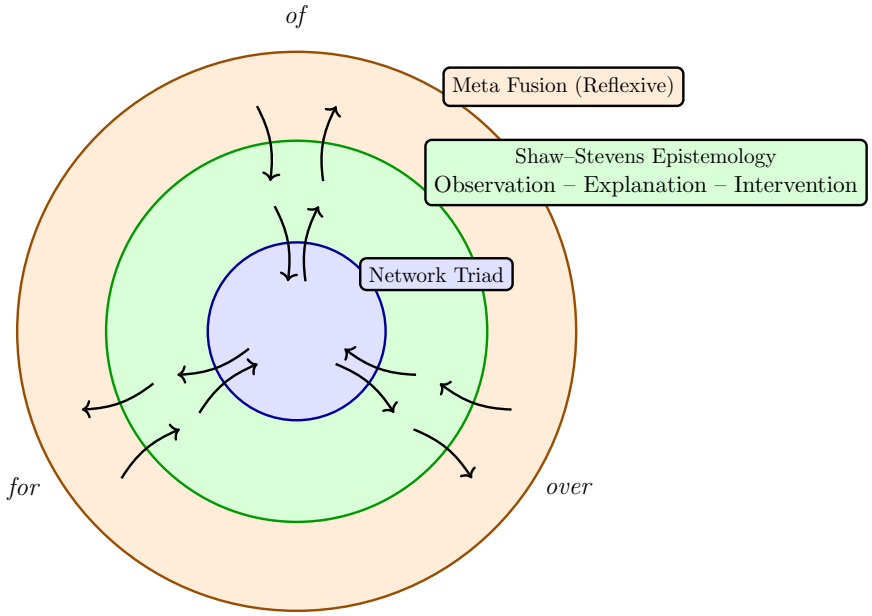


Figure 2.9: Meta-fusion as three nested layers. The inner *triad* (of–for–over) sits within the Shaw–Stevens epistemology, itself enclosed by a reflexive meta-layer. Red arrows indicate two-way feedback: methodology shapes epistemology and vice versa.

We also claim that the network triad becomes a data agent within the autopoietic system, with a task of generating knowledge through data discoveries that update the state of the data universe.

We stress that autopoiesis should not be taken as solipsism. For Maturana and Varela, a living system maintains its identity by remaining open to energy and information from the environment. We invoke the need to the same principle to apply inside the proposed ecosystem that finds its essence in the engagement with social, political, and ethical realities.

In this meta fusion, closure and openness are no longer opposites but phases of a single process.

Dimension	**Fusion within the Triad**	**Fusion with Shaw–Stevens**	**Meta-Fusion (Reflexive)**
Ontological	Data, model, and domain are unified into a single representation. The triad triggers an ontology of relations where there is no more distinction between object and context.	From processes to paradigms: the triad supplies a concrete ontology compatible with the Shaw-Stevens agenda.	Mutual embedding of data ontology and epistemology. The network becomes a co-agent in data universe updates.
Methodological	Inference, geometry, and observation co-evolve. Empirical and generative logic fold into one another through network dynamics.	Hierarchical constructions test and inform the critique of LSD. Network hierarchy offers a tool for plural and adaptive modelling.	Recursive design of hypotheses: methodology interrogates its own assumptions. Epistemology becomes a design principle guiding computational practice.
Ethical / Operational	Transparency and interpretability are the essence of model design; interpretability and trustworthiness acts as moral constraints within computation.	Accountability and responsibility emerge as conditions for explanation. Hence, ethics becomes an epistemic dimension.	Co-governance of models and meanings. Data ethics and epistemic humility are folded within methodological loops, ensuring that closure remains open to critique.

Table 2.8: Schematic synthesis of the three levels of fusion: within the triad, between the triad and the Shaw-Stevens paradigm, and their reflexive meta-fusion. Each row corresponds to an ontological, methodological, or ethical dimension of interaction.

Chapter 3

Temporally-Evolving Generalised Networks and their Kernels[1]

3.1 Introduction

Data indexed on networks have increasingly attracted the interest of the statistical and machine learning communities. The driver for such an interest is the need to represent complex systems where quantifying the interactions between components covers a fundamental importance. Networks provide an, albeit abstract, flexible framework that may describe and sometimes reveal relationships and dependencies among variables stemming from diverse applications, ranging from transportation infrastructures and communication systems to biological structures and climate-related processes. Within the realm of space-time statistics, networks naturally arise in several applications, since they allow to extend the classic modelling stochastic processes machinery over irregular, yet structured, and often non-Euclidean spaces (Borovitskiy et al., 2023; Bolin et al., 2024).

Traditionally, the statistical literature has focused on data defined over static networks, where the topology is assumed fixed throughout the analysis (Tsonis and Roebber, 2004). Such an assumption is deemed unrealistic for many applications where the underlying system evolves dynamically over time. In several examples, the connectivity structure changes over time: edges may appear or vanish; nodes may emerge or disappear. Further, the geometry governing the relationships between entities might change as a function of time. As a consequence, the temporally-evolving nature of many networks poses challenges to modelling, inference and prediction.

There has been some work related to purely spatial stochastic processes continuously defined over networks or for space-time processes where the network topology does not change as a function of time (Anderes et al., 2020; Bolin et al., 2024; Porcu et al., 2023).

The attention has been specially devoted to Gaussian processes. One group of

[1]This chapter is a version of the paper Filosi et al. (2025a), available on arXiv:2309.15855.

authors (Anderes et al., 2020; Porcu et al., 2023, 2022) directly models the covariance function as a function of some *distance* that is properly defined over a network, while another group of authors (Bolin and Lindgren, 2011; Bolin et al., 2024) obtains valid covariance functions on the basis of a given class of SPDEs. Both approaches have positive features. The SPDEs group works with more general topologies that allow, for instance, for multiple edges connecting any two vertices of the network; however, this point is not very relevant as the results in Anderes et al. (2020) imply that multiple edges can be achieved at the expense of adding arbitrary degree-two vertices over the network, without altering its intrinsic topology. Further, Bolin et al. (2025) achieve a class of Gaussian random fields having a covariance function that is arbitrarily differentiable at the origin, with the advantages implied by such property in terms of kriging prediction. The approach by Anderes et al. (2020) can instead count on generality, as a wealth of parametric examples is available thanks to this approach. However, none of the covariance functions proposed in Anderes et al. (2020) is differentiable at the origin. Another advantage of working with covariance functions is that Porcu et al. (2023) provide a method based on direct construction that allows to create several parametric classes of nonseparable space-time covariance functions.

However, none of these approaches considers topologies that can evolve over time. Several real-life applications motivate this research field. For instance, in neuroscience, functional brain connectivity evolves throughout development, learning, and disease progression (Medaglia et al., 2015; Uddin and Karlsgodt, 2018); in environmental science, river networks change due to erosion, sediment transport, or climatic shifts (James and Roulet, 2007; Masselink et al., 2017); in engineering, transportation and energy infrastructures adapt through expansion, failure, or technological advances (Newman et al., 2011).

Chapter 2 has classified the role of networks in climate data into three paradigms: networks *of* data, networks *over* which data are defined, and networks *for* data. Networks *of* data are constructed using similarity (synchronous) measures among observations. Networks *for* data pertain to the use of network-based models in machine learning, such as graph neural networks. Finally, networks *over* data—the focus of the current Chapter, called geophysical networks therein—describe domains where the network topology is pre-defined, such as river systems or transportation grids. Within the third paradigm, the temporal evolution of the network remains a critical and unexplored challenge in statistical modelling.

This paper works under the paradigm of covariance-based constructions for Gaussian processes defined on temporally-evolving networks. There are several intricacies inherent to this challenge: the notion of distance —a cornerstone in defining covariance structures— is itself dynamic, as paths between points may lengthen, shorten, or be redefined over time. Second, the integration of temporal dynamics into spatial structures is a non-trivial task. Separable covariance models treating time and space independently are definitely inadequate to describe scenarios where spatial distances depend explicitly on time horizon. Third, covariance functions

must be positive definite, and such a task entails a fair amount of mathematical work.

We provide a comprehensive framework for modelling stochastic processes over temporally-evolving networks, while extending the existing theory of graphs with Euclidean edges (Anderes et al., 2020) to temporal dynamics. Our constructions accounts for either linear and periodic temporal evolutions, reflecting systems that evolve arbitrarily or cyclically over time. Our first argument to advocate in favour of a periodic construction is that it suits perfectly to several real-world phenomena, where both linear-time evolution (*e.g.* long-term trends) and cyclic oscillations (*e.g.* seasonal components) might happen. To make an example, consider temperatures in a given geographical area: apparently there might be strong correlations between: (i) contiguous spatial points at a given time, which are represented by means of spatial edges; (ii) the same points considered at contiguous times, which are represented by means of temporal edges between temporal layers and (iii) the same points considered at the same periods of the year, which are considered in the model as they are exactly the same point in the temporally evolving graph. In addition, in many real-world applications, the network underlying a system is only partially observable. As a consequence, it could be hard or impossible to specify the whole time-evolving (not periodic) network in cases of long time series. The periodic assumption, when all in all reasonable, may be a great help in this circumstance as well.

Our method embeds temporal dynamics through the addition of temporal edges into the network structure, and allows for semi-metrics that capture both spatial and temporal proximities in a unified manner. Clearly, the semi-metrics cover a fundamental part to provide the new covariance structures. Our contribution is not a mere technical generalisation, but responds to concrete statistical needs: in hydrology, the dynamic connectivity of catchments affects the propagation of pollutants or the prediction of flood events (James and Roulet, 2007); in neuroscience, evolving functional networks influence the interpretation of longitudinal brain imaging studies (Medaglia et al., 2015); in engineering, the resilience of infrastructures under dynamic conditions requires models that incorporate changing connectivity patterns Newman et al. (2011).

To illustrate how our framework may apply to a concrete scientific problem, consider longitudinal brain connectivity studies using fMRI data (Damaraju et al., 2014). Within this framework, dynamic functional connectivity (dFC) is associated with evolving patterns of interaction between brain regions through the so-called time-resolved correlation networks (Hutchison et al., 2013; Preti et al., 2017). Existing models make use of sliding-window correlations, or of vector autoregressive approaches applied to static graphs (Lindquist et al., 2014), which fail to account for continuous topological changes such as the emergence or dissolution of edges, or shifts in network topology.

By contrast, our framework models the data via a temporally-evolving graph: each layer represents the functional network at a given time and temporal edges can link corresponding nodes (or regions) across successive layers with weights that

reflect continuity or similarity. We can then define a Gaussian process over this evolving network using resistance-based semi-distances (see Section 3.4), allowing us to model neural activity as a continuous stochastic process indexed by a time-varying domain. For statisticians, this approach may pave the way for inference tasks such as predicting connectivity or activity on unobserved regions or times, quantifying uncertainty across layers and detecting anomalous structural changes, which are key for understanding developmental trajectories or disease progression (Medaglia et al., 2015).

Another example may come from modelling vascular networks in tumour imaging. Tumour angiogenesis — where new vasculature forms in response to tumour growth — can be seen as a dynamic stochastic process, with vessel topology and perfusion patterns evolving over time (Folkman, 2002; Kim et al., 2016). By modelling perfusion as a Gaussian process over the evolving vascular graph, our approach may allow for interpolation at unsampled sites, identification of emergent or collapsing vascular pathways, and quantification of structural and functional uncertainty. Hence, the new framework proposed here provides potentially valuable insights for prognosis or therapy planning in oncology.

As a final contextualisation of our method, urban transportation systems are inherently dynamic: the topology of routes, road availabilities, and patterns associated with passenger flows change over time: construction, accidents, policy interventions (*e.g.*, road closures or congestion pricing), or daily temporal rhythms such as rush hours (Lilleborge et al., 2025). In metropolitan areas, for instance, public transit networks — buses, trains, and subways — have a temporally dynamical connectivity with nodes (stations, stops) and edges (routes or paths) are active only at specific time intervals or vary in travel time and reliability throughout the day (Zheng et al., 2014; Rico et al., 2023).

According to our framework, the transportation system may be modelled as a time-indexed sequence of graphs, where each layer corresponds to the transportation network at a given time (*e.g.*, every 15 minutes). Edges across temporal layers translate into continuity or transitions in accessibility, *e.g.*, based on predicted travel times or real-time availability of connections. Our dynamical covariance functions can model quantities such as commuter flow, vehicle congestion, or service reliability and aligns with the realities of smart-city infrastructure and urban analytics.

The contributions of this work are organised as follows. Section 3.2 recalls the main mathematical objects that will be used. Section 3.3 builds the skeleton of our construction, *i.e.* time-evolving graphs, which are exploited in Sections 3.4 and 3.5, where time-evolving graphs with Euclidean edges are defined for the linear time and circular time cases, respectively. Section 3.6 delves into the construction of the reproducing kernel Hilbert spaces of the processes underlying our model. Section 3.7 illustrates how it is possible to build kernels on time-evolving graphs and present some examples, whilst Section 3.8 explores some numerical experiment. Finally, Section 3.9 provides conclusions and final remarks.

3.2 Mathematical Background

This material is largely expository and provides the mathematical background and notation used in the next sections.

Definition 3.1 (Moore-Penrose generalised inverse)**.** Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ be a matrix. Then its *Moore-Penrose generalised inverse* is the unique matrix $\mathbf{M}^- \in \mathbb{R}^{n\times n}$ such that:

- $\mathbf{M}\mathbf{M}^-\mathbf{M} = \mathbf{M}$ and $\mathbf{M}^-\mathbf{M}\mathbf{M}^- = \mathbf{M}^-$,
- $\mathbf{M}\mathbf{M}^- = (\mathbf{M}\mathbf{M}^-)^\top$ and $\mathbf{M}^-\mathbf{M} = (\mathbf{M}^-\mathbf{M})^\top$.

Definition 3.2 (g-embedding, isometric embedding)**.** Let (X, d) be a semi-distance space and $g : D_X^d \to [0, +\infty)$ a function, where D_X^d denotes the diameter of X. Then (X, d) is said to have a g-embedding into a Hilbert space $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$, written $(X, d) \overset{g}{\hookrightarrow} \mathcal{H}$, if there exists a map $\psi : X \to \mathcal{H}$ such that, for all $x, y \in X$,

$$g(d(x, y)) = \|\psi(x) - \psi(y)\|_{\mathcal{H}} .$$

If g is the identity map, then it is called *isometric embedding*.

3.2.1 Gaussian Random Fields over Semi-Distance Spaces

Let us begin with a brief introduction about Gaussian random fields (Stein, 1999). Let X be a non-empty set and let $k : X \times X \to \mathbb{R}$ be a function. Then k is a *positive semi-definite* function (or a *kernel*, or a *covariance function*) if and only if it is symmetric and, for all $n \in \mathbb{N}^+$, $x_1, \ldots, x_n \in X$ and $a_1, \ldots, a_n \in \mathbb{R}$, $\sum_{i=1}^n \sum_{j=1}^n a_i a_j k(x_i, x_j) \geq 0$. If, in addition, whenever the above relation is an equality, then necessarily $a_1 = \cdots = a_n = 0$, k is *(strictly) positive definite*.

We denote Z a real-valued random field over X, *videlicet*: for each $x \in X$, $Z(x)$ is a real-valued random variable. Then Z is called *Gaussian* if, for all $n \in \mathbb{N}^+$ and $x_1, \ldots, x_n \in X$, the random vector $\boldsymbol{Z} := (Z(x_1), \ldots, Z(x_n))^\top$, with $\top$ denoting the transpose operator, follows a n-variate Gaussian distribution.

A Gaussian random field Z on X is completely determined by its first two moments: the mean function $\mu_Z : X \to \mathbb{R}$, $x \mapsto \mathbb{E}(Z(x))$, with $\mathbb{E}$ denoting the expectation operator, and the covariance function (kernel) $k_Z : X \times X \to \mathbb{R}$, $(x_1, x_2) \mapsto \mathbb{C}\mathrm{ov}(Z(x_1), Z(x_2))$. A necessary and sufficient condition for a function k_Z to be a covariance function (a kernel) of some random field Z is to be symmetric and positive semi-definite.

A mapping $d : X \times X \to \mathbb{R}$ is called *semi-distance* on X (or, equivalently, (X, d) is a *semi-distance space*) if the following conditions hold for each $x, y \in X$: $d(x, y) \geq 0$, $d(x, y) = 0 \iff x = y$ and $d(x, y) = d(y, x)$. In addition, d is called a *distance* on X (or, equivalently, (X, d) is a *distance space*) if it is a semi-distance and the triangle inequality holds, namely, for all $x, y, z \in X$: $d(x, y) + d(y, z) \geq d(x, z)$. The covariance function k_Z is called isotropic for the semi-distance space (X, d) if there

exists a mapping $\psi : D_X^d \to \mathbb{R}$ such that $k_Z(x, y) = \psi(d(x, y))$, for $x, y \in X$. Here, $D_X^d := \{d(x_1, x_2) : x_1, x_2 \in X\}$ is the diameter of X. Theorem 3.1 in Section 3.7 provides a useful tool to construct isotropic kernels on arbitrary domains. For a Gaussian random field Z on X, we define its *variogram* $\gamma_Z : X \times X \to \mathbb{R}$ through

$$\gamma_Z(x_1, x_2) := \mathbb{V}\mathrm{ar}\left(Z(x_1) - Z(x_2)\right), \qquad x_1, x_2 \in X, \tag{3.1}$$

with $\mathbb{V}\mathrm{ar}$ denoting the variance operator. Schoenberg (1942) proves that γ_Z is a variogram if and only if the mapping $\exp(-\gamma_Z(\cdot, \cdot))$ is positive definite on $X \times X$.

Let (X_1, d_1) and (X_2, d_2) be two semi-distance spaces. Then, the triple $(X_1 \times X_2, d_1, d_2)$ is called a *product semi-distance space*. Menegatto et al. (2020) define *isotropy* over a product semi-distance space through continuous functions $\psi : D_{X_1}^{d_1} \times D_{X_2}^{d_2} \to \mathbb{R}$ such that, for $(x_1, x_2), (x_1', x_2') \in X_1 \times X_2$,

$$((x_1, x_2), (x_1', x_2')) \mapsto \psi(d_1(x_1, x_1'), d_2(x_2, x_2')), \tag{3.2}$$

is positive definite. The above definition naturally arises from spatio-temporal settings: suppose we have a *static* semi-distance space (X, d) that represents some spatial structure and (T, d_T) representing time, where $T \subseteq \mathbb{R}$ and, usually, $d_T(t_1, t_2) = |t_1 - t_2|$. In such a case, Equation (3.2) can be re-adapted to define kernels. This is the setting adopted by Porcu et al. (2023) and by Tang and Zimmerman (2024).

Later on, we deviate from the literature and we instead consider a distance space X_t that evolves over time, $t \in T$. Hence, our domain is written as

$$\{(x_t, t) \, : \, x_t \in X_t, \, t \in T\},$$

where t describes *time*, and the graph coordinate x_t is constrained on the space X_t. Such a framework entails a way more sophisticated construction to equip such a space with a proper distance.

3.2.2 Graphs with Euclidean Edges

We start with a formal definition of graphs with Euclidean edges. We slightly deviate from the definition provided by Anderes et al. (2020), for the reasons that will be clarified subsequently. We will use the notation and the definitions presented in the Subsection Standard Graph Theory and Notation (pag. 15).

For the sake of simplicity, henceforth we may denote edges as $e \in E$, without explicitly referencing the vertices they connect.

Definition 3.3 (Graph with Euclidean edges)**.** Consider a simple, connected and weighted graph $G = (V, E, w)$, where $w : E \to \mathbb{R}^+$ represents the weight mapping. Then, G is called a *graph with Euclidean edges* provided that the following conditions hold.

1. Edge sets. Each edge $e \in E$ is associated to the compact segment, also denoted by e, $[0, \ell(e)]$, where $\ell(e) := w(e)^{-1}$ may be interpreted as the *length* of the edge e.

2. Linear edge coordinates. For each edge $e \in E$ there is a bijection, $\varphi_e : e \to [0, \ell(e)]$. For each point $x \in [0, \ell(e)]$, we define $u := \varphi_e^{-1}(x)$. In particular, we call $\underline{u} := \varphi_e^{-1}(0)$ and $\overline{u} := \varphi_e^{-1}(\ell(e))$ and require that $\underline{u}$ and $\overline{u}$ are the two endpoints of e in V. Thus, each point u is uniquely identified by the triple $(\underline{u}, \overline{u}, \delta(u))$, where $\delta(u) := \frac{\varphi_e(u)}{\ell(e)} = \varphi_e(u)\, w(e)$ is the relative position of $\varphi_e(u) \in [0, \ell(e)]$.

Henceforth, we shall assume the existence of a total order relation on the set of vertices V and that every edge is represented through the ordered pair (v_1, v_2), where $v_1 < v_2$. In particular, for each $u \in e$, the endpoints of e, $\underline{u}$ and $\overline{u}$ satisfy the relation $\underline{u} < \overline{u}$. Although this assumption is not essential for the following, we think that in practice it simplifies the exposition, as each point on a given edge e is written in exactly one way and, therefore, there is only one quantity $\delta(u)$. See Figure 3.1 (right) for an illustration.

Notice that our Definition 3.3 does not require any *distance consistency* opposed to Anderes et al. (2020, Definition 1, (d)). The reason is that our setting does not need a bridge between geodesic and resistance metrics. In addition, we have restricted the space of possible bijections from each edge onto closed intervals to be linear: the main reason is that the focus of this paper is not to explore isometric embeddings, but to provide suitable topological structures evolving over time, and stochastic processes attached to them. We stress that the framework introduced

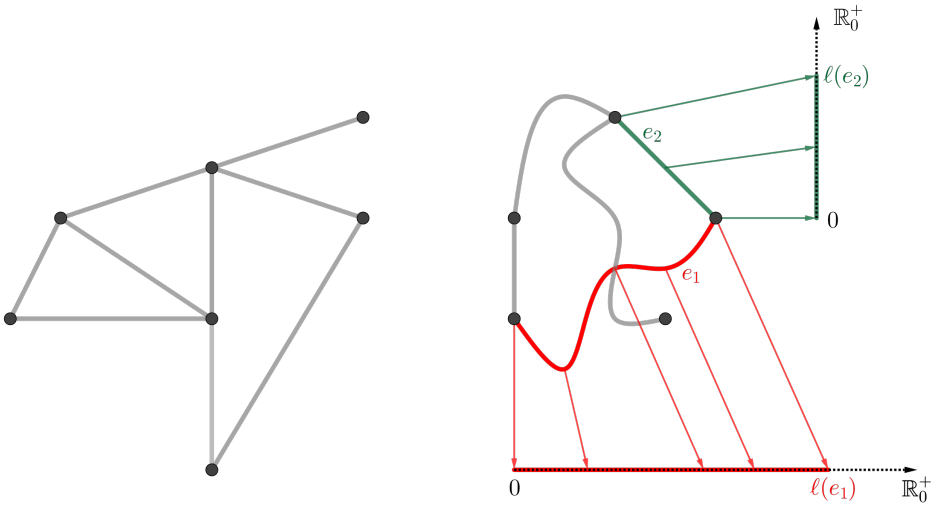


Figure 3.1: Left: a linear network. Right: a graph with Euclidean edges, where the bijections φ_{e_1} and φ_{e_2} between the edges e_1 and e_2 and their respective real segments $[0, \ell(e_1)]$ and $[0, \ell(e_2)]$ are stressed.

through Definition 3.3 is more general than linear networks, for at least two reasons: (i) in our framework, the weights of each edge be chosen independently from the others, and (ii) our framework needs no restriction on the network structure, *e.g.*, as

shown in Figure 3.1 (right), edges may cross without sharing the crossing point.

3.2.3 Graph Laplacian and Resistance Metric

The resistance metric has been widely used in graph analysis, as it is more natural than the shortest-path metric when, for instance, considering flows or transport networks, where multiple roads between two given points may share the total flow (Jorgensen and Pearse, 2010). Below we introduce the effective resistance distance for an undirected and connected graph, we briefly report its definition and its mathematical construction.

Let $G = (V, E, w)$ be a simple, weighted and connected graph and let $\mathbf{W}$ its adjacency matrix, that is: $\mathbf{W}(v_1, v_2) = w((v_1, v_2))$, where we set $w((v_1, v_2)) = 0$ whenever $v_1 \not\sim v_2$ (*i.e.*, v_1 is not adjacent to v_2). In addition, for each node $v \in V$, we define its *degree* as the sum of the weights of the edges adjacent to it. Let $\mathbf{D}$ be the degree matrix of G, *i.e.* the diagonal matrix where each diagonal element is the degree of the corresponding vertex. Then, the *Laplacian matrix* (or simply *Laplacian*) of G is the matrix $\mathbf{L} := \mathbf{D} - \mathbf{W}$.

Laplacian matrices enjoy several properties (Devriendt, 2022): they are symmetric, diagonally dominant (thus positive semi-definite) and singular with exactly one null eigenvalue, corresponding to the eigenvector $\mathbf{1}_n$. Furthermore, they have non-positive off-diagonal entries and positive main-diagonal entries.

A graph $G = (V, E)$ is called a *resistor graph* if the edges $e \in E$ represent electrical resistors and the nodes represent contact points. Given a resistor graph, the *effective resistance distance* R between two vertices is defined as the voltage drop between them when injecting one Ampère of current in one and withdrawing one Ampère from the other. Several mathematical formulations of this concept have been provided, and the reader is reminded, among others, to Jorgensen and Pearse (2010, Subsection 2.1). Throughout, we follow Ghosh et al. (2008). Let $G = (V, E)$ be a resistor graph. For each $v_1 \sim v_2 \in V$, let $r(v_1, v_2) \in \mathbb{R}^+$ denote the resistance of the resistor that connects v_1 and v_2. In addition, for each $v_1, v_2 \in V$, define the weight (which plays the role of the physical conduttance) $w((v_1, v_2)) := 1/r((v_1, v_2))$ if $v_1 \sim v_2$ and $w((v_1, v_2)) := 0$ if $v_1 \not\sim v_2$. Let $\mathbf{L}$ be the Laplacian matrix of G with the above-defined weights, $\mathbf{L}^-$ its Moore-Penrose generalised inverse (see Definition 3.1). Finally let $\boldsymbol{e}_{v_i}$ denote the vector with all zeroes, except a one at position v_i. Then the effective resistance distance R between two nodes v_1 and v_2 is given by

$$R(v_1, v_2) = (\boldsymbol{e}_{v_1} - \boldsymbol{e}_{v_2})^\top \mathbf{L}^- (\boldsymbol{e}_{v_1} - \boldsymbol{e}_{v_2}).$$

3.3 Time-Evolving Graphs with Euclidean Edges

In this Section we provide the construction of time-evolving graphs with Euclidean edges. While providing full specifications below, we start with some description and a graphical representation to give an intuition of how these graphs work. The *rationale* behind this construction is shown in Figure 3.3: we build a new graph containing

Step 1. Define a time-evolving graph as a properly defined sequence of graphs indexed by discrete time instants.

Step 2. Define *connected equivalent simple* time-evolving graphs by completing a time evolving graph through a set of edges that connect the same nodes at different time instants.

Step 3. Over the connected equivalent simple graph, we can now define, for every time t, a graph with Euclidean edges, G_t.

Step 4. Define a time-evolving graph of order 1 to exploit computational advantages.

Box 3.2: A sketch of our construction.

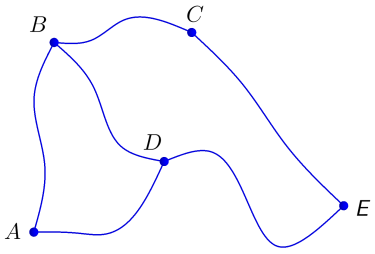
all the graphs a different time instants, with additional edges connecting the same nodes at different times. However, the mathematical description of such an object is quite involved as it requires several steps and substantial notation. To guide the reader in following, we provide a sketch of our procedure in Box 3.2. **Step 1.** is completely general and does not require any topological structure on every *marginal* graph G_t, for a given time t. Yet, having graphs with Euclidean edges that evolve over time requires some more work, and this fact justifies **Step 2.**, which allows for connectivity, being one of the properties *sine qua non* of a graph with Euclidean edges. **Step 4.** is not mathematically necessary to guarantee the validity of the structure, but it is justified by computational and intuitive reasons as explained throughout.

Step 1. starts with the definition of *equivalent simple graph*. The intuitive idea behind the construction formalised next is depicted in Figure 3.3: consider m layers, each representing a different temporal instant (namely a graph G_t), and connect them by means of additional intra-time edges, which account for the time-dependency of the graphs. Each vertex v in a given layer is uniquely identified by a labelling function $s(v)$ so that it is possible to track its evolution among different layers. In Figure 3.3 the set of labels is $\{A, B, C, D, E, F\}$.

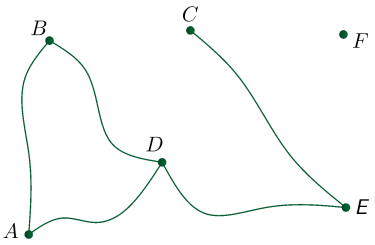
Definition 3.4 (Time-evolving graph)**.** Let $T = \{0, ..., m-1\}$ be a (finite) collection of time instants. To every time instant $t \in T$ we associate a simple undirected and weighted graph $G_t = (V_t, E_t, w_t)$, with $V_t \cap V_{t'} = \emptyset$ whenever $t \neq t'$. For an edge $e_t \in E_t$, the corresponding weight is denoted $w(e_t) := w_t(e_t)$. We use $n_t := |V_t|$ for the number of vertices at time t.

Let $G = \{G_0, ..., G_{m-1}\}$ be the associate finite collection of these graphs. Call $V := \bigcup_t V_t$ the set of vertices, $n := |V|$ the total number of vertices, and $E_S := \bigcup_t E_t$ the set of *spatial* edges. Finally, if $v \in V$, whenever convenient we shall write $t(v)$ for the unique value t such that $v \in V_t$.

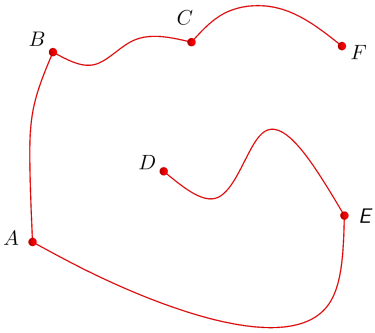
Let $s : V \to S$ be a mapping from V, where S is a set of labels, with the following properties:



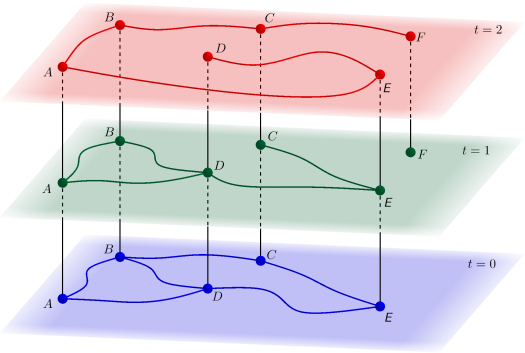
(a) Graph at $t = 0$.



(b) Graph at $t = 1$.



(c) Graph at $t = 2$.



(d) Equivalent simple graph.

Figure 3.3: An example of an equivalent simple graph (bottom-right), with $m = 3$, $S = \{A, B, C, D, E, F\}$, $n_1 = 5$, $n_2 = n_3 = 6$. The coloured edges belong to E_S, whilst the black ones belong to E_T. The temporal *slices* at time instants $t = 0, 1, 2$ are reported, respectively, at quadrants (a), (b) and (c).

1. for all $t \in T$ and for all $v_1 \neq v_2 \in V_t$, necessarily $s(v_1) \neq s(v_2)$, meaning that two vertices at the same time cannot have the same label;

2. for $t_1 \neq t_2$, two vertices $v_1 \in V_{t_1}$ and $v_2 \in V_{t_2}$ are considered (in the resulting time-evolving graph) the same vertex at different times if and only if $s(v_1) = s(v_2)$.

We call the triple $\boldsymbol{G} = (T, G, s)$ a *time-evolving graph*.

While Definition 3.4 provides a flexible framework to manage graphs that evolve over time, we are going to merge its underlying idea with the one of graph with Euclidean edges presented in Subsection 3.2.2.

Step 2. of our procedure intends to *complete* the time-evolving graph as in Definition 3.4 so to ensure spatio-temporal connectivity.

Definition 3.5 (Equivalent simple time-evolving graph)**.** Let $\boldsymbol{G}$ be a time-evolving graph. We define its *equivalent simple* time-evolving graph, $\widetilde{\boldsymbol{G}} = (V, \widetilde{E})$ as the graph with edges $\widetilde{E} := E_S \cup E_T$, with E_T a set of additional edges (called *temporal* edges throughout) that connect the same nodes at different time instants. More precisely, we have:

$$E_T \subseteq \{(v_1, v_2) \in V \times V \,:\, s(v_1) = s(v_2),\, t(v_1) \neq t(v_2)\}\,.$$

To each new edge $e = (v_1, v_2) \in E_T$ a weight $w(e) > 0$ is assigned, while all the other weights remain unchanged.

Remark 3.1. Some comments are in order.

- It is reasonable to choose a weight that depends only on the temporal distance $|t(v_1) - t(v_2)|$. One possibility is to choose $w(e) := \alpha\,|t(v_1) - t(v_2)|^{-r}$, with $\alpha, r > 0$ given scale factors. Although we assume this particular expression with $r := 1$ in all the following examples, we stress that any choice leads to a valid model as long as $w(e) > 0$.

- Notice that an equivalent simple time-evolving graph is a graph with Euclidean edges.

- We assume $\widetilde{\boldsymbol{G}}$ to be *connected*. If this is not the case, we henceforth consider each connected component independently from the others.

Step 2. ensures that it becomes feasible to assign, to each temporal label $t \in T$, a graph with Euclidean edges to the equivalent simple graph associated with a given time-evolving graph (**Step 3.**). We note that the choice of the temporal edges needs care. Indeed, the set of possible temporal edges for a fixed label $s \in S$ may grow quadratically in the number of considered temporal instants m. An interesting class of time-evolving graph that allows to control the structure of temporal edges is the time-evolving graph of order p, formalised below.

Definition 3.6 (Time-evolving graph of order p)**.** Let $p \in \mathbb{N}^+$. A *time-evolving graph of order p* is a time-evolving graph $\boldsymbol{G} = (V, E)$, where

$$E_T \subseteq \{(v_1, v_2) \in V \times V \,:\, s(v_1) = s(v_2),\, 1 \leq |t(v_1) - t(v_2)| \leq p\}\,. \tag{3.3}$$

This class of models restricts the sets of temporal edges so that they cannot link directly nodes more than p time steps away. We stress that mathematical results following hold in the general case, even for the case of non adjacent layers. Yet, being an order p graph simplifies both interpretation and computational aspects. In fact, it allows for a plain representation of an equivalent simple graph. Furthermore, there are non-negligible computational reasons. Indeed, for large networks, order p graphs have a sparse Laplacian matrices (block $(2p+1)$-diagonal) of the associated equivalent simple graph. This entails huge computational savings in both terms of storage and computation. Finally, allowing edges between non-adjacent layers could lead to a huge number of weights α's. In particular, under the assumption of temporally homogeneous weights (the weight of a temporal edges depends only on the temporal distance between the layers it connects), we would have $m-1$ possible weights if order $m-1$ graph is assumed. For a large collection of time instants, this can become computationally unfeasible.

We call a time-evolving graph of order p $\boldsymbol{G}$ *temporally complete* when the set E_T is identically equal to the set in the right hand side of (3.3). Albeit such a property is not required to prove our theoretical results, it is operationally useful as it allows to avoid removing or adding temporal edges.

3.4 Resistance Metrics for Linear Time

We start this section by noting that defining the classic resistance metrics between nodes of the temporally evolving graph is not an issue. Yet, we are dealing with a graph where (semi-)distances should be computed between any pair of points lying continuously over the edges.

For the case of static graphs with Euclidean edges, Anderes et al. (2020) provide an ingenious construction that allows for a suitable continuously-defined metric on the basis of Brownian bridges and their variograms.

The idea is to follow a similar path, by defining a Gaussian process that is continuously indexed over the edges of an equivalent simple connected graph associated with a given time-evolving graph.

Before going into technical details, we present a brief outline. Following Anderes et al. (2020), we are going to define a semi-distance on all the points of the graph, namely its vertices and the points on its edges. To this aim, we define a Gaussian process Z on every point of the time-evolving equivalent simple graph and then *define* the semi-distance between two points as the variogram of such process, *i.e.*, for each $u_1, u_2 \in \widetilde{\boldsymbol{G}}$:

$$d(u_1, u_2) := \gamma_Z(u_1, u_2), \tag{3.4}$$

with γ_Z as being defined through (3.1). In such a way, we can directly apply Theorem 3.1 stated in Section 3.7 to obtain kernels. Here, $Z := Z_V + Z_E$ is the sum of two independent Gaussian processes defined on the equivalent simple graph $\widetilde{\boldsymbol{G}}$. The process Z_V accounts for the structure of the graph (namely its vertices and the weights of its edges) and plays the role of major source of variability, whilst Z_E adds some variability on the edges and accounts for the temporal relationship between the same edge at different times. The following remark delves into the motivation behind the definitions of the processes Z_V and Z_E, to give the reader a (hopefully) clear roadmap and useful intuition.

Remark 3.2. The *rationale* behind the definitions of Z_V and Z_E is closely related to the desired semi-metric properties. Indeed, the objective in Anderes et al. (2020) is to build an extension to the classical resistance distance and, to achieve such a result, they define the analogous of a Wiener process on the graph. To have an intuition about such a choice, consider the standard Wiener process $(W_t)_t$ on the positive real line $\mathbb{R}_0^+$: we have $\mathbb{C}\mathrm{ov}\,(W_t, W_s) = \min(t, s)$ and, as a consequence, $\gamma_W(t, s) := \mathbb{V}\mathrm{ar}\,(W_t - W_s) = |t - s|$, *i.e.* the variogram is the Euclidean distance, which, considering the positive real line as a uniform electric resistor, coincides with the effective resistance distance. Perhaps surprisingly, this continues to hold shifting from the positive real line to a graph with Euclidean edges. Clearly, the definition of the Wiener process cannot be applied directly to a graph, which has several differences with $\mathbb{R}_0^+$: it is compact and does not have any intrinsic order or sum operation. As a consequence, the Wiener process $(W_t)_t$ needs to be adapted to the new topology and Anderes et al. (2020) defined it via the conditional independence structure offered by the (modified) Laplacian, which entails the topology of the graph. Therefore, they define Z_V on the vertices as the Gaussian random vector with precision matrix $\mathbf{L}$: in such a way, for $v_1 \not\sim v_2 \in V$, $Z_V(v_1)$ and $Z_V(v_2)$ are conditional independent given $Z_V(V \setminus \{v_1, v_2\})$. Furthermore, it has been shown that the inverse Laplacian matrix of a resistor graph offers an efficient way to compute all the effective resistance distances between its nodes, as we mentioned in Subsection 3.2.3. Therefore, since we define Z_E at vertices to be zero and since the linear operation on the inverse Laplacian matrix to obtain the effective resistance is actually the variogram, it is natural to define Z_V having precision matrix $\mathbf{L}$.

Once the process is defined on the vertices V, it remains to define it on the edges. The idea is to build, for each edge $e = (v_1, v_2) \in E$, a Brownian bridge linking $Z_V(v_1)$ and $Z_V(v_2)$, thus summing a linear interpolation between $Z_V(v_1)$ and $Z_V(v_2)$, and adding a standard Brownian bridge on $[0, \ell(e)]$.

Clearly, this construction needs to be adapted when we shift from the purely spatial case of Anderes et al. (2020) to the time-evolving setting studied in this Chapter. Indeed, both Z_V and Z_E need some adaptation to cope with the time-evolving dynamic. Regarding Z_V, the (modified) Laplacian matrix of the equivalent simple graph offers a nice conditional independence structure: while, for a given time t, the conditional independence structure is the same as there was no time, the links between adjacent time instants of an (order 1) equivalent simple graph provide the

same conditional independence structure among times. More specifically Z_V at layer t_1 is independent from Z_V at layer t_3 given Z_V at layer t_2, for each $t_1 < t_2 < t_3$.

The construction of Z_E extends the original construction by Anderes et al. (2020) as well. On each edge at a given time, its covariance structure is the same, however, we add some correlation (governed by the choice of k_T) between times. This is motivated by the following: imagine that the sampling time of the graph has a scale much smaller than the process on it. Then, for two adjacent time instants, say t and $t+1$, it is reasonable to assume that the process on a given edge e does not change too much. This is achieved by having a high correlation of the final process (and thus a low variogram of $Z = Z_V + Z_E$). Therefore, a high correlation of Z_E for adjacent time will fix this issue.

3.4.1 Formal Construction of Z_V and Z_E

We start with the definition of the process $Z_V = \left\{ Z_V(u) \,:\, u \in \widetilde{\boldsymbol{G}} \right\}$. First, we define the process Z_V at the vertices $V = \{v_1, \dots, v_n\}$ as a zero-mean multivariate normal random variable:

$$\boldsymbol{Z}_V\big|_V := (Z_V(v_1), \dots, Z_V(v_n))^\top \sim \mathcal{N}_n\left(\mathbf{0},\, (\mathbf{L}^\star)^{-1}\right),$$

where $\mathbf{L}^\star := \mathbf{L} + \boldsymbol{x}\boldsymbol{x}^\top$ with $\mathbf{L}$ the Laplacian matrix associated to the graph $\widetilde{\boldsymbol{G}}$ and $\boldsymbol{x}$ being a vector such that ${\mathbf{1}_n}^\top \boldsymbol{x} \neq 0$. The role of the inverse Laplacian matrix is explained in Remark 3.2, whilst the reason underlying the addition of $\boldsymbol{x}\boldsymbol{x}^\top$ is that $\mathbf{L}$ is singular, whilst $\mathbf{L}^\star$ is strictly positive definite. This will turn out to be necessary in the construction of the reproducing kernel Hilbert space, delved into in 3.6. However, we stress that the semi-distance and the resulting covariance functions remain the same for $\boldsymbol{x} = \mathbf{0}$ (thus $\mathbf{L}^\star = \mathbf{L}$), as shown in Proposition 3.2 below. Notice that Anderes et al. (2020) chose $\boldsymbol{x}$ as a vector of the canonical base of $\mathbb{R}^n$. The process $\boldsymbol{Z}_V\big|_V$ is then extended to the whole equivalent simple graph via a sheer linear interpolation:

$$Z_V(u) := (1 - \delta(u))\, Z_V(\underline{u}) + \delta(u) Z_V(\overline{u}), \tag{3.5}$$

where $u = (\underline{u}, \overline{u}, \delta(u))$, with $e = (\underline{u}, \overline{u})$ and $\delta(u)$ as in Definition 3.3. The choice is motivated by the fact that the resulting semi-distance is an extension of the classic resistance distance.

The construction of Z_E is a bit more complex, as Z_E is piecewise defined on a suitable partition of E. For each $e = (v_1, v_2) \in E_S$, we define the *lifespan* of e, written $\mathrm{ls}(e)$, as the maximal connected set of time instants, t, for which the edge e exists. More formally, $\mathrm{ls}(e)$ is defined as the maximal (with respect to the inclusion partial order) subset of T such that:

1. $t(v_1) \in \mathrm{ls}(e)$;
2. $\mathrm{ls}(e)$ is connected, that is $\forall t_1 < t_2 \in \mathrm{ls}(e)$, $\{t_1, \dots, t_2\} \subseteq \mathrm{ls}(e)$;

3. $\forall t \in \mathrm{ls}(e)$, there exists $(v_1', v_2') \in E_S$ such that $s(v_1') = s(v_1)$, $s(v_2') = s(v_2)$ and $t(v_1') = t(v_2') = t$.

Figure 3.3 allows to visualise the situation. The lifespan of the edge (A, B) at time $t = 0$ is $\{0, 1, 2\}$; the lifespan of (C, E) at time $t = 1$ is $\{0, 1\}$ and the one of (B, C) at time $t = 2$ is $\{2\}$.

We now define the *life* of e (denoted $\mathrm{lf}(e)$) as the set of edges that represent e at different times and have the same lifespan $\mathrm{ls}(e)$, that is:

$$\mathrm{lf}(e) := \left\{(v_1', v_2') \in E_S : s(v_1') = s(v_1),\, s(v_2') = s(v_2),\, t(v_1') = t(v_2') \in \mathrm{ls}(e)\right\}.$$

For convenience, we define the life for temporal edges as well: if $e \in E_T$, $\mathrm{lf}(e) := \{e\}$. The set $\{\mathrm{lf}(e) : e \in E_S\}$ forms a partition of all the spatial edges E_S, and $\{\mathrm{lf}(e) : e \in E_T\}$ is a partition of E_T. For convenience, we set $\Lambda := \left\{\mathrm{lf}(e) : e \in \widetilde{E}\right\}$ the set of all lives and we indicate with λ a generic element of Λ. Finally, we write $u \in \lambda$ whenever the edge containing u belongs to λ.

The main idea is to consider the life of each spatial and temporal edge and define a suitable process on it, being independent from the others. Let us consider a spatial edge $e \in E_S$ and its lifespan $\mathrm{ls}(e)$. Consider now the set $\mathrm{ls}(e) \times [0, 1]$ and define on it a zero-mean Gaussian process $B(t, \delta)$ whose covariance function is given by $k_B((t_1, \delta_1), (t_2, \delta_2)) := k_T(|t_1 - t_2|)\, k_{BB}(\delta_1, \delta_2)$, with $t_1, t_2 \in \mathrm{ls}(e)$ and $\delta_1, \delta_2 \in [0, 1]$. Here k_T is a temporal kernel defined on $\mathbb{N}$ such that $k_T(0) = 1$ and $k_{BB}(\delta_1, \delta_2) := \min(\delta_1, \delta_2) - \delta_1\delta_2$ is the kernel of the standard Brownian bridge on $[0, 1]$. Notice that the spatial marginals of the process B are standard Brownian bridges. We stress that the process $B(t, \delta)$ is only needed for the definition of the process Z_E on $\mathrm{lf}(e)$, denoted $Z_E\big|_{\mathrm{lf}(e)}$, as follows: given an edge $e' = (\underline{u}, \overline{u}) \in \mathrm{lf}(e)$ and given a point $u = (\underline{u}, \overline{u}, \delta)$ on it,

$$Z_E\big|_{\mathrm{lf}(e)}(u) := \sqrt{\ell(e')}\, B(t(\underline{u}), \delta). \tag{3.6}$$

Finally, for each temporal edge $e = [0, \ell(e)] \in E_T$, we define the process Z_E on it as an independent (from both Z_V and Z_E on E_S) Brownian bridge on $[0, \ell(e)]$, having covariance function given by $\mathbb{C}\mathrm{ov}\left(Z_E\big|_e(\delta_1), Z_E\big|_e(\delta_2)\right) = \ell(e)\,(\min(\delta_1, \delta_2) - \delta_1\delta_2)$. This concludes the construction of the process on the whole set of the edges.

3.4.2 Mathematical Properties of the Construction

We remind the reader that the process Z is Gaussian, being the sum of two independent Gaussian processes. Hence, the finite dimensional distributions of Z are completely specified through the second order properties, namely the covariance function. The following result provides an analytical expression for the covariance function associated with Z.

Proposition 3.1. *Let $u_1, u_2 \in \widetilde{\boldsymbol{G}}$, with $u_i = (\underline{u}_i, \overline{u}_i, \delta_i)$, $i = 1, 2$. We recall that δ_i is the relative distance of u_i from $\underline{u}_i$, as stated in Definition 3.3. Let $Z = Z_V + Z_E$,*

with Z_V as defined through (3.5) and Z_E as defined in the previous subsection (see 3.6). Then,

$$k_Z(u_1,u_2) = \boldsymbol{\delta}_1^\top \left(\mathbf{L}^\star\right)^{-1} \boldsymbol{\delta}_2 + \mathbb{1}_{\mathrm{lf}(e_1)=\mathrm{lf}(e_2)} \sqrt{\ell(e_1)\ell(e_2)}\, k_T\left(|t(\underline{u}_1) - t(\underline{u}_2)|\right) \left(\min\left(\delta_1,\delta_2\right) - \delta_1\delta_2\right), \quad (3.7)$$

where $\boldsymbol{\delta}_i$ ($i \in \{1,2\}$) are the $n-$dimensional vectors whose entries are given by

$$\boldsymbol{\delta}_j := \begin{cases} 1-\delta_i & \text{if } j = \underline{u}_i, \\ \delta_i & \text{if } j = \overline{u}_i, \\ 0 & \text{otherwise.} \end{cases} \quad (3.8)$$

We stress that the first line of Equation (3.7) is the covariance function of Z_V, whilst the second line is the covariance of Z_E.

Proof. **of Proposition 3.1**
From the definition of Z_V, we have that

$$\begin{aligned} \mathbb{C}\mathrm{ov}\left(Z_V(u_1), Z_V(u_2)\right) &= \left(1-\delta_1\right)\left(1-\delta_2\right)\mathbf{L}^-[\underline{u}_1,\underline{u}_2] + \left(1-\delta_1\right)\delta_2\mathbf{L}^-[\underline{u}_1,\overline{u}_2] \\ &\quad + \delta_1\left(1-\delta_2\right)\mathbf{L}^-[\overline{u}_1,\underline{u}_2] + \delta_1\delta_2\mathbf{L}^-[\overline{u}_1,\overline{u}_2] \\ &= \boldsymbol{\delta}_1^\top \mathbf{L}^- \boldsymbol{\delta}_2. \end{aligned} \quad (3.9)$$

Let us now consider the process Z_E. For the sake of simplicity, here we set $e_1 := (\underline{u}_1, \overline{u}_1)$ and $e_2 := (\underline{u}_2, \overline{u}_2)$. By construction, the covariance between $Z_E(u_1)$ and $Z_E(u_2)$ is null whenever $\mathrm{lf}(e_2) \neq \mathrm{lf}(e_1)$. Furthermore,

$$\mathbb{C}\mathrm{ov}\left(Z_E(u_1), Z_E(u_2)\right) = \ell(e_1)\left(\min(\delta_1,\delta_2) - \delta_1\delta_2\right)$$

if $e_1 = e_2 \in E_T$ and

$$\mathbb{C}\mathrm{ov}\left(Z_E(u_1), Z_E(u_2)\right) = \sqrt{\ell(e_1)\ell(e_2)}\, k_T\left(|t(\underline{u}_1) - t(\underline{u}_2)|\right)\left(\min\left(\delta_1,\delta_2\right) - \delta_1\delta_2\right)$$

if $e_1, e_2 \in E_S$ and $\mathrm{lf}(e_1) = \mathrm{lf}(e_2)$. By noticing that the covariance expression for former case is actually a special case of the one of the latter (recall that $k_T(0) = 1$), we can summarise the covariance function of Z_E for each couple of points $u_1, u_2 \in \boldsymbol{G}$ as follows:

$$k_{Z_E}(u_1,u_2) = \mathbb{1}_{\mathrm{lf}(e_1)=\mathrm{lf}(e_2)} \sqrt{\ell(e_1)\ell(e_2)}\, k_T\left(|t(\underline{u}_1) - t(\underline{u}_2)|\right)\left(\min\left(\delta_1,\delta_2\right) - \delta_1\delta_2\right). \quad (3.10)$$

□

As an immediate consequence of Proposition 3.1, we obtain the independence of the variogram (thus, of the distance) from the choice of $\boldsymbol{x}$.

Proposition 3.2. *The variogram of Z_V, $(u_1,u_2) \mapsto \gamma_{Z_V}(u_1,u_2)$, does not depend on the choice of $\boldsymbol{x}$.*

In order to prove Proposition 3.2, we need the following Lemma.

Lemma 3.1. *Let* $\mathbf{L}$ *be a* $n \times n$ *Laplacian matrix, and let* $\boldsymbol{x} \in \mathbb{R}^n$ *such that* $\mathbf{1}_n{}^\top \boldsymbol{x} \neq 0$. *Let, in addition,* $\delta_1, \delta_2 \in [0,1]$ *and* $\underline{u}_1, \overline{u}_1, \underline{u}_2, \overline{u}_2 \in \{1, \dots, n\}$. *Finally, define* $\boldsymbol{\delta}_i$ *(*$i \in \{1,2\}$*) as in Equation (3.8). Then it holds:*

$$(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \left(\mathbf{L} + \boldsymbol{x}\boldsymbol{x}^\top\right)^{-1} (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) = (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \mathbf{L}^- (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)$$

Proof of Lemma 3.1. First, let us show that the thesis is meaningful, that is that $\mathbf{L} + \boldsymbol{x}\boldsymbol{x}^\top$ is strictly positive definite. Clearly, it is positive semidefinite, as sum of two positive semidefinite matrices. In addition, for $\boldsymbol{y} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, we have

$$\boldsymbol{y}^\top \left(\mathbf{L} + \boldsymbol{x}\boldsymbol{x}^\top\right) \boldsymbol{y} = \boldsymbol{y}^\top \mathbf{L} \boldsymbol{y} + \boldsymbol{y}^\top \boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{y}.$$

Now, if $\boldsymbol{y}$ does not belong to the kernel of $\mathbf{L}$, that is $\mathbf{L}\boldsymbol{y} \neq 0$, we have that $\boldsymbol{y}^\top \mathbf{L} \boldsymbol{y}$ is strictly positive. If, instead, $\boldsymbol{y} \in \ker(\mathbf{L})$, then, by the properties of Laplacian matrices stated in Subsection 3.2.3, $\boldsymbol{y} = c\,\mathbf{1}_n$, where $c \in \mathbb{R} \setminus \{0\}$. In such a case we have $\boldsymbol{y}^\top \left(\mathbf{L} + \boldsymbol{x}\boldsymbol{x}^\top\right) \boldsymbol{y} = \boldsymbol{y}^\top \mathbf{L} \boldsymbol{y} + \boldsymbol{y}^\top \boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{y} = 0 + c^2 \left(\mathbf{1}_n{}^\top \boldsymbol{x}\right)^2 > 0$, since $\mathbf{1}_n{}^\top \boldsymbol{x} \neq 0$ by hypothesis.

Now, let us show that the equality holds. In order to express the Moore-Penrose inverse (which coincides with the standard inverse) of the rank-1 update $\mathbf{L} + \boldsymbol{x}\,\boldsymbol{x}^\top$, we exploit Theorem 1 of Meyer (1973) with $\mathbf{A} := \mathbf{L}$ and $\boldsymbol{c} := \boldsymbol{d} := \boldsymbol{x}$. Firstly, let us show that the hypotheses of the theorem are satisfied: we have to prove that $\boldsymbol{x} \notin R(\mathbf{L})$, where R denotes the range or column space. $R(\mathbf{L})$ is perpendicular to the kernel of $\mathbf{L}$, that is (recall that $\mathbf{L}$ is a Laplacian matrix) $\ker(\mathbf{L}) = \{h\,\mathbf{1}_n : h \in \mathbb{R}\}$. Now, $\boldsymbol{x} \in R(\mathbf{L}) \iff \boldsymbol{x} \perp \ker(\mathbf{L}) \iff \mathbf{1}_n{}^\top \boldsymbol{x} = 0$, yet $\mathbf{1}_n{}^\top \boldsymbol{x} \neq 0$ by hypothesis. Therefore $\boldsymbol{x} \notin R(\mathbf{L})$, *i.e.* we can apply the above-mentioned theorem. In the following computations, we have adopted the notation used by Meyer (1973, Section 2), *i.e.*:

$$\begin{aligned}
&\boldsymbol{k} := \mathbf{L}^- \boldsymbol{x} && \boldsymbol{h} := \boldsymbol{x}^\top \mathbf{L}^- \\
&\boldsymbol{u} := \left(\mathbf{I}_n - \mathbf{L}\mathbf{L}^-\right) \boldsymbol{x} = \frac{1}{n} \mathbf{1}_{n \times n} \boldsymbol{x} && \boldsymbol{v} := \boldsymbol{x}^\top \left(\mathbf{I}_n - \mathbf{L}^- \mathbf{L}\right) = \frac{1}{n} \boldsymbol{x}^\top \mathbf{1}_{n \times n} \\
&\beta := 1 + \boldsymbol{x}^\top \mathbf{L}^- \boldsymbol{x}.
\end{aligned}$$

In addition, we have $\|\boldsymbol{u}\|^2 = \|\boldsymbol{v}\|^2 = \boldsymbol{u}^\top \boldsymbol{u} = \frac{1}{n^2} \mathbf{1}_n{}^\top (\mathbf{1}_n{}^\top \boldsymbol{x})(\mathbf{1}_n{}^\top \boldsymbol{x}) \mathbf{1}_n = \frac{1}{n} (\mathbf{1}_n{}^\top \boldsymbol{x})^2 \neq 0$. Let us finally apply Equation (3.1) of Meyer (1973):

$$\begin{aligned}
\left(\mathbf{L} + \boldsymbol{x}\boldsymbol{x}^\top\right)^- &= \mathbf{L}^- - \boldsymbol{k}\boldsymbol{u}^- - \boldsymbol{v}^- \boldsymbol{h} + \beta \boldsymbol{v}^- \boldsymbol{u}^- \\
&= \mathbf{L}^- - \frac{\boldsymbol{k}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2} - \frac{\boldsymbol{v}^\top \boldsymbol{h}}{\|\boldsymbol{v}\|^2} + \beta \frac{\boldsymbol{v}^\top \boldsymbol{u}^\top}{\|\boldsymbol{v}\|^2 \|\boldsymbol{u}\|^2} \\
&= \mathbf{L}^- - \frac{\mathbf{L}^- \boldsymbol{x} \mathbf{1}_n{}^\top}{\mathbf{1}_n{}^\top \boldsymbol{x}} - \frac{\mathbf{1}_n \boldsymbol{x}^\top \mathbf{L}^-}{\mathbf{1}_n{}^\top \boldsymbol{x}} + \left(1 + \boldsymbol{x}^\top \mathbf{L}^- \boldsymbol{x}\right) \frac{\mathbf{1}_{n \times n}}{\left(\mathbf{1}_n{}^\top \boldsymbol{x}\right)^2}.
\end{aligned}$$

To conclude, it is now sufficient to prove that

$$(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \left(\left(1 + \boldsymbol{x}^\top \mathbf{L}^- \boldsymbol{x}\right) \frac{\mathbf{1}_{n\times n}}{\left(\mathbf{1}_n{}^\top \boldsymbol{x}\right)^2} - \frac{\mathbf{L}^- \boldsymbol{x} \mathbf{1}_n{}^\top}{\mathbf{1}_n{}^\top \boldsymbol{x}} - \frac{\mathbf{1}_n \boldsymbol{x}^\top \mathbf{L}^-}{\mathbf{1}_n{}^\top \boldsymbol{x}} \right) (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) = 0.$$

It is possible to see this by noticing that $\mathbf{1}_n{}^\top (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) = \mathbf{1}_n{}^\top \boldsymbol{\delta}_1 - \mathbf{1}_n{}^\top \boldsymbol{\delta}_2 = 1 - 1 = 0$ and similarly $(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \mathbf{1}_n = 0$, therefore each term of the above equation equals 0. □

Proof. **of Proposition 3.2** The proof relies on Lemma 3.1. By the proof of Proposition 3.1, we have:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(Z_V(u_1) - Z_V(u_2)\right) &= k_{Z_V}(u_1, u_1) + k_{Z_V}(u_2, u_2) - 2k_{Z_V}(u_1, u_2) \\
&= \boldsymbol{\delta}_1^\top (\mathbf{L}^\star)^{-1} \boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^\top (\mathbf{L}^\star)^{-1} \boldsymbol{\delta}_2 - 2\boldsymbol{\delta}_1^\top (\mathbf{L}^\star)^{-1} \boldsymbol{\delta}_2 \\
&= (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top (\mathbf{L}^\star)^{-1} (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) \\
&= (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \mathbf{L}^- (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2),
\end{aligned}$$

where the last step follows from Lemma 3.1. Notice that the last expression does not depend on $\boldsymbol{x}$, therefore for all values of $\boldsymbol{x}$ the variogram is the same. This settles the proof. □

Remark 3.3. Notice that the process Z_E is defined on all the vertices V (it is zero) and that the expression (3.10) is meaningful even when any of the points u_1 and u_2 belongs to V. Indeed, if, say, $u_1 \in V$, then $\delta_1 \in \{0, 1\}$ regardless of which incident edge (u, v) is taken in the expression $u_1 = (u, v, \delta)$. As a consequence, the last factor in (3.10) vanishes and the covariance is therefore null. Finally, notice that, since Z_V and Z_E are independent, the covariance function of Z is simply the sum of (3.9) and (3.10).

While noting that this construction is completely general, we also point out that conditional independence properties, whenever aimed, can be achieved through a proper choice of the temporal kernel k_T. A reasonable choice for k_T is the correlation function of an autoregressive process of order one: $k_T(h) = \phi^{|h|}$, where $\phi \in (-1, 1)$ is a free parameter and $h \in \mathbb{Z}$ is the lag. Notice that the special case $\phi = 0$, for which $k_T(h) = \mathbb{1}_{h=0}$, corresponds to the static resistance metric provided by Anderes et al. (2020). Figure 3.4 depicts some realisations for the process Z_E over an edge for different values of the parameter ϕ.

The parameter ϕ for the edges plays a similar role of the parameter α for the nodes: their are both closely related to the inter-dependency of the process at different times. Indeed, ϕ measures how much the process Z_E is correlated between two times $t_1, t_2 \in \mathrm{ls}(e)$. Analogously, the value α as weight of the temporal edge $e \in E_T$, is related to the partial correlation of the endpoints of e given everything else. As a consequence, it is natural to choose a high value of ϕ for high values of α and vice-versa. We notice that it is natural to choose non-negative values for ϕ, as we

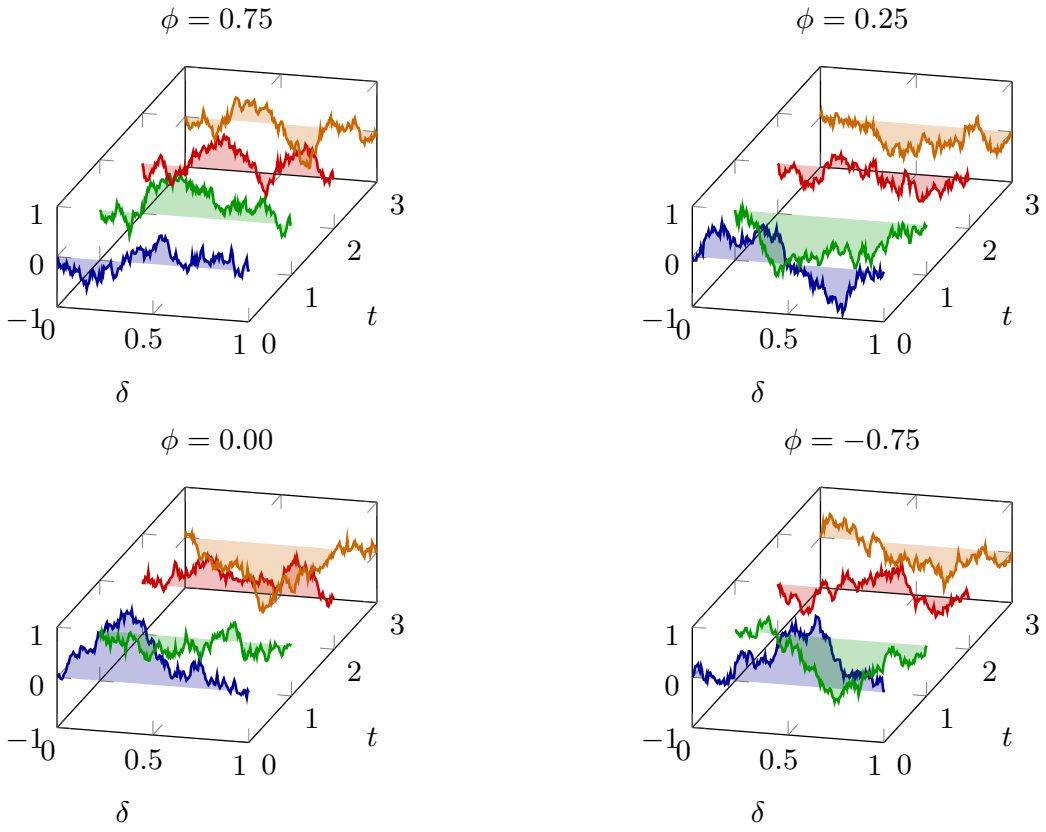


Figure 3.4: Draws from the process Z_E on an edge with lifespan $\{0, 1, 2, 3\}$, for several values of the parameter ϕ.

usually expect a non-negative correlation between the values of Z_E for close times. Using equation (3.4), we get the following expression for the semi-distance between any to points $u_1, u_2 \in \widetilde{\boldsymbol{G}}$.

$$d(u_1, u_2) = k_Z(u_1, u_1) + k_Z(u_2, u_2) - 2k_Z(u_1, u_2). \tag{3.11}$$

The formal statement below provides a complete description of the space $(\widetilde{\boldsymbol{G}}, d)$, that is, the time-evolving graph $\widetilde{\boldsymbol{G}}$ equipped with the semi-distance d.

Proposition 3.3. *Let d be the mapping defined at (3.11). Then, the pair $(\widetilde{\boldsymbol{G}}, d)$ is a semi-distance space.*

Proof. **of Proposition 3.3** Symmetry, non-negativeness and the implication $u_1 = u_2 \implies d(u_1, u_2) = 0$ follow immediately from (3.4). Therefore, we just need to show that $d(u_1, u_2) = 0 \implies u_1 = u_2$. From (3.4), if $d(u_1, u_2) = 0$, then $Z(u_1) = Z(u_2)$ almost surely. As a consequence:

$$Z_V(u_1) - Z_V(u_2) = -Z_E(u_1) + Z_E(u_2).$$

Being Z_V and Z_E independent, necessarily $Z_V(u_1) - Z_V(u_2) = 0$ and $-Z_E(u_1) + Z_E(u_2) = 0$, hence $Z_V(u_1) = Z_V(u_2)$ and $Z_E(u_1) = Z_E(u_2)$ a.s.. Now $Z_V(u_1)$ and $Z_V(u_2)$ are linear combinations of $\boldsymbol{Z}_V(V)$, that is: $Z_V(u_1) = \boldsymbol{x}_1^\top \boldsymbol{Z}_V(V)$ and $Z_V(u_2) = \boldsymbol{x}_2^\top \boldsymbol{Z}_V(V)$ for some $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^N$, where $N = |V|$. More specifically, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ have the following structure (here we assume that the vertices V are ordered, so that $\underline{u}_i$ comes before $\overline{u}_i$, for $i \in \{1, 2\}$):

$$\begin{cases} \boldsymbol{x}_1^\top = \begin{bmatrix} \mathbf{0}^\top & 1 - \delta_e(u_1) & \mathbf{0}^\top & \delta_e(u_1) & \mathbf{0}^\top \end{bmatrix} \\ \boldsymbol{x}_2^\top = \begin{bmatrix} \mathbf{0}^\top & 1 - \delta_e(u_2) & \mathbf{0}^\top & \delta_e(u_2) & \mathbf{0}^\top \end{bmatrix}, \end{cases}$$

where the $\mathbf{0}$'s represent vectors of zeroes of the appropriate length (possibly zero). Since $Z_V(u_1) = Z_V(u_2)$ a.s., it follows that $(\boldsymbol{x}_1 - \boldsymbol{x}_2)^\top Z_V(V) = 0$ a.s., that is: $\boldsymbol{x}_1 - \boldsymbol{x}_2 = \lambda \mathbf{1}_N$ for some $\lambda \in \mathbb{R}$. Hence,

$$\lambda = \lambda \frac{\mathbf{1}_N^\top \mathbf{1}_N}{N} = \frac{1}{N} \mathbf{1}_N^\top (\boldsymbol{x}_1 - \boldsymbol{x}_2) = \frac{1}{N} \left(\mathbf{1}_N^\top \boldsymbol{x}_1 - \mathbf{1}_N^\top \boldsymbol{x}_2 \right) = 0.$$

This means that $\boldsymbol{x}_1 = \boldsymbol{x}_2$, that is $u_1 = u_2$. □

One might ask whether a stronger assertion holds for the pair $(\widetilde{\boldsymbol{G}}, d)$ as defined above. The next statement provides a negative answer. A counterexample is given by the graph depicted in Figure 3.5

Proposition 3.4. *Let d be the mapping defined at (3.11). Then, the pair $(\widetilde{\boldsymbol{G}}, d)$ is not, in general, a metric space.*

Proof. **of Proposition 3.4** Consider the equivalent simple graph represented in Figure 3.5, where all the weights of are 1, exception made for the edges (A_0, B_0) and A_2, B_2, which have weights ε and $\frac{1}{\varepsilon}$ respectively, for a sufficiently small $\varepsilon > 0$. Considering the vertices in the order $A_0, B_0, A_1, \ldots, B_2$, the Laplacian matrix $\mathbf{L}$ is

$$\mathbf{L} = \begin{bmatrix} 1+\varepsilon & -\varepsilon & -1 & 0 & 0 & 0 \\ -\varepsilon & 1+\varepsilon & 0 & -1 & 0 & 0 \\ -1 & 0 & 3 & -1 & -1 & 0 \\ 0 & -1 & -1 & 3 & 0 & -1 \\ 0 & 0 & -1 & 0 & 1+\frac{1}{\varepsilon} & -\frac{1}{\varepsilon} \\ 0 & 0 & 0 & -1 & -\frac{1}{\varepsilon} & 1+\frac{1}{\varepsilon} \end{bmatrix}.$$

Let now consider the points $P := \left(0, A_0, B_0, \frac{1}{2}\right)$, $Q := \left(1, A_1, B_1, \frac{1}{2}\right)$ and $R := \left(2, A_2, B_2, \frac{1}{2}\right)$. We will show that, for any $\gamma > 0$, for ε sufficiently small, it holds

$$d(P, Q) + d(Q, R) < d(P, R). \tag{3.12}$$

First, let us rewrite and simplify a bit (3.12). Notice that here all the δ's are $\frac{1}{2}$.

$$(3.12) \iff k_Z(P, P) + k_Z(Q, Q) - 2k_Z(P, Q)$$

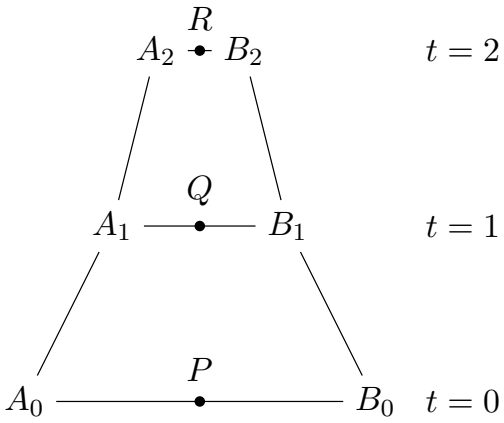


Figure 3.5: An example of equivalent simple graph for which the semi-distance defined at (3.11) does not satisfy the triangle inequality. Here the length of the top edge (A_2, B_2) becomes vanishingly small, while length of the bottom edge (A_0, B_0) grows to infinity.

$$
\begin{aligned}
&+ k_Z(Q,Q) + k_Z(R,R) - 2k_Z(Q,R) \\
&< k_Z(P,P) + k_Z(R,R) - 2k_Z(P,R) \\
\iff & k_Z(Q,Q) - k_Z(P,Q) - k_Z(Q,R) < -k_Z(P,R) \\
\iff & \frac{1}{4}\mathbf{1}_2^\top \mathbf{L}^- \left[(A_1,B_1),(A_1,B_1)\right] \mathbf{1}_2 + \sqrt{1\cdot 1}\,\gamma^0 \cdot \frac{1}{4} \\
&- \frac{1}{4}\mathbf{1}_2^\top \mathbf{L}^- \left[(A_0,B_0),(A_1,B_1)\right] \mathbf{1}_2 - \sqrt{\frac{1}{\varepsilon}\cdot 1}\,\gamma^1 \cdot \frac{1}{4} \\
&- \frac{1}{4}\mathbf{1}_2^\top \mathbf{L}^- \left[(A_1,B_1),(A_2,B_2)\right] \mathbf{1}_2 - \sqrt{1\cdot \varepsilon}\,\gamma^1 \cdot \frac{1}{4} \\
&< -\frac{1}{4}\mathbf{1}_2^\top \mathbf{L}^- \left[(A_0,B_0),(A_2,B_2)\right] \mathbf{1}_2 - \sqrt{\frac{1}{\varepsilon}\cdot \varepsilon}\,\gamma^2 \cdot \frac{1}{4} \\
\iff & \mathbf{1}_2^\top \left(\mathbf{L}^- \left[(A_1,B_1),(A_1,B_1)\right] + \mathbf{L}^- \left[(A_0,B_0),(A_2,B_2)\right]\right) \mathbf{1}_2 \\
&+ \mathbf{1}_2^\top \left(-\mathbf{L}^- \left[(A_0,B_0),(A_1,B_1)\right] - \mathbf{L}^- \left[(A_1,B_1),(A_2,B_2)\right]\right) \mathbf{1}_2 \\
&< -1 + \frac{\gamma}{\sqrt{\varepsilon}} + \gamma\sqrt{\varepsilon} - \gamma^2.
\end{aligned}
\tag{3.13}
$$

Notice that the right-hand size of the last inequality (3.13) is not limited for any $\gamma > 0$ when $\varepsilon \to 0^+$. As a consequence, it is sufficient to show that the left-hand size is limited when $\varepsilon \to 0^+$. Indeed the left-hand side of the last inequality is a sheer signed sum of 16 elements of the matrix $\mathbf{L}^-$. This sum is surely not greater than

$$
16 \max_{i,j\in\{1,\ldots,6\}} \left|\mathbf{L}^-[i,j]\right| \leq 16 \max_{i\in\{1,\ldots,6\}} \left|\mathbf{L}^-[i,i]\right|.
$$

Now, in our case, the main diagonal of $\mathbf{L}^-$ is given by:

$$\operatorname{diag}(\mathbf{L}^-) = \left[\frac{10\varepsilon^2+39\varepsilon+34}{36\,(\varepsilon^2+3\varepsilon+1)}, \frac{10\varepsilon^2+39\varepsilon+34}{36\,(\varepsilon^2+3\varepsilon+1)}, \frac{10\varepsilon^2+27\varepsilon+10}{36\,(\varepsilon^2+3\varepsilon+1)}, \frac{10\varepsilon^2+27\varepsilon+10}{36\,(\varepsilon^2+3\varepsilon+1)}, \frac{34\varepsilon^2+39\varepsilon+10}{36\,(\varepsilon^2+3\varepsilon+1)}, \frac{34\varepsilon^2+39\varepsilon+10}{36\,(\varepsilon^2+3\varepsilon+1)}\right].$$

As all the entries are continuous functions of $\varepsilon \in [0,1]$, they are limited. Since the maximum of (a finite number of) limited functions on the same domain is limited, the left-hand of (3.13) is limited as well. This concludes the proof. □

Remark 3.4. Although in general our extension to the classic resistance distance is not a metric, it retains some of its properties: (i) $(V, d\big|_V)$ is a metric space and (ii) for all t, $(G_t, d\big|_{G_t})$ coincides with the restriction on G_t of the resistance metric of Anderes et al. (2020) computed on the whole graph $\widetilde{\boldsymbol{G}}$, hence it is a metric and it is invariant to splitting edges and merging edges at degree 2 vertices (Anderes et al., 2020, Propositions 2 and 3).

3.4.3 Time-Evolving Linear Networks

Anderes et al. (2020) defined graphs with Euclidean edges as a generalisation of linear networks, and Euclidean trees with a given number of leaves. For both cases, edges are linear. This case is not especially interesting for the framework proposed here. The reason is that a simple isometric embedding arguments as in Tang and Zimmerman (2024) proves that one can embed a time-evolving linear network in $\mathbb{R} \times \mathbb{R}^2 = \mathbb{R}^3$, where the first component indicates time. As a consequence, it is immediate to build a vast class of covariance functions on a time-evolving linear network by a sheer restriction of a given covariance function defined on $\mathbb{R}^3$. However, such a method does not take into account the *structure* of the graph: two points that are close in $\mathbb{R}^3$ but far in the time-evolving graph could have a high correlation. Section 3.7 illustrates how to build kernels over the special topologies proposed in this Chapter. Apparently, the choices are more restrictive than the ones available for the case of linear networks, but they ensure that the spatio-temporal structure is taken into account.

3.5 Circular Time and Periodic Graphs

Perhaps the main drawback of using the resistance semi-distance in the layer graphs that express the spatio-temporal variability is that, when adding one or more new layers, the semi-distances between the points of the previous layers may change (decrease). Indeed, whenever new paths between a couple of points are added, the effective semi-distance between such points decreases, as the current meets less resistance. This presents a critical interpretation problem: for a given time series, let new data be added on a daily basis. Then, inference routines may provide different

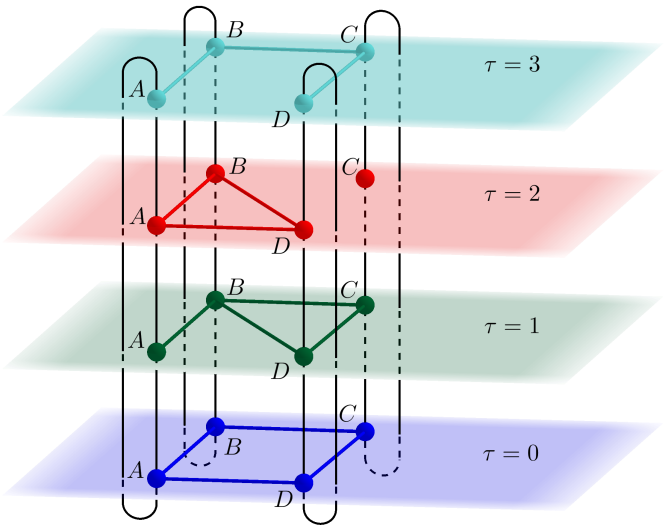


Figure 3.6: An example of an equivalent simple graph for a periodic time-evolving graph with $m = 4$ and $S = \{A, B, C, D\}$. The coloured edges belong to E_S, whilst the black ones belong to E_T.

results when compared to the results of same inference techniques applied to the updated time series. Indeed, as the semi-distances may vary, the covariances between the same space-time points may vary as well.

Here, we consider time-evolving *periodic* networks, *i.e.* time-evolving networks whose evolution repeats after a fixed amount of time instants (number of layers). Not only does this construction solve the above-mentioned issue, but it also suits many phenomena whose evolution present both linear and periodic components.

Definition 3.7 (Time-evolving periodic graph)**.** Let $\boldsymbol{G} = \{G_0, G_1, \ldots\}$ be a countable sequence of graphs. Then, $\boldsymbol{G}$ is a *time-evolving periodic graph* if there exists a natural number $m \geq 3$ such that, for all $t \in \mathbb{N}$, $G_t = G_{t+m}$.

Its equivalent simple periodic graph $\widetilde{\boldsymbol{G}}$ is built by connecting $G_0, \ldots, G_{m-1}$ via a proper set of temporal edges E_T. A special case is when the set of edges is

$$E_T = \{(v_1, v_2) \in V \times V \,:\, s(v_1) = s(v_2),\, |t(v_1) - t(v_2)| \equiv \pm 1 \pmod{m}\},$$

which provides a temporally-complete periodic graph of order 1.

Each point in the resulting space-time is denoted by its true time $t \in \mathbb{R}_0^+$, by the endpoints of the edge e it lies on and by the relative distance $\delta(u)$ from the first one: we write $u = (t, \underline{u}, \overline{u}, \delta(u))$, where $e = (\underline{u}, \overline{u})$. Notice that $t \in \mathbb{N}$ whenever u belongs to a temporal layer, while t is not integer if u belongs to the inner part of a temporal edge $e \in E_T$. Given a point $u \in V \cup \bigcup E_S$, we sometimes write $\tau(u)$ as the unique layer $\tau \in \{0, \ldots, m-1\}$ that contains u, *i.e.* $\tau(u) \equiv t(u) \pmod{m}$.

We start by noting that the previously-mentioned issue about linear time-evolving graphs is overcome by this construction. Indeed, once the full periodic structure

has been established, the Laplacian matrix needs be computed only once, regardless of how many new time points are added. A second remark comes from the metric construction, which necessarily needs to be adapted to a periodic process. Otherwise, some counter-intuitive properties can arise. Suppose the semi-distance $d(u_1, u_2)$ is defined as in (3.4). Then, for any couple of points $u_1 = (t_1, \underline{u}_1, \overline{u}_1, \delta_e(u_1))$ and $u_2 = (t_2, \underline{u}_2, \overline{u}_2, \delta_e(u_2))$ with $\underline{u}_1 = \underline{u}_2$, $\overline{u}_1 = \overline{u}_2$ and $\delta_e(u_1) = \delta_e(u_2)$, even when $t_1 \neq t_2$, the semi-distance would be identically equal to zero. Hence, a different definition for the process Z is necessary.

For a point $u = (t, \underline{u}, \overline{u}, \delta(u)) \in \widetilde{\boldsymbol{G}}$, we define the process Z for the periodic graph as $Z(u) := Z_V(u) + Z_E(u) + \beta W(t)$, where $\beta > 0$ is a given parameter, W is a standard Wiener process, Z_V is the same process as in the linear-time case, while Z_E, albeit similar, presents some difference with respect to the construction given in Subsection 3.4.1, aimed to capture the time structure of the periodic graph.

Let $e = (v_1, v_2) \in E_S$: we define the *lifespan* of e as the maximal subset of $\{0, \dots, m-1\}$ such that:

1. $\tau(v_1) \in \mathrm{ls}(e)$;
2. $\mathrm{ls}(e)$ is connected, *i.e.* if $\tau_1 < \tau_2 \in \mathrm{ls}(e)$, then $\{\tau_1, \dots, \tau_2\} \subseteq \mathrm{ls}(e)$ or $\{\tau_2, \dots, m-1, 0, \dots, \tau_1\} \subseteq \mathrm{ls}(e)$
3. $\forall \tau \in \mathrm{ls}(e)$, $\exists (v_1', v_2') \in E_S$ such that $s(v_1') = s(v_1)$, $s(v_2') = s(v_2)$ and $\tau(v_1') = \tau(v_2') = \tau$.

Figure 3.6 depicts this situation. Here, the lifespan of the edge (A, B) at time $\tau = 0$ is $\{0, 1, 2, 3\}$; the lifespan of (A, D) at time $\tau = 2$ is $\{2\}$; the lifespan of (C, D) at time $\tau = 3$ is $\{0, 1, 3\}$. The definition of life of any edge $e \in E$ remains unchanged: if $e \in E_S$,

$$\mathrm{lf}(e) := \big\{(v_1', v_2') \in E_S : s(v_1') = s(v_1),\, s(v_2') = s(v_2),\, \tau(v_1') = \tau(v_2') \in \mathrm{ls}(e)\big\}$$

while, if $e \in E_T$, $\mathrm{lf}(e) := \{e\}$. The definition of Z_E is now identical to the one of the linear-time graph, except the choice of the temporal kernel k_T. Indeed, we ought to consider that the time is now cyclic in the dependence structure of the temporal layers $\tau \in \{0, ..., m-1\}$. It is reasonable to model the process underlying the temporal kernel k_T by means of a graphical model, as it embodies the idea of conditional independence. For a given spatial edge $e \in E_S$, we distinguish two cases: whether the lifespan of e is the whole temporal set $T = \{0, ..., m-1\}$ or not.

3.5.1 The Lifespan Coincides with T

In this case, we define the covariance matrix of a zero-mean Gaussian random vector $Z_T : \{0, ..., m-1\} \to \mathbb{R}$ via its precision matrix. More precisely, let G_T be the circulant graph with m nodes (labelled by $\tau \in \{0, ..., m-1\}$) and m edges between adjacent nodes, as shown in Figure 3.7. We associate each edge with a given weight

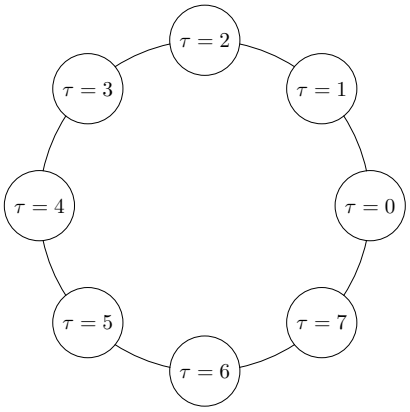


Figure 3.7: Conditional dependence structure of the process Z_T for $m = 8$.

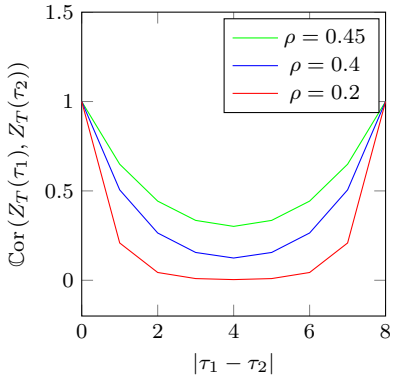
$\rho \in \left[0, \frac{1}{2}\right)$, which represents the partial correlation between subsequent times. As a consequence, the precision matrix $\mathbf{\Theta}_{Z_T}$ is the circulant matrix

$$\mathbf{\Theta}_{Z_T} = \kappa \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 & -\rho \\ -\rho & 1 & -\rho & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\rho \\ -\rho & 0 & 0 & \dots & -\rho & 1 \end{bmatrix}.$$

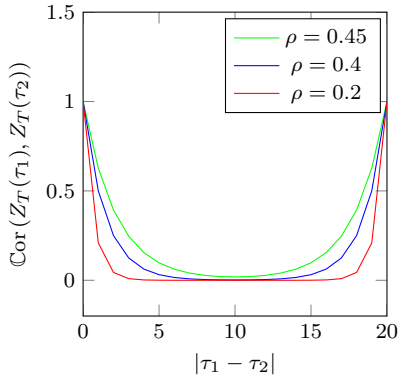
Here, $\kappa > 0$ is a normalising constant whose role is to make the covariance matrix $\mathbf{\Sigma}_{Z_T} := (\mathbf{\Theta}_{Z_T})^{-1}$ a correlation matrix (namely the variances of every entry of Z_T should be 1). Notice that the matrix $\mathbf{\Sigma}_{Z_T}$ is a symmetric circulant matrix: as a consequence, it is possible to store only its first column, which will be denoted by $\sigma_{Z_T} \in \mathbb{R}^m$. In Figure 3.8, the values of the vector σ_{Z_T} are plotted for some values of m and ρ. The next Figure show two examples of correlation values σ_{Z_T}, as defined in Subsection 3.5.1, for some values of the parameter ρ and for two periods ($m = 8, 20$).

3.5.2 The Lifespan does not Coincide with T

In this case, the life of the edge e is interrupted. Thus, it is reasonable to consider the different parts of the life of e as independent. To this aim, we consider the subgraph of G_T that represents the evolution of the edge e. More precisely, we remove from G_T all the nodes τ for which the edge e does not exist and we remove from G_T all the edges whose at least one endpoint has been eliminated. Next, we consider all the connected components of the so-obtained graph and define an autoregressive model on each of them, independently from the others (similarly to the linear case of



The correlations of Z_T for some values of ρ and $m = 8$.



The correlations of Z_T for some values of ρ and $m = 20$.

Figure 3.8: Some examples of the correlation functions k_T in the case $\mathrm{ls}(e) = \{0, \dots, m-1\}$.

Section 3.4.2). More precisely, we define the covariance matrix of the process Z_T as a block-diagonal matrix whose diagonal blocks are of the form

$$\begin{bmatrix} 1 & \phi & \dots & \phi^{j-1} \\ \phi & 1 & \dots & \phi^{j-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{j-1} & \phi^{j-2} & \dots & 1 \end{bmatrix},$$

being $\phi \in (-1, 1)$ the lag-1 correlation and j the number of times τ that belong to $\mathrm{ls}(e)$, *i.e.* $j := |\mathrm{ls}(e)|$.

3.5.3 Second-Order Properties of Z in the Circular Case

The following result illustrates the analytic expression for the covariance function associated with Z in the construction of the distance associated with $\widetilde{\boldsymbol{G}}$ in the periodic case.

Proposition 3.5. *Let $u_1 = (t_1, \underline{u}_1, \overline{u}_1, \delta_e(u_1)) \in \widetilde{\boldsymbol{G}}$ and $u_2 = (t_2, \underline{u}_2, \overline{u}_2, \delta_e(u_2)) \in \widetilde{\boldsymbol{G}}$. Then the kernel of the process Z defined on $\widetilde{\boldsymbol{G}}$ enjoys the following representation:*

$$\begin{aligned} k_Z(u_1, u_2) = {} & \boldsymbol{\delta}_1^\top \left(\mathbf{L}^\star\right)^{-1} \boldsymbol{\delta}_2 + \beta^2 \min(t_1, t_2) \\ & + \mathbb{1}_{\mathrm{lf}(e_1)=\mathrm{lf}(e_2)} \sqrt{\ell(e_1)\ell(e_2)}\, k_T\left(\tau(\underline{u}_1), \tau(\underline{u}_2)\right) \left(\min\left(\delta_1, \delta_2\right) - \delta_1\delta_2\right) \quad (3.14) \end{aligned}$$

where $\boldsymbol{\delta}_i$ ($i \in \{1, 2\}$) are defined in Equation (3.8).

Notice that the unique differences with the expression (3.7) are the different choices for the temporal kernel k_T and the additional addend $\beta^2 \min(t_1, t_2)$. The

latter ensures that the same points at different times have a strictly positive semi-distance, as, combining equations (3.11) and (3.14) for such points u_1 and u_2, we get $d(u_1, u_2) = \beta^2 |t_1 - t_2|$. We conclude this section with a formal assertion regarding the mapping d as being introduced for the case of a periodic graph $\widetilde{\boldsymbol{G}}$.

Proposition 3.6. *$(\widetilde{\boldsymbol{G}}, d)$ is a semi-distance space.*

Proof. **of Proposition 3.6**
The proof is very similar to the one of Proposition 3.3. Also in this case, we get symmetry, non-negativeness and the implication $u_1 = u_2 \implies d(u_1, u_2) = 0$ immediately from (3.4). Let us show that $d(u_1, u_2) = 0 \implies u_1 = u_2$. From (3.4), if $d(u_1, u_2) = 0$, then $Z(u_1) = Z(u_2)$ almost surely. As a consequence:

$$Z_V(u_1) - Z_V(u_2) = -Z_E(u_1) + Z_E(u_2) - \beta W(t_1) + \beta W(t_2).$$

Since Z_V is independent from Z_E and W, it must be $Z_V(u_1) = Z_V(u_2)$ a.s.. Following the same argument of the proof of Proposition 3.3, we obtain $(\underline{u}_1, \overline{u}_1, \delta_1) = (\underline{u}_2, \overline{u}_2, \delta_2)$. It remains to be shown that $t_1 = t_2$. Again, using $Z(u_1) = Z(u_2)$, and the independence between W_t and both Z_V and Z_E, we obtain $\beta W(t_1) = \beta W(t_2)$ a.s., that is $t_1 = t_2$. □

3.6 Reproducing Kernel Hilbert Space Construction

In this section, we provide a constructive definition of the reproducing kernel Hilbert space (RKHS) for the kernel k_Z of the process $Z = Z_V + Z_E$. Given its constructive nature, we give full details and then summarise them in Proposition 3.7 at the end of this Section. We divide the construction in four steps (plus one).

Step 1: Construction of the RKHS of Z_E on a Life λ
Recall from Subsection 3.4.1 that $\Lambda := \left\{ \mathrm{lf}(e) \,:\, e \in \widetilde{E} \right\}$ is the set of lives of all the edges of $\widetilde{\boldsymbol{G}}$. In this step, we characterise the RKHS of the kernel k_E on any $\lambda \in \Lambda$. For the sake of simplicity, we set $n_\lambda := |\lambda|$: the number of edges that share the life λ. Recall that, for a life λ, $k_E\big|_\lambda : \widetilde{\boldsymbol{G}}\big|_\lambda \times \widetilde{\boldsymbol{G}}\big|_\lambda \to \mathbb{R}$ has the following expression:

$$k_E\big|_\lambda(u_1, u_2) = \sqrt{\ell(e_1)\ell(e_2)} \cdot k_T(|t(e_1) - t(e_2)|) \cdot (\min(\delta_1, \delta_2) - \delta_2\delta_2)\,.$$

It is patent that such a kernel is the product of a kernel defined on a finite set (λ) and the kernel k_{BB}, defined on $[0, 1]$. As a consequence, following (Berlinet and Thomas-Agnan, 2004, Theorem 13), in order to find the RKHS of $k_E\big|_\lambda$, it is sufficient to build the tensor Hilbert product between the RKHSs of

$$k_\lambda^{(M)} : \lambda^2 \to \mathbb{R} \qquad k_\lambda^{(M)}(e_1, e_2) := \sqrt{\ell(e_1)\ell(e_2)} \cdot k_T(|t(e_1) - t(e_2)|), \quad (3.15)$$

$$k_{BB} : [0, 1]^2 \to \mathbb{R} \qquad k_{BB}(\delta_1, \delta_2) := \min(\delta_1, \delta_2) - \delta_2\delta_2. \quad (3.16)$$

Regarding the former, since λ has finite cardinality n_λ, (3.15) can be expressed in the following matrix form:

$$k_\lambda^{(M)}(e_1, e_2) := [\mathbf{M}_\lambda]_{e_1,e_2},$$

where $\mathbf{M}_\lambda \in \mathbb{R}^{n_\lambda \times n_\lambda}$ is defined as

$$[\mathbf{M}_\lambda]_{e_1,e_2} := \sqrt{\ell(e_1)\ell(e_2)} \cdot k_T(|t(e_1) - t(e_2)|), \qquad e_1, e_2 \in \lambda.$$

By introducing the definitions

$$\mathbf{K}_T := [k_T(|t(e_1) - t(e_2)|)]_{e_1,e_2\in\lambda} \in \mathbb{R}^{n_\lambda \times n_\lambda}$$
$$\boldsymbol{w}^\top := \left[\sqrt{\ell(e_1)}, \ldots, \sqrt{\ell(e_{n_\lambda})}\right] \in \mathbb{R}^{n_\lambda},$$

it is possible to compactly write $\mathbf{M}_\lambda$ as

$$\mathbf{M}_\lambda = \mathbf{K}_T \circ \boldsymbol{w}\boldsymbol{w}^\top = \operatorname{diag}(\boldsymbol{w})\, \mathbf{K}_T\, \operatorname{diag}(\boldsymbol{w}),$$

where $\circ$ denotes the Hadamard matrix product. This last expression shows that, if we assume k_T to be a strictly positive definite kernel (and, thus, $\mathbf{K}_T$ is strictly positive definite), then $\mathbf{M}_\lambda$ is strictly positive definite as well. This step is crucial in the definition of the RKHS of $k_\lambda^{(M)}$, which is given next.

Remark 3.5. The RKHS of $k_\lambda^{(M)}$ is $\mathcal{H}_\lambda^{(M)} := \left(\mathbb{R}^{n_\lambda}, \langle\cdot,\cdot\rangle_\lambda^{(M)}\right)$, where

$$\langle v_1, v_2\rangle_\lambda^{(M)} := v_1^\top \mathbf{M}_\lambda^{-1} v_2.$$

This is proved by noticing that $\langle\cdot,\cdot\rangle_\lambda^{(M)}$ is clearly symmetric, bilinear and positive definite, and that

$$\langle k_\lambda^{(M)}(v_1, \cdot), v_2\rangle_\lambda^{(M)} = \langle v_1^\top \mathbf{M}_\lambda, v_2\rangle_\lambda^{(M)} = v_1^\top \mathbf{M}_\lambda \mathbf{M}_\lambda^{-1} v2 = v_1^\top v_2.$$

Next, we need to characterise the RKHS of the kernel k_{BB} in (3.16). By the same argument in the proof of (Anderes et al., 2020, Lemma 3.B), the RKHS of k_{BB} is $\mathcal{H}_{BB} := (\mathcal{F}_{BB}, \langle\cdot,\cdot\rangle_{BB})$, where $\mathcal{F}_{BB}$ is the set of functions $f : [0,1] \to \mathbb{R}$ that are absolutely continuous and such that $f' \in L^2([0,1])$, and

$$\langle f, g\rangle_{BB} := \int_0^1 f'(\delta) g'(\delta)\, \mathrm{d}\delta.$$

Now we have all the ingredients to characterise the RKHS of $k_E\big|_\lambda$.

Remark 3.6. Consider the Hilbert tensor product space $\mathcal{H}_\lambda := \mathcal{H}_\lambda^{(M)} \otimes \mathcal{H}_{BB}$, that is: the completion of the set of functions

$$f_\lambda^{(M)} \otimes f_{BB} : \lambda \times [0,1] \to \mathbb{R}$$

$$(e, \delta) \mapsto f_\lambda^{(M)}(e)\, f_{BB}(\delta),$$

where $f_\lambda^{(M)} \in \mathbb{R}^{n_\lambda}$ and $f_{BB} \in \mathcal{F}_{BB}$. The scalar product of $\mathcal{H}_\lambda$ is defined on the functions $f_\lambda^{(M)} \otimes f_{BB}$ as:

$$\langle f_\lambda^{(M)} \otimes f_{BB}, g_\lambda^{(M)} \otimes g_{BB} \rangle_\lambda := \langle f_\lambda^{(M)}, g_\lambda^{(M)} \rangle_\lambda^{(M)} \cdot \langle f_{BB}, g_{BB} \rangle_{BB}, \tag{3.17}$$

and then extended by linearity. For the sake of completeness, we report an explicit way to build $\mathcal{H}_\lambda$. Assume that $\{\eta_i\}_{i=1}^{n_\lambda}$ and $\{\phi_j\}_{j=1}^{\infty}$ denote two orthonormal bases of $\mathcal{H}_\lambda^{(M)}$ and $\mathcal{H}_{BB}$ respectively, then $\mathcal{H}_\lambda$ is the space generated by the (orthonormal) basis $\{\eta_i \otimes \phi_j\}_{i,j}$. Therefore, given two generic elements of $\mathcal{H}_\lambda$

$$f = \sum_{i,j} b_{ij} \eta_i \otimes \phi_j \qquad\qquad g = \sum_{i,j} c_{ij} \eta_i \otimes \phi_j,$$

their scalar product is defined as:

$$\begin{aligned} \langle f, g \rangle_\lambda &= \langle \sum_{i,j} b_{ij} \eta_i \otimes \phi_j, \sum_{i',j'} c_{i'j'} \eta_{i'} \otimes \phi_{j'} \rangle_\lambda \\ &= \sum_{i,j} \sum_{i',j'} b_{i,j} c_{i',j'} \langle \eta_i \otimes \phi_j, \eta_{i'} \otimes \phi_{j'} \rangle_\lambda \\ &= \sum_{i,j} \sum_{i',j'} b_{i,j} c_{i',j'} \langle \eta_i, \eta_{i'} \rangle_\lambda^{(M)} \cdot \langle \phi_j, \phi_{j'} \rangle_{BB} = \sum_{i,j} b_{i,j} c_{i,j}, \end{aligned} \tag{3.18}$$

where the last step relies on the orthonormality of $\{\eta_i\}_{i=1}^{n_\lambda}$ and $\{\phi_j\}_{j=1}^{\infty}$.

This concludes the construction of the RKHS of $k_E\big|_\lambda$.

Step 2: RKHS of k_E on $\widetilde{\boldsymbol{G}}$

Once the RKHSs of k_E have been defined on each life $\lambda \in \Lambda$, it is possible to characterise the RKHS of k_E on the whole set of edges (i.e. on $\widetilde{\boldsymbol{G}}$). Indeed, as k_E is null when computed among points on edges that do not share lives, the RKHS of k_E on $\widetilde{\boldsymbol{G}}$ will be the sheer direct sum of them:

$$\mathcal{H}_\Lambda := \bigoplus_{\lambda \in \Lambda} \mathcal{H}_\lambda.$$

Clearly, the elements of direct sum are the tuples of size $|\Lambda|$ of functions, each belonging to the respective $\mathcal{H}_\lambda$. However, there is clearly a bijection between such tuples and the set of functions defined on the union of the lives $\lambda's$. Therefore, $\mathcal{H}_\lambda$ can be interpreted as the set of functions $f : \widetilde{\boldsymbol{G}} \to \mathbb{R}$ such that their restrictions on each life λ belong to the respective $\mathcal{H}_\lambda$. In formulae: $\mathcal{H}_\Lambda = (\mathcal{F}_\Lambda, \langle \cdot, \cdot \rangle_\Lambda)$, with:

$$\mathcal{F}_\Lambda := \left\{ f : \widetilde{\boldsymbol{G}} \to \mathbb{R} \,:\, f \text{ is A.C., } f' \in L^2(\widetilde{\boldsymbol{G}}),\, f(V) = \mathbf{0} \right\}$$

$$\langle f, g\rangle_\Lambda := \sum_{\lambda\in\Lambda} \langle f\big|_\lambda, g\big|_\lambda\rangle_\lambda.$$

By construction (recall, by Berlinet and Thomas-Agnan 2004, Theorem 5, that the direct sum of RKHSs is the RKHS of the sum of the kernels), $\mathcal{H}_\Lambda$ is the RKHS of the kernel k_E on the whole graph $\widetilde{\boldsymbol{G}}$.

Step 3: RKHS of k_V on $\widetilde{\boldsymbol{G}}$
Since the kernel k_V is exactly the same defined in Anderes et al. (2020), this step is an adaptation. Consider $\mathcal{H}_V := (\mathcal{F}_V, \langle\cdot,\cdot\rangle_V)$, where:

$$\begin{aligned}
\mathcal{F}_V &:= \left\{ f : \widetilde{\boldsymbol{G}} \to \mathbb{R} \,:\, \forall e \in \widetilde{E},\, f\big|_e(u) = (1-\delta) f(\underline{u}) + \delta f(\overline{u}) \right\}, \\
\langle f, g\rangle_V &:= f(V)^\top \boldsymbol{x}\boldsymbol{x}^\top g(V) + \sum_{e\in\widetilde{E}} \frac{1}{\ell(e)} \int_0^1 f\big|_e'(\delta) g\big|_e'(\delta)\,\mathrm{d}\delta \qquad (3.19)\\
&= f(V)^\top \boldsymbol{x}\boldsymbol{x}^\top g(V) + \sum_{e\in\widetilde{E}} \frac{\left(f\big|_e(1) - f\big|_e(0)\right)\left(g\big|_e(1) - g\big|_e(0)\right)}{\ell(e)}.
\end{aligned}$$

Notice that the sum in the second line is the same quantity indicated by $\int_{\underline{e}}^{\overline{e}} f_e'(t) g_e'(t)\,\mathrm{d}t$ in Anderes et al. (2020). It is straightforward to show that $\langle\cdot,\cdot\rangle_V$ is a scalar product on $\mathcal{F}_V$: clearly it is symmetric and bilinear, moreover, for an $f \in \mathcal{F}_V$, we have:

$$\langle f, f\rangle_V = \left(\boldsymbol{x}^\top f(V)\right)^2 + \sum_{e\in\widetilde{E}} \frac{\left(f\big|_e(1) - f\big|_e(0)\right)^2}{\ell(e)} \geq 0,$$

where $\langle f, f\rangle_V = 0$ if and only if $\boldsymbol{x}^\top f(V) = 0$ and $\forall e \in \widetilde{E}$, $f\big|_e(1) = f\big|_e(0)$. From the latter we get that f is constant at all vertices, that is $f(V) = c\mathbf{1}_n$ for some $c \in \mathbb{R}$, and we conclude thanks to the former:

$$0 = \boldsymbol{x}^\top f(V) = c\,\boldsymbol{x}^\top \mathbf{1}_n,$$

that implies $c = 0$, since $\boldsymbol{x}^\top \mathbf{1}_n \neq 0$.

Step 4: RKHS of k on $\widetilde{\boldsymbol{G}}$ in the Non Periodic Case
Also this step is quite straightforward, since, being k on $\widetilde{\boldsymbol{G}}$ the sheer sum of k_V and k_E (Equation (3.7)), the RKHS of k_Z will be the direct sum of the RKHSs of k_V and k_E:

$$\mathcal{H} = \mathcal{H}_V \oplus \mathcal{H}_E.$$

Now, analogously to what noticed in Step 2, it is possible to observe that the set of couples $(f_V, f_E) \in \mathcal{H}_V \times \mathcal{H}_E$ is isomorphic to the set $\mathcal{F}$ of functions $f : \widetilde{\boldsymbol{G}} \to \mathbb{R}$ that are absolutely continuous on every edge and $f\big|_e'$ belongs to L^2. Summarising, if we define the two operators $\mathcal{P}_V, \mathcal{P}_\lambda : \mathcal{F} \to \mathcal{F}$ as follows:

$$\mathcal{P}_V(f)(u) := (1-\delta) f(\underline{u}) + \delta f(\overline{u}) \qquad (3.20)$$

$$\mathcal{P}_\lambda(f)(u) := \begin{cases} f(u) - \mathcal{P}_V(f)(u) & \text{if } u \in \lambda \\ 0 & \text{otherwise,} \end{cases} \tag{3.21}$$

then it is possible to express the RKHS of k_Z as follows: $\mathcal{H} = (\mathcal{F}, \langle\cdot,\cdot\rangle_\mathcal{H})$, where

$$\langle f, g\rangle_\mathcal{H} = \langle \mathcal{P}_V f, \mathcal{P}_V g\rangle_V + \sum_{\lambda\in\Lambda} \langle \mathcal{P}_\lambda f, \mathcal{P}_\lambda g\rangle_\lambda.$$

Notice that the operators $\mathcal{P}_V$ and $\mathcal{P}_\lambda$ are the same operators defined in Anderes et al. (2020), where our $\mathcal{P}_\lambda$ is simply the union of their $\mathcal{P}_e$, indexed by $e \in \lambda$. Therefore, we obtain that $\mathcal{P}_V$ and $\mathcal{P}_E$ are orthogonal projectors and self-adjoint.

Step 5: RKHS of k on $\widetilde{G}$ in the Periodic Case
The expression for the kernel k in case of a periodic time-evolving graph, given in Equation (3.14), has an additional term to ensure that the process Z varies at different times, namely $\beta^2 \min(t_1, t_2)$. Being this the kernel of a standard Wiener process multiplied by a constant $\beta > 0$, its RKHS is given by $\mathcal{H}_W := (\mathcal{F}_W, \langle\cdot,\cdot\rangle_W)$, where

$$\mathcal{F}_W := \left\{ f : \mathbb{R}_0^+ \to \mathbb{R} \,:\, f \text{ is A.C., } f(0) = 0,\, f' \in L^2(\mathbb{R}_0^+) \right\}$$
$$\langle f, g\rangle := \frac{1}{\beta^2} \int_0^{+\infty} f'(x) g'(x)\, \mathrm{d}x.$$

As a consequence, to obtain the RKHS of k in the case of a periodic time-evolving graph, it is sufficient to do the direct sum of $\mathcal{H}$ and $\mathcal{H}_W$. This concludes the construction of the RKHS for both cases.

We summarise the construction in the following Proposition.

Proposition 3.7. *Let $\widetilde{G}$ be an equivalent simple graph of either a time evolving graph or a periodic time evolving graph. Denote $\mathcal{F}$ the set of functions $f : \widetilde{G} \to \mathbb{R}$ that are continuous w.r.t. the shortest-path metric and that, for each edge $e \in \widetilde{E}$ the restriction f_e of f on e is absolutely continuous and f_e' belongs to $L^2(e)$. Further, define the quadratic form $\langle\cdot,\cdot\rangle_\mathcal{H} : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ as follows:*

$$\langle f, g\rangle_\mathcal{H} = \langle \mathcal{P}_V f, \mathcal{P}_V g\rangle_V + \sum_{\lambda\in\Lambda} \langle \mathcal{P}_\lambda f, \mathcal{P}_\lambda g\rangle_\lambda, \tag{3.22}$$

where:

- *$\langle\cdot,\cdot\rangle_V$ is defined in Equation (3.19);*
- *$\langle\cdot,\cdot\rangle_\lambda$ is defined in Equations (3.17) and (3.18);*
- *$\mathcal{P}_V$ and $\mathcal{P}_\lambda$ are defined in Equations (3.20) and (3.21) respectively.*

Then, the space $(\mathcal{F}, \langle\cdot,\cdot\rangle_\mathcal{H})$ is an infinite dimensional Hilbert space with reproducing kernel given by (3.7). The construction of the RKHS of the kernel (3.14) is obtained by the direct sum of $\mathcal{H}$ with the RKHS of a (scaled) Wiener process, as explained in Step 5 above.

3.7 Kernels and Covariance Functions

Here, we explore the possibility of defining a large class of covariance functions on the equivalent simple graph. Indeed, variograms can be composed with certain classes of functions to create valid covariance functions associated with semi-distance spaces. A function $\psi : [0, +\infty) \to \mathbb{R}$ is called *completely monotone* if it is continuous on $[0, +\infty)$, infinitely differentiable on $(0, +\infty)$ and for each $i \in \mathbb{N}$ it holds $(-1)^i \, \psi^{(i)}(x) \geq 0$, where $\psi^{(i)}$ denotes the i^{th} derivative of ψ and $\psi^{(0)} := \psi$. By the Bernstein's theorem (Bernstein, 1929), completely monotone functions are the Laplace transforms of positive and bounded measures. Some example of parametric families of completely monotone functions are presented in Table 3.1.

Type	$\psi(x)$	Parameter range
Power exponential	$e^{-\beta x^{\alpha}}$	$0 < \alpha \leq 1$, $\beta > 0$
Matérn	$\frac{2^{1-\alpha}}{\Gamma(\alpha)}(\beta x)^{\alpha} K_r(\beta x)$	$0 < \alpha \leq \frac{1}{2}$, $\beta > 0$
Generalised Cauchy	$(\beta x^{\alpha} + 1)^{-\xi/\alpha}$	$0 < \alpha \leq 1$, $\beta > 0$, $\xi > 0$
Dagum	$1 - \left(\frac{\beta x^{\alpha}}{1 + \beta x^{\alpha}}\right)^{\xi/\alpha}$	$0 < \alpha, \xi \leq 1$, $\beta > 0$
	$\frac{\log(1 + \beta x)}{\beta x}$	$\beta > 0$
	$(1 + \alpha \log(1 + \beta x))^{-\xi}$	$\alpha, \beta, \xi > 0$
	$e^{-\alpha}(1 + \beta x)^{\frac{\alpha}{\beta x}}$	$\alpha, \beta > 0$
	$\frac{1}{(\alpha + \xi)^{\mu}}\left(\alpha + \frac{\xi}{1 + \beta x}\right)^{\mu}$	$\alpha, \beta, \xi, \mu > 0$

Table 3.1: Examples of completely monotone functions $0 < x \mapsto \psi(x)$ such that $\psi(0^+) = 1$. Here, K_r denotes the modified Bessel function of the second kind. The first four rows are from Anderes et al. (2020, Table 1), whilst the others have been taken from Miller and Samko (2001).

By Theorem 6 and Corollary 1 in Anderes et al. (2020) it is immediate to show the next result, which allows to define covariance functions on arbitrary domains. As a sheer consequence, we obtain in Proposition 3.9 a positive definite kernel on the time-evolving graph.

Theorem 3.1. *Let Z be a stochastic process defined on a set X such that $\mathbb{E}\left(Z^2(x)\right) < +\infty$ for all $x \in X$. Define $d : X \times X \to \mathbb{R}_0^+$ as*

$$d(x_1, x_2) := \gamma_Z(x_1, x_2) = \mathbb{V}\mathrm{ar}\left(Z(x_1) - Z(x_2)\right). \tag{3.23}$$

In addition, let ψ be a non-constant completely monotone function on $[0, +\infty)$. Then the following holds:

1. *$(X, d) \overset{\sqrt{\cdot}}{\hookrightarrow} \mathcal{H}$ for some Hilbert space $\mathcal{H}$, that is: there exist a Hilbert space $\mathcal{H}$ and a function $\xi : X \to \mathcal{H}$ such that, given $x_1, x_2 \in X$: $\sqrt{d(x_1, x_2)} = \|\xi(x_1) - \xi(x_2)\|_{\mathcal{H}}$;*

2. *the function* $(x_1, x_2) \mapsto \psi(d(x_1, x_2))$ *is positive semi-definite;*

3. *if, in addition,* d *is a semi-distance on* X *(see Proposition 3.8 below for a useful characterisation), then* $(x_1, x_2) \mapsto \psi(d(x_1, x_2))$ *is strictly positive definite.*

Proof. **of Theorem 3.1**

1. Define $\tilde{d} := \sqrt{d}$. Since $d = \tilde{d}^2$ is a variogram, it is conditionally negative semidefinite: as a consequence, by Anderes et al. (2020, Theorem 6), $(X, \tilde{d}) \overset{id}{\hookrightarrow} \mathcal{H}$ for some Hilbert space $\mathcal{H}$. Therefore $(X, d) \overset{\sqrt{\cdot}}{\hookrightarrow} \mathcal{H}$.

2. Follows immediately from the previous point and Anderes et al. (2020, Corollary 1).

3. If $Z(x_1) = Z(x_2)$ almost surely implies $x_1 = x_2$, then (X, d) is a semi-distance space by Proposition 3.8. Thus, by Anderes et al. (2020, Corollary 1), $(x_1, x_2) \mapsto C(d(x_1, x_2))$ is strictly positive definite.

□

Proposition 3.8. *Let* Z*,* X *and* d *as in Theorem 3.1. Then:*

1. (X, d) *is a semi-distance space if and only if, for all* $x_1, x_2 \in X$*,* $Z(x_1) = Z(x_2)$ *almost surely implies* $x_1 = x_2$*;*

2. d *satisfies the triangular inequality if and only if, for all* $x_1, x_2, x_3 \in X$*, it holds:*
$$\mathbb{C}ov(Z(x_1) - Z(x_2), Z(x_3) - Z(x_2)) \geq 0.$$

Proof. **of Proposition 3.8**

1. (X, d) is a semi-distance space iff $d(x_1, x_2) = 0$ implies $x_1 = x_2$. But $d(x_1, x_2) = 0$ is equivalent to $Z(x_1) = Z(x_2)$ almost surely.

2. Without loss of generality, we can restrict the proof to a zero-mean process X. Indeed, both the variance and the covariance do not change if we change the mean of their arguments. The proof consists in the following chain of equivalences. Let $x_1, x_2, x_3 \in X$.

$$\begin{aligned}
& d(x_1, x_2) + d(x_2, x_3) \geq d(x_1, x_3) \\
\iff & \mathbb{V}\mathrm{ar}\left(Z(x_1) - Z(x_2)\right) + \mathbb{V}\mathrm{ar}\left(Z(x_2) - Z(x_3)\right) \geq \mathbb{V}\mathrm{ar}\left(Z(x_1) - Z(x_3)\right) \\
\iff & \mathbb{V}\mathrm{ar}\, Z(x_1) + \mathbb{V}\mathrm{ar}\, Z(x_2) - 2\,\mathbb{C}\mathrm{ov}\left(Z(x_1), Z(x_2)\right) \\
& \quad + \mathbb{V}\mathrm{ar}\, Z(x_2) + \mathbb{V}\mathrm{ar}\, Z(x_3) - 2\,\mathbb{C}\mathrm{ov}\left(Z(x_2), Z(x_3)\right) \geq \\
& \quad \mathbb{V}\mathrm{ar}\, Z(x_1) + \mathbb{V}\mathrm{ar}\, Z(x_3) - 2\,\mathbb{C}\mathrm{ov}\left(Z(x_1), Z(x_3)\right) \\
\iff & \mathbb{V}\mathrm{ar}\, Z(x_2) - \mathbb{C}\mathrm{ov}\left(Z(x_1), Z(x_2)\right) - \mathbb{C}\mathrm{ov}\left(Z(x_2), Z(x_3)\right) \geq \\
& \quad - \mathbb{C}\mathrm{ov}\left(Z(x_1), Z(x_3)\right)
\end{aligned}$$

$$\iff \mathbb{E}\left(Z^2(x_2) + Z(x_1)Z(x_3) - Z(x_1)Z(x_2) - Z(x_2)Z(x_3)\right) \geq 0$$
$$\iff \mathbb{E}\left((Z(x_2) - Z(x_1))(Z(x_2) - Z(x_3))\right) \geq 0$$
$$\iff \mathbb{C}\mathrm{ov}\left(Z(x_2) - Z(x_1), Z(x_2) - Z(x_3)\right) \geq 0$$

□

Proposition 3.9. *Let $\psi : [0, \infty) \to \mathbb{R}$ be a continuous, completely monotone function on the positive real line, with $\psi(0) < \infty$. Let $d : \widetilde{\boldsymbol{G}} \times \widetilde{\boldsymbol{G}} \to \mathbb{R}$ be the mapping defined at (3.11). Then, the function $k_Z(u_1, u_2) = \psi\left(d(u_1, u_2)\right)$, where $u_1, u_2 \in \widetilde{\boldsymbol{G}}$, is a strictly positive definite function.*

Proposition 3.9 provides a very easy recipe to build kernels over time evolving graphs, whatever the temporal structure (linear or periodic). Any element from the Table 3.1 is a good candidate for such a composition. We do not report the corresponding algebraic forms, instead we illustrate how these covariance works through two practical examples. We believe that the free parameters and the large number of analytically-tractable completely monotone functions provide a wide range of models that could fit several real-world frameworks.

3.7.1 Linear Time

We start by considering the graph in Figure 3.9. Here, we have $m = 3$ time instants. Further, $V = \{A_0, B_0, C_0, D_0, A_1, B_1, C_1, D_1, A_2, C_2, D_2\}$. We focus on the semi-distances as well as the covariances between the points $A_0 = (A_0, B_0, \delta(A_0) = 0)$, $P := (C_0, D_0, \delta(P) = 0.8)$ and $Q := (C_2, D_2, \delta(Q) = 0.5)$. All the spatial edges E_S have weight 1, whilst the temporal edges have weight $\alpha > 0$. Finally, we use $k_T(h) = \phi^{|h|}$ as defined in Subsection 3.4.2, with $\phi \in (-1, 1)$ free parameter. Figure 3.9 clearly shows the effect the temporal edge parameter α plays on the semi-distances: while it has a considerable impact on the semi-distances $d(A_0, Q)$ and $d(P, Q)$, it shows a negligible effect on $d(A_0, P)$. This is reasonable given the graph structure: A_0 and P belong to the same layer ($t = 0$) and they are connected from both the paths A_0, D_0, P and A_0, B_0, C_0, P, which completely lie on $t = 0$ (and therefore they do not change with α). On the other hand, Q can be connected to A_0 and P only via paths that include temporal edges. As a consequence, if $\alpha \to 0^+$, both $d(A_0, Q)$ and $d(P, Q)$ will go to infinity.

The plot on the right of Figure 3.9 shows the effect of the correlation parameter ϕ as well. Whilst it does not influence the semi-distances concerning A_0 (since it is a vertex), as it increase, it reduces the semi-distances between P and Q. Clearly, the effect is more significant for large values of the parameter α. Indeed, when α is small, the semi-distances between nodes at different time instants are large. As a consequence, the second line of equation (3.7) becomes negligible when compared to the first one. Figure 3.10 shows the resulting effect of the parameter α on the correlations between A_0, P and Q generated by the composition of two completely monotone functions taken from Table 3.1 and the semi-distances shown in Figure 3.9.

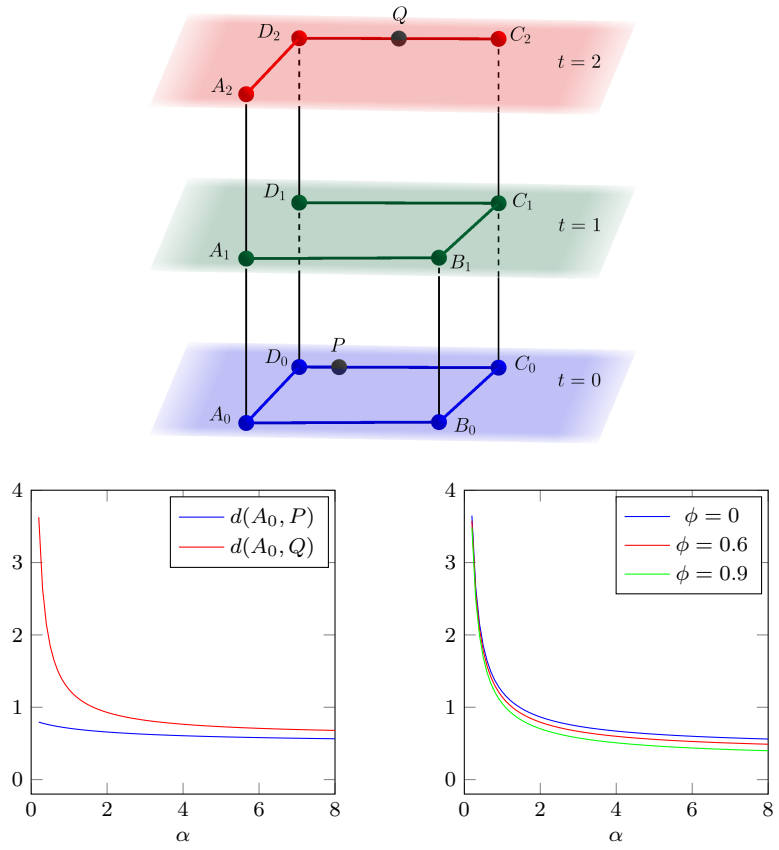


Figure 3.9: Top: equivalent simple graph taken as an example for the linear-time case. Bottom: semi-distances between the points A_0, P and Q. Notice that while $d(A_0, P)$ and $d(A_0, Q)$ (left) do not depend on ϕ, the semi-distance $d(P, Q)$ (right) decreases as ϕ increases.

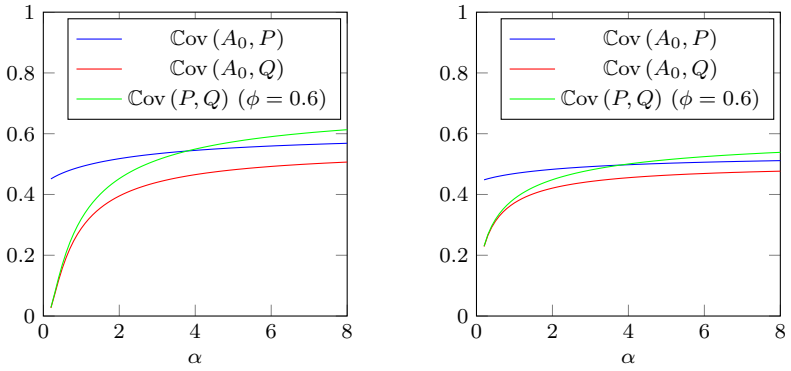


Figure 3.10: Covariances generated by the distances in Figure 3.9 between the points A_0, P and Q. Left: exponential kernel with parameters $(\alpha = 1, \beta = 1)$ (see Table 3.1). Right: generalised Cauchy kernel with parameters $(\alpha = 1, \beta = 5, \xi = 0.5)$ (see Table 3.1).

3.7.2 Circular Time

We are going to analyse the time-evolving periodic graph depicted in Figure 3.11. In this case, we have $m = 8$ and $V = \{A_0, B_0, A_1, B_1, \ldots, A_7, B_7\}$. We will compare the semi-distances and the covariances between the points $P_0 := (0, A_0, B_0, \delta(P_0) = 0.5)$ and $P_t := (t, A_\tau, B_\tau, \delta(P_t) = 0.5)$, where $t \in \mathbb{N}$ and $\tau \equiv t \pmod m$. Here, all the spatial edges have weight 1, whilst the temporal ones have weight $\alpha > 0$. We use the temporal kernel k_T as described in Subsection 3.5.1, with ρ and β free parameters.

The semi-distances and covariances in Figure 3.12 show the effect of the parameter β on our construction. It adds a linear component $\beta^2 t$, which allows to calibrate the semi-distance (and, as a result, the covariance) between the points at different time instants. In such a way, the effect of the periodicity is increased by setting a low β and becomes negligible when β grows. Furthermore, notice the spikes the covariance functions show: they perfectly embody the periodic setting of a process, as introduced in Section 3.5. We notice that, although isotropic covariances are decreasing functions of the spatial (semi-)distances, Figures 3.12 and 3.13 show valid covariance functions, as the semi-distance of our setting is completely different from the Euclidean distance on $\mathbb{R}^n$ as it takes into account the spatio-temporal structure of the time-evolving graph.

In Figure 3.13, it is possible to visualise the effect of the partial correlation parameter $\rho \in [0, \frac{1}{2})$. First, notice that its role is particularly significant when the weight α is high. Indeed, for low α's, the covariance structure of the vertices given by the inverse Laplacian matrix is dominant. Yet, when α is high, the nodes at different time instants are considered close to each other and the resulting (semi-)distances given by the sheer first line of (3.14) are low. Thus, the parameter ρ (which enters in the second line of (3.14)) has a greater influence. Clearly, the greater is ρ, the lower is the semi-distance, as it correlates different Brownian bridge realisations.

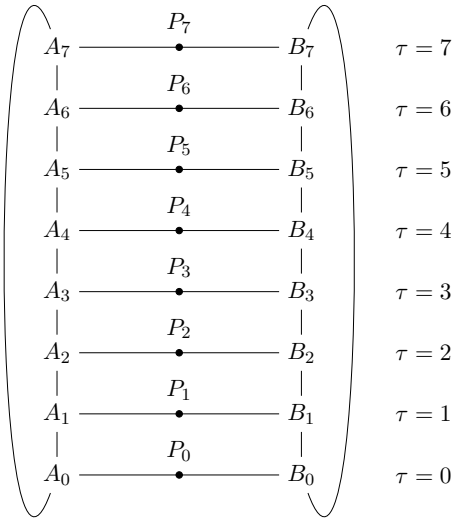


Figure 3.11: Equivalent simple graph taken as an example for the circular-time case.

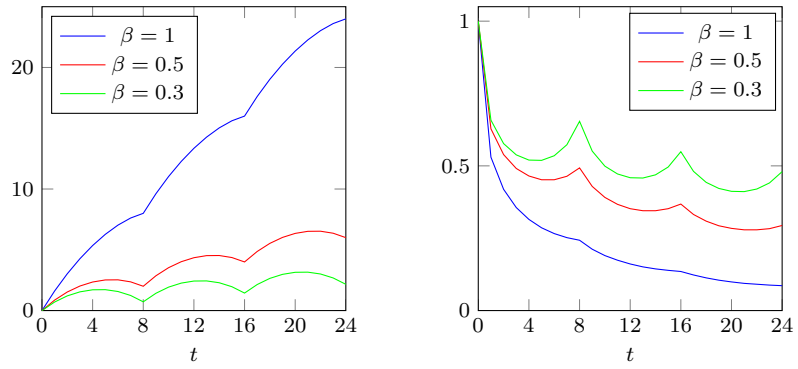


Figure 3.12: Semi-distances (left) and covariances (right) for the graph in 3.11 between the points P_0 and P_t for $\rho = 0.45$ and $\alpha = 1$. Covariances have been generated via the exponential kernel with parameters $\alpha = 0.5$ and $\beta = 0.5$ (see Table 3.1).

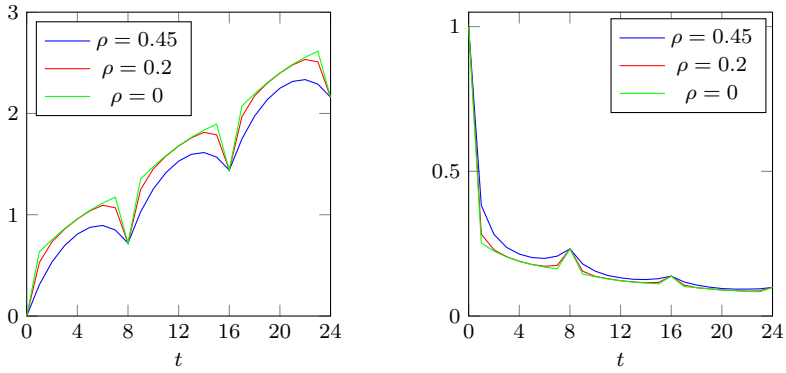


Figure 3.13: Semi-distances (left) and covariances (right) for the graph in 3.11 between the points P_0 and P_t for $\alpha = 10$ and $\beta = 0.3$. Covariances have been generated via the Dagum kernel with parameters $\alpha = 1$, $\beta = 2$ and $\xi = 0.5$ (see Table 3.1).

3.7.3 A Note on Continuous Time

There may be several approaches to extend this work to continuous-time settings. The first, and perhaps most intuitive, is to perform a limit operation, increasing the number of time instants for a fixed time interval (either linear or periodic). However, this would increase the number of vertices and, as a consequence, the dimension of the Laplacian matrix L to be inverted, leading to extremely high computational burdens. Another approach could be to use the values of the process Z on temporal edges E_T connecting time $t \in T$ to $t + 1 \in T$ to define the process and, thus, the (semi-)metric on the whole interval $[t, t + 1]$ for each couple of points belonging to both G_t and G_{t+1}.

3.8 Numerical Experiments

In order to illustrate our approach, we performed a numerical experiment to show how it can handle a time-evolving graph. More specifically, we considered the periodic time-evolving graph $\widetilde{\boldsymbol{G}}_1$ depicted in Figure 3.14: at any $\tau \in \{0, 1, 2\}$, it has two vertices, say A and B; yet, while at both $\tau = 0$ and $\tau = 2$ they are connected by an edge of weight 1, at $\tau = 1$ they are not connected. We then set parameters $\alpha := 1$, $\beta := 0.5$ and $\gamma := 0.6$. Notice that in this graph the parameter ρ has no effect, since there is no edge whose lifespan coincides with $\{0, 1, 2\}$. We compared $\widetilde{\boldsymbol{G}}_1$ with $\widetilde{\boldsymbol{G}}_2$, that is a "regularised" version of $\widetilde{\boldsymbol{G}}_1$: we assume that at time $\tau = 1$ the vertices A and B are indeed connected by the edge with weight 1. We considered a regular grid of points on $\widetilde{\boldsymbol{G}}_1$, including all the vertices and 3 equally-spaced points on each spatial edge and an evolution of 2 complete periods ($t \in \{0, \dots, 5\}$), resulting in 24 spatio-temporal points: 6 vertices and 6 points on the edges, times 2 periods. We considered the same grid on $\widetilde{\boldsymbol{G}}_2$ as well. Next, we computed their 24×24 semi-distance matrix $\mathbf{D}$

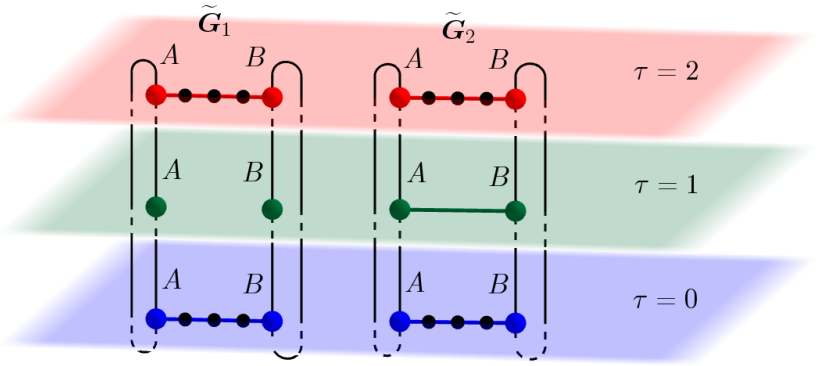


Figure 3.14: The two periodic time-evolving graphs used for the numerical experiment. Coloured points are vertices, while black ones are points on the spatial edges taken as regular grid.

and covariances matrix $\mathbf{\Sigma} = \psi(\mathbf{D})$, where we used the completely monotone function $\psi(x) := e^{-x}$. We simulated $N := 10^4$ iid samples from the multivariate normal distribution $\mathcal{N}_{24}(\mathbf{0}, \mathbf{\Sigma})$. Next, we performed a maximum likelihood estimation of the parameters $\alpha, \beta, \gamma, \rho$ under both $\widetilde{\boldsymbol{G}}_1$ and $\widetilde{\boldsymbol{G}}_2$ and compared the results (see Table 3.2). Clearly, the maximum likelihood estimates under the true underlying graph

	$\theta_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_2}$
α	1.0000	0.9980	0.7825
β	0.5000	0.5006	0.5076
γ	0.6000	0.6059	–
ρ	–	–	0.2882
$\log L$	-16934.6	-16934.2	-18991.3
$\mathbb{E}\lvert\Delta\mathbf{D}\rvert$	0.0000	0.0016	0.0857
$\max\lvert\Delta\mathbf{D}\rvert$	0.0000	0.0037	0.8600
$\mathbb{E}\lvert\Delta\mathbf{\Sigma}\rvert$	0.0000	0.0004	0.0232
$\max\lvert\Delta\mathbf{\Sigma}\rvert$	0.0000	0.0009	0.3042

Table 3.2: Comparison between the maximum likelihood estimates under $\widetilde{\boldsymbol{G}}_1$ and $\widetilde{\boldsymbol{G}}_2$. Blank spaces denote that the parameter is not influent in the model.

structure $\widetilde{\boldsymbol{G}}_1$ are extremely close to the true ones, whilst the "regularised" graph $\widetilde{\boldsymbol{G}}_2$ clearly is not able to capture the actual data structure. This behaviour can also be deduced from the log-likelihood and the summary statistics that report the mean and the maximum discrepancies of the matrices $\Delta\mathbf{D} := \hat{\mathbf{D}} - \mathbf{D}$ and $\Delta\mathbf{\Sigma} := \hat{\mathbf{\Sigma}} - \mathbf{\Sigma}$.

Such matrices are also represented in Figures 3.15 and 3.16. We point out that the effect of the introduction of edge connecting A and B at time $\tau = 1$ in the graph $\widetilde{\boldsymbol{G}}_2$ is particularly significant in the semi-distances and, thus, in the covariances between such points. However, it also influences the general spatio-temporal structure of the graph and, as a consequence, of the process.

Additional numerical examples can be found at the end of this Chapter (p. 108).

3.9 Conclusion: Statistical Implications and Broader Impact

We have contributed with a first attempt to define distances and kernels for a class of graphs whose topology can evolve over time, a setting of growing importance across a range of scientific and engineering domains. This work generalises Anderes et al. (2020), who contributed through static graphs and provides a new tool to space-time analysis over networks by introducing dynamics into metrics and consequently on the covariance functions, to provide a more faithful representation of certain types of networks.

The examples provided in the introduction explain the relevance of having adaptive topologies, and consequently adaptive covariance functions, in a wealth of practical real-life situations (among them, urban mobility systems, active river networks, or biological connectivity structures) and hence a rigorous and valid support for statistical modelling over temporally-adaptive networks.

The methodology we proposed might be outperformed in some real-world cases by *ad hoc* models, since these are specifically built to capture the intrinsic dynamics of the relative phenomena. Nevertheless, to our best knowledge, this is the first approach that provides a unified framework for time-evolving graphs. As such, it opens many aspects for future researches. The next step will be to inspect the finite sample properties of our covariance functions on real-life datasets, which is beyond our scope for this paper. A second step will be concerned about how to efficiently simulate Gaussian processes over dynamical networks, adapting, for instance, the methods in Alegría et al. (2025) to the time-evolving setting.

A fundamental aspect for future work will be to understand statistical inference aspects, maximum likelihood estimation, and kriging misspecification, under fixed domain asymptotics (Bevilacqua et al., 2019; Bolin and Wallin, 2024) as well as Bayesian inference and its computational aspects.

Through examples from neuroscience, hydrology, and transportation science, we have illustrated several fields that could benefit from this contribution. For instance, modelling dynamic functional connectivity in fMRI studies, evolving discharge patterns in stream networks, or real-time passenger flow in urban transit systems all benefit from our ability to define stochastic processes over changing topologies.

Finally, our approach might be useful to integrate climate networks - called Tsonis networks in Chapter 2 - typically constructed through measures of statistical

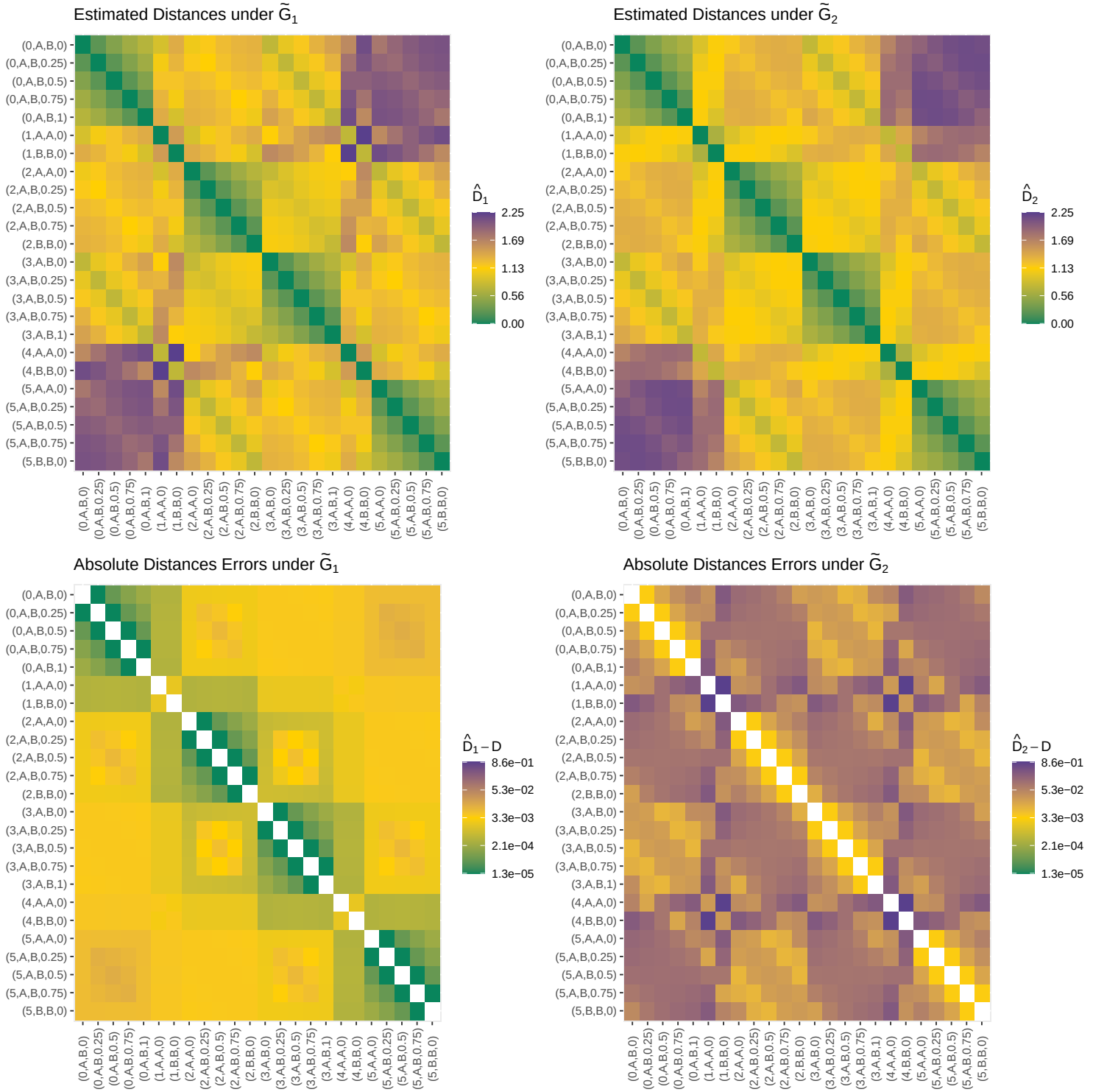


Figure 3.15: Estimated distance matrices $\hat{\mathbf{D}}$ (top) and distance errors $\hat{\mathbf{D}} - \mathbf{D}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

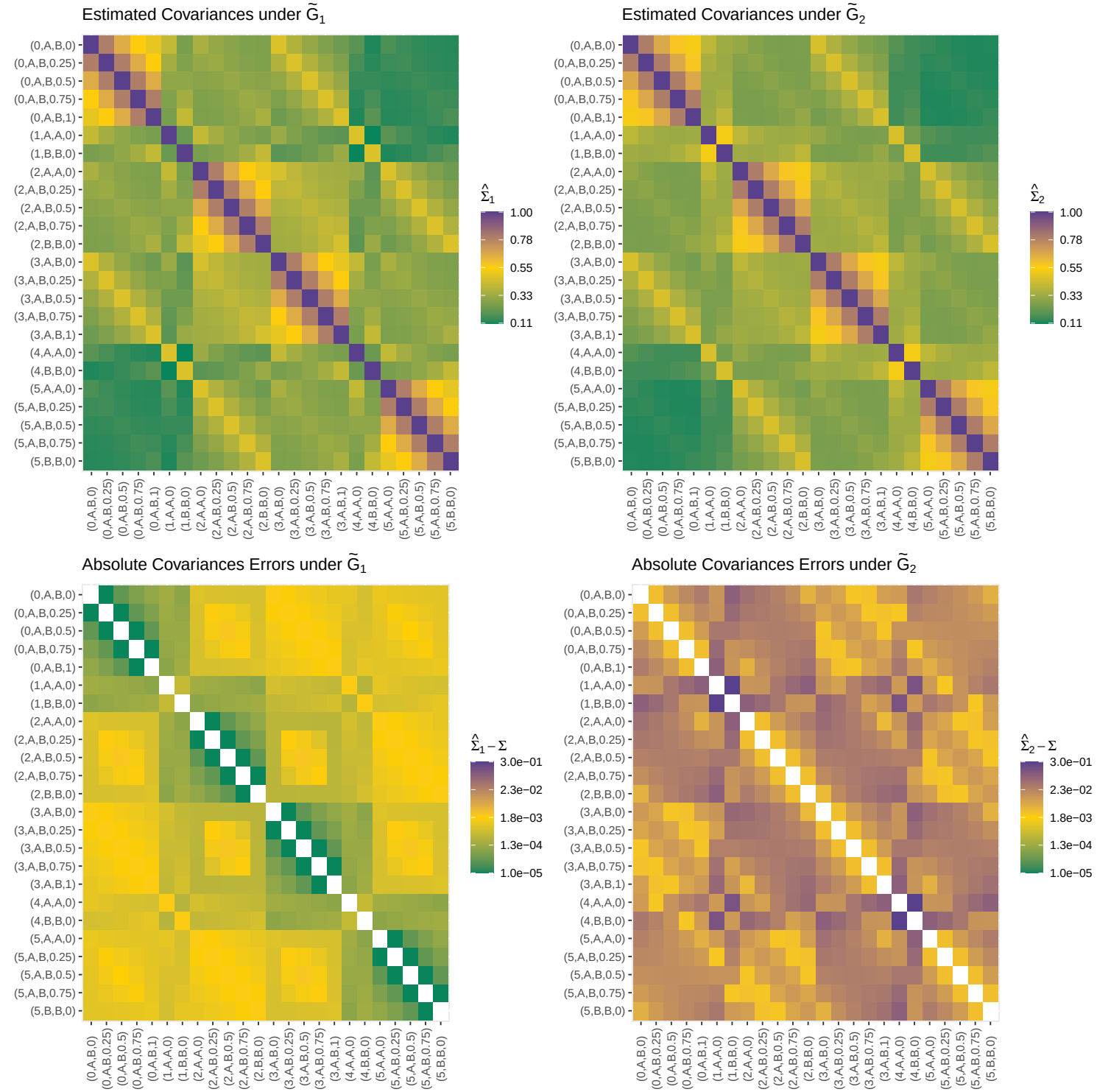


Figure 3.16: Estimated covariance matrices $\hat{\mathbf{\Sigma}}$ (top) and covariances errors $\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

similarity, with geophysical networks - such as river basins - to improve the statistical understanding of one or the other. Merging networks having different topologies and different temporal evolution is a key of success to understand certain climate phenomena (Tsonis and Roebber, 2004).

We believe the methodological advances proposed in this work will be useful in a growing number of real-world problems as it offers a new statistical lens on temporally-evolving systems and provides both theoretical rigour and practical tools to model, infer, and understand data in non-stationary and structurally dynamic environments.

Acknowledgements

The authors are grateful to Havard Rue for insightful discussions during the preparation of this manuscript. This project has also been sustained by the prompt help of Valeria Simoncini and Valter Moretti.

Additional Numerical Experiments

Here, we present the results of some numerical experiments similar to the one in Section 3.8. In total, we performed 6 different simulations and estimations. Each of them is a combination of a regular grid on a time-evolving graph (and relative parameters α, β, γ, ρ) and of a completely monotonic function ψ used to build the covariance matrix. Such functions are: $\psi_1(x) := e^{-x}$ and $\psi_2(x) := \frac{1}{1+x}$.

Regarding graphs, the first Example uses the one presented in Figure 3.14. The second Example uses a periodic graph with period $m = 3$ and 3 vertices (A, B and C). At time $\tau = 0$, the edge set is $\{(A, B), (B, C)\}$. At time $\tau = 1$, the only connection is (A, B) and there is no vertex C. Finally, at time $\tau = 2$, there are no connections and only vertices A and B. The third Example uses the graph presented in Figure 3.6. Table 3.3 summarises the parameters used for each graph. For computational reasons, having the third graph 16 nodes, we have reduced the number of considered periods to 1 and the number of sample points on the regular grid (breaks) to 1 per edge. The performed experiment is the same described in Section 3.8, always with 10^4

Parameter	1st Example	2nd Example	3rd Example
breaks	3	1	1
periods	2	2	1
α	1	0.8	1.5
β	0.5	0.2	0.7
γ	0.6	0.9	0.5
ρ	0.4	0.3	0.2

Table 3.3: The parameters used as true values (to be estimated) in the additional simulation studies (see Table 3.4 for the detailed estimates).

iid samples. Results (tables and plots) are presented next. As general remarks, the maximum likelihood estimation returns very good results if the underlying graph is indeed the one that generated data (denoted by $\widetilde{\boldsymbol{G}}_1$), whilst the "regularised" graphs ($\widetilde{\boldsymbol{G}}_2$) provide bad approximations of their counterparts. In addition, it is patent that the largest discrepancies in both distances and covariances arise in nodes that in the time-evolving layers show low connection with other nodes (*e.g.* node $(2, C, C, 0)$ in the 3rd Example). This is expected as these are the points at which the topology is most altered by the "regularisation".

	$\psi_1(x) = e^{-x}$			$\psi_2(x) = \frac{1}{1+x}$			
1st Example		$\theta_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_2}$		$\theta_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_2}$
	α	1.0000	0.9980	0.7825	α	1.0000	0.9842	0.7980
	β	0.5000	0.5006	0.5076	β	0.5000	0.4977	0.4977
	γ	0.6000	0.6059	–	γ	0.6000	0.5826	–
	ρ	–	–	0.2882	ρ	–	–	0.2691
	$\log L$	-16934.6	-16934.2	-18991.3	$\log L$	-5224.5	-5221.8	-6375.3
	$\mathbb{E}\lvert\boldsymbol{\Delta D}\rvert$	0.0000	0.0016	0.0857	$\mathbb{E}\boldsymbol{\Delta D}$	0.0000	0.0042	0.0818
	$\max\lvert\boldsymbol{\Delta D}\rvert$	0.0000	0.0037	0.8600	$\max\boldsymbol{\Delta D}$	0.0000	0.0160	0.8701
	$\mathbb{E}\lvert\boldsymbol{\Delta\Sigma}\rvert$	0.0000	0.0004	0.0232	$\mathbb{E}\boldsymbol{\Delta\Sigma}$	0.0000	0.0009	0.0158
	$\max\lvert\boldsymbol{\Delta\Sigma}\rvert$	0.0000	0.0009	0.3042	$\max\boldsymbol{\Delta\Sigma}$	0.0000	0.0037	0.2110
2nd Example		$\theta_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_2}$		$\theta_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_2}$
	α	0.8000	0.7979	0.5303	α	0.8000	0.7993	0.5534
	β	0.2000	0.1997	0.2072	β	0.2000	0.2007	0.2038
	γ	0.9000	0.9176	–	γ	0.9000	0.9033	–
	ρ	–	–	0.4900	ρ	–	–	0.4900
	$\log L$	18823.6	18826.3	15863.6	$\log L$	83408.7	83410.2	81474.3
	$\mathbb{E}\boldsymbol{\Delta D}$	0.0000	0.0011	0.2296	$\mathbb{E}\boldsymbol{\Delta D}$	0.0000	0.0008	0.1885
	$\max\boldsymbol{\Delta D}$	0.0000	0.0082	1.0702	$\max\boldsymbol{\Delta D}$	0.0000	0.0021	1.0768
	$\mathbb{E}\boldsymbol{\Delta\Sigma}$	0.0000	0.0004	0.0651	$\mathbb{E}\boldsymbol{\Delta\Sigma}$	0.0000	0.0002	0.0391
	$\max\boldsymbol{\Delta\Sigma}$	0.0000	0.0047	0.3329	$\max\boldsymbol{\Delta\Sigma}$	0.0000	0.0007	0.2340
3rd Example		$\theta_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_2}$		$\theta_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_1}$	$\hat{\theta}_{\widetilde{\boldsymbol{G}}_2}$
	α	1.5000	1.4938	0.6817	α	1.5000	1.4954	0.6824
	β	0.7000	0.7106	0.5764	β	0.7000	0.6874	0.5471
	γ	0.5000	0.4907	–	γ	0.5000	0.4955	–
	ρ	0.2000	0.1999	0.2797	ρ	0.2000	0.1985	0.2775
	$\log L$	-44309.0	-44304.7	-46854.9	$\log L$	-22261.4	-22258.5	-23969.1
	$\mathbb{E}\boldsymbol{\Delta D}$	0.0000	0.0197	0.1140	$\mathbb{E}\boldsymbol{\Delta D}$	0.0000	0.0213	0.1515
	$\max\boldsymbol{\Delta D}$	0.0000	0.0493	0.5021	$\max\boldsymbol{\Delta D}$	0.0000	0.0521	0.6011
	$\mathbb{E}\boldsymbol{\Delta\Sigma}$	0.0000	0.0042	0.0270	$\mathbb{E}\boldsymbol{\Delta\Sigma}$	0.0000	0.0034	0.0262
	$\max\boldsymbol{\Delta\Sigma}$	0.0000	0.0079	0.1786	$\max\boldsymbol{\Delta\Sigma}$	0.0000	0.0066	0.1274

Table 3.4: Summary of the MLE results in the three examples (rows) times the two completely monotonic functions used for the evaluation of the covariance (columns).

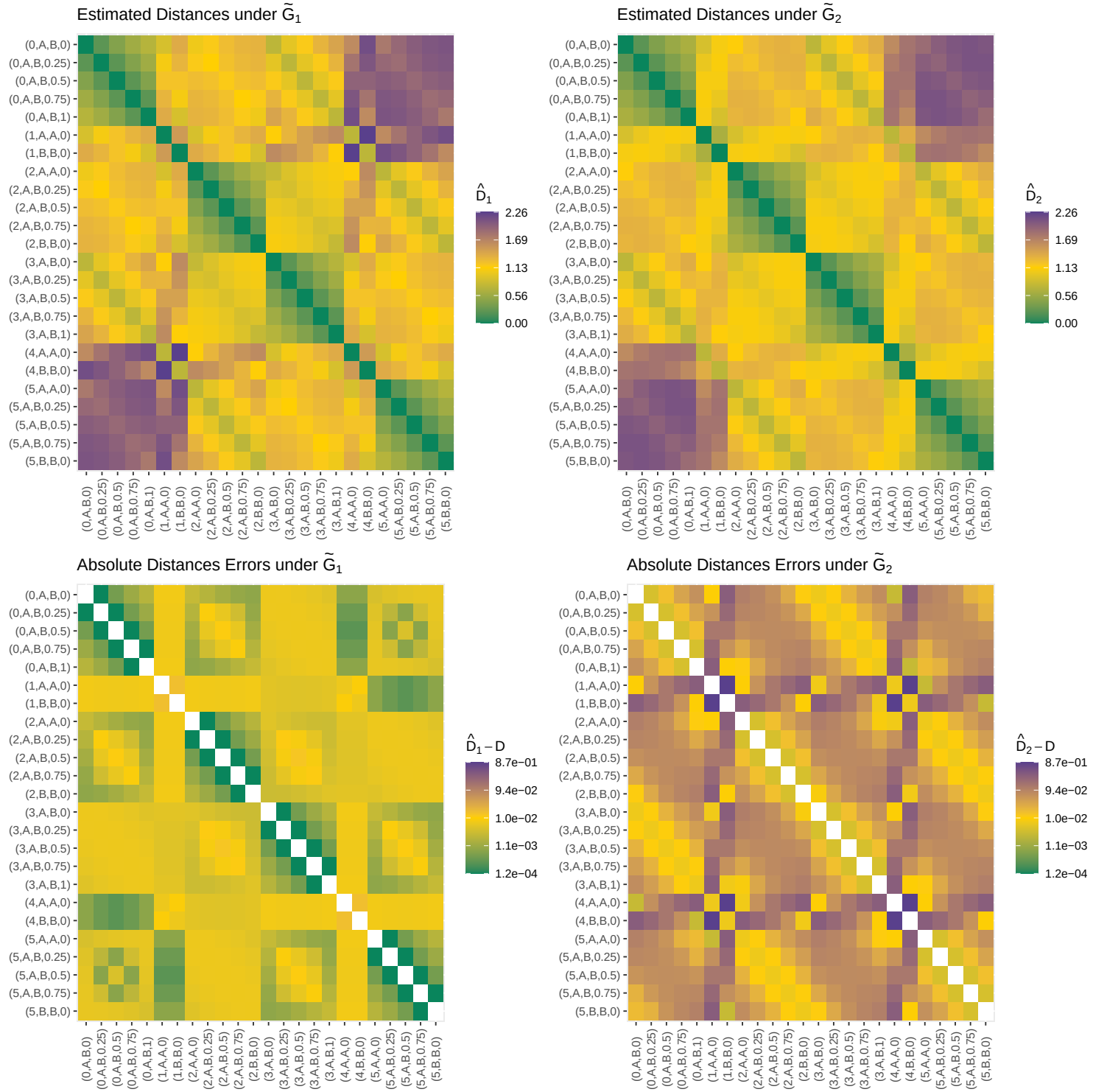


Figure 3.17: **1st Example with function** $\psi_2(x) = \frac{1}{1+x}$**.** Estimated distance matrices $\hat{\mathbf{D}}$ (top) and distance errors $\hat{\mathbf{D}} - \mathbf{D}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

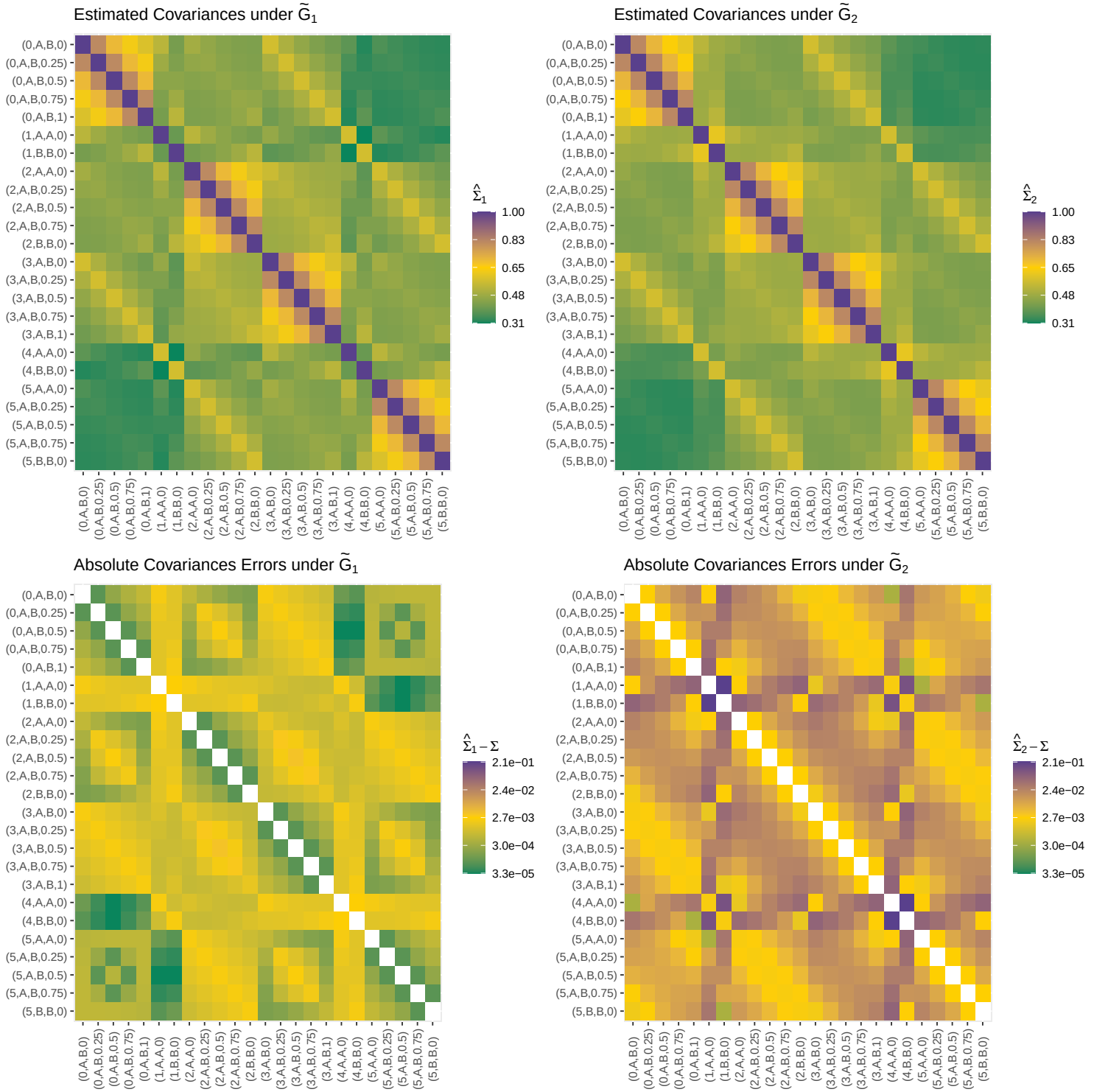


Figure 3.18: **1st Example with function** $\psi_2(x) = \frac{1}{1+x}$**.** Estimated covariance matrices $\hat{\boldsymbol{\Sigma}}$ (top) and covariances errors $\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

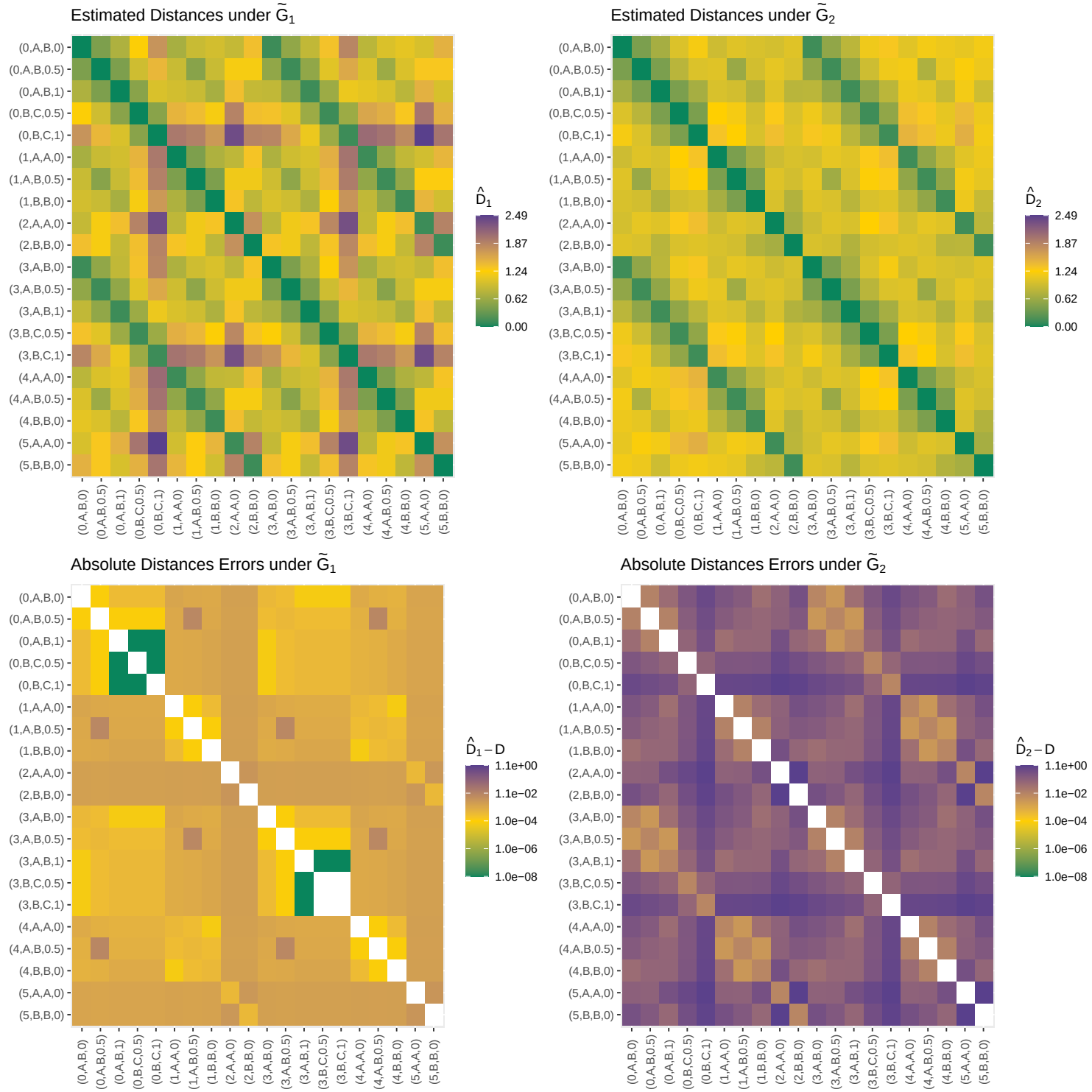


Figure 3.19: **2nd Example with function** $\psi_1(x) = e^{-x}$**.** Estimated distance matrices $\hat{\mathbf{D}}$ (top) and distance errors $\hat{\mathbf{D}} - \mathbf{D}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

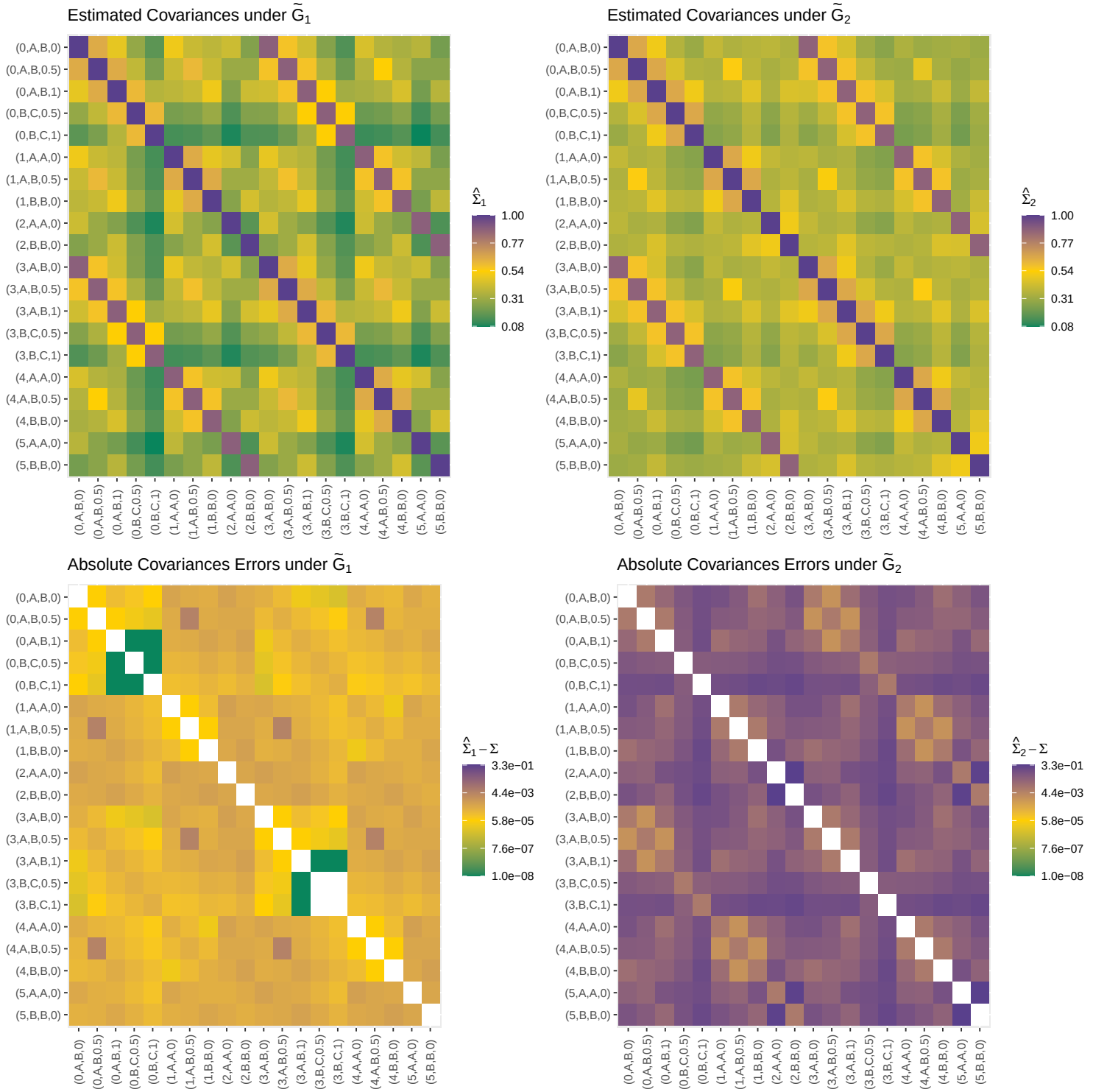


Figure 3.20: **2nd Example with function** $\psi_1(x) = e^{-x}$**.** Estimated covariance matrices $\hat{\boldsymbol{\Sigma}}$ (top) and covariances errors $\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

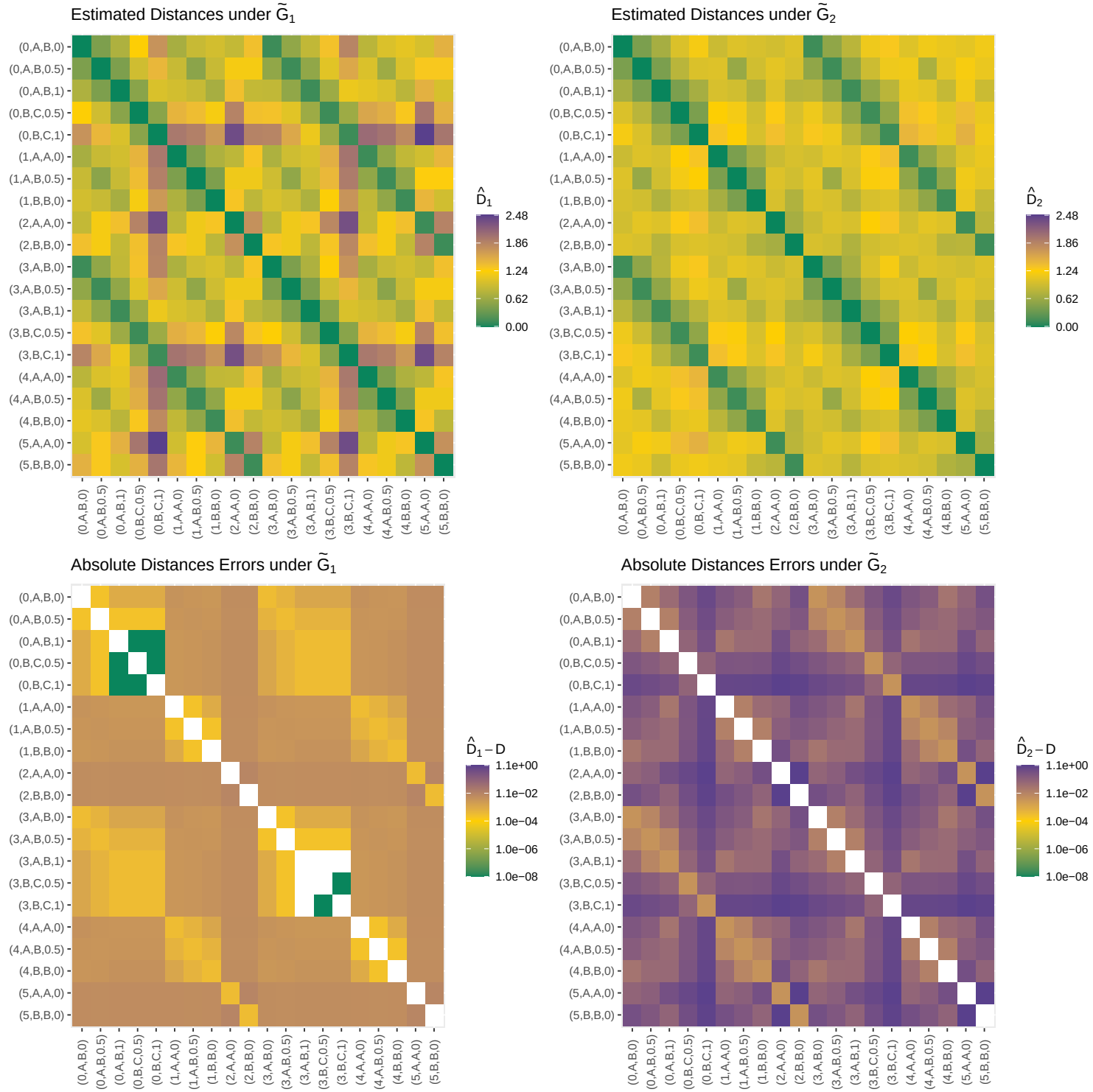


Figure 3.21: **2nd Example with function** $\psi_2(x) = \frac{1}{1+x}$**.** Estimated distance matrices $\hat{\mathbf{D}}$ (top) and distance errors $\hat{\mathbf{D}} - \mathbf{D}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

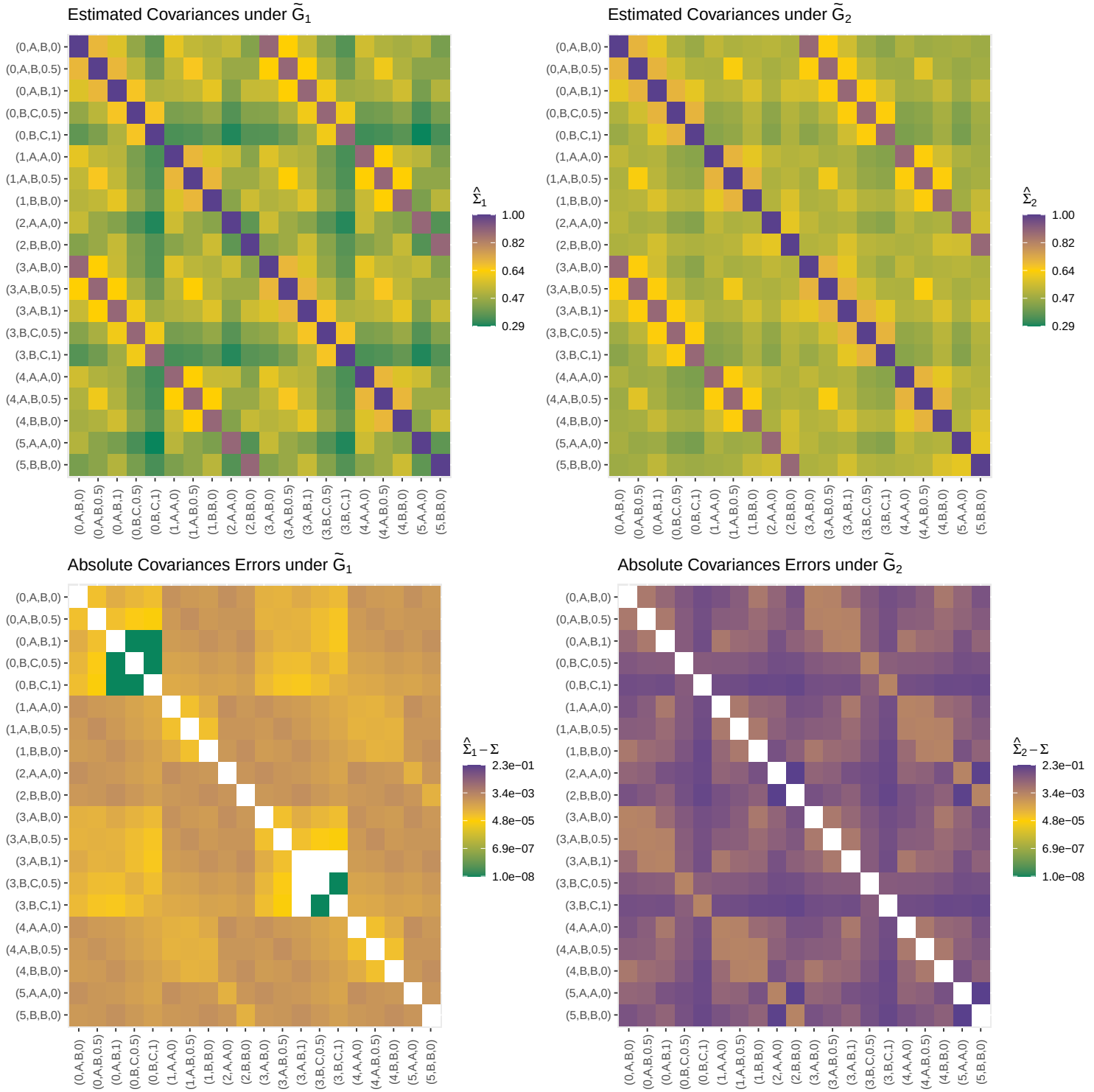


Figure 3.22: **2nd Example with function** $\psi_2(x) = \frac{1}{1+x}$**.** Estimated covariance matrices $\hat{\boldsymbol{\Sigma}}$ (top) and covariances errors $\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

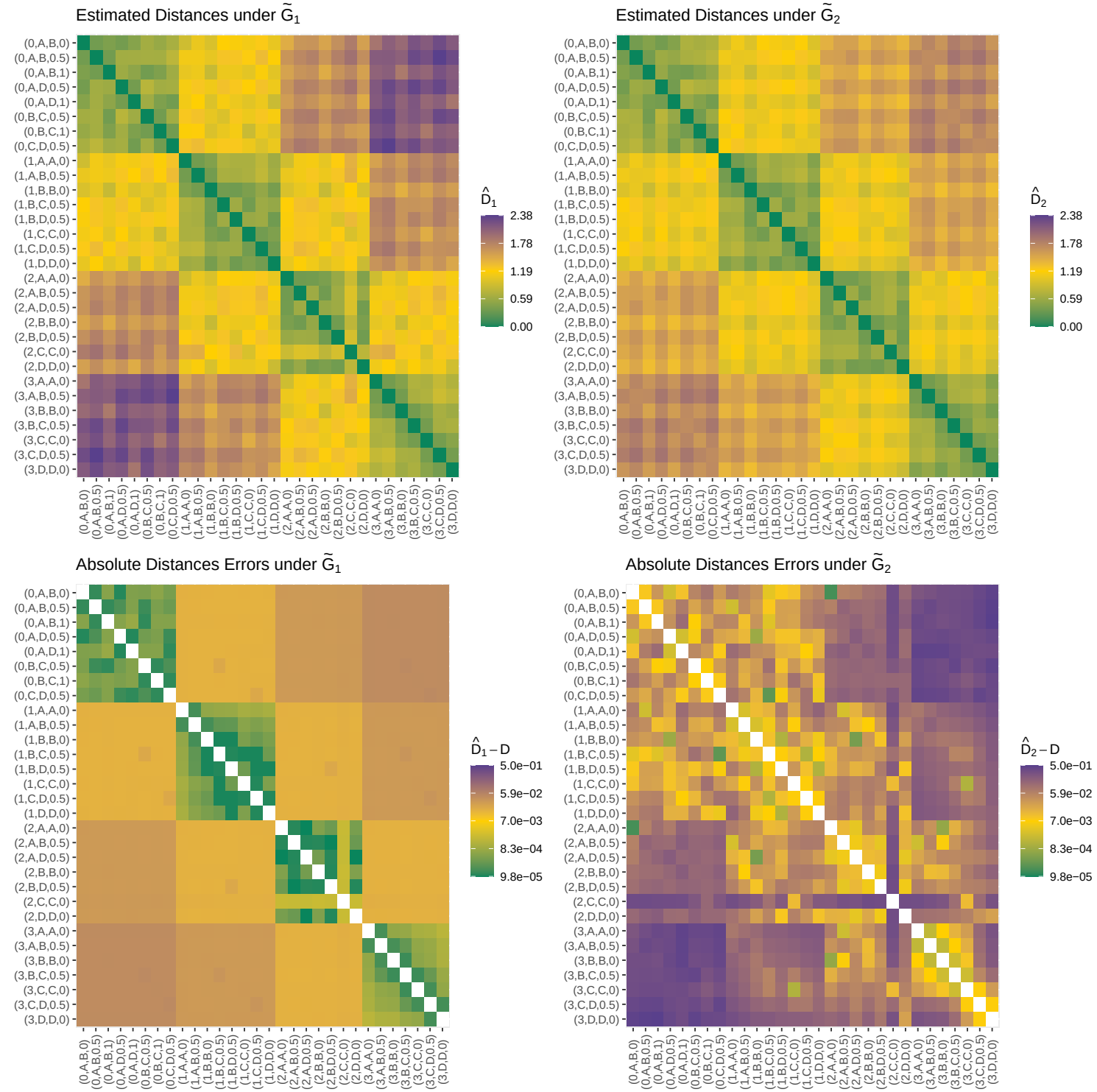


Figure 3.23: **3rd Example with function** $\psi_1(x) = e^{-x}$**.** Estimated distance matrices $\hat{\mathbf{D}}$ (top) and distance errors $\hat{\mathbf{D}} - \mathbf{D}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

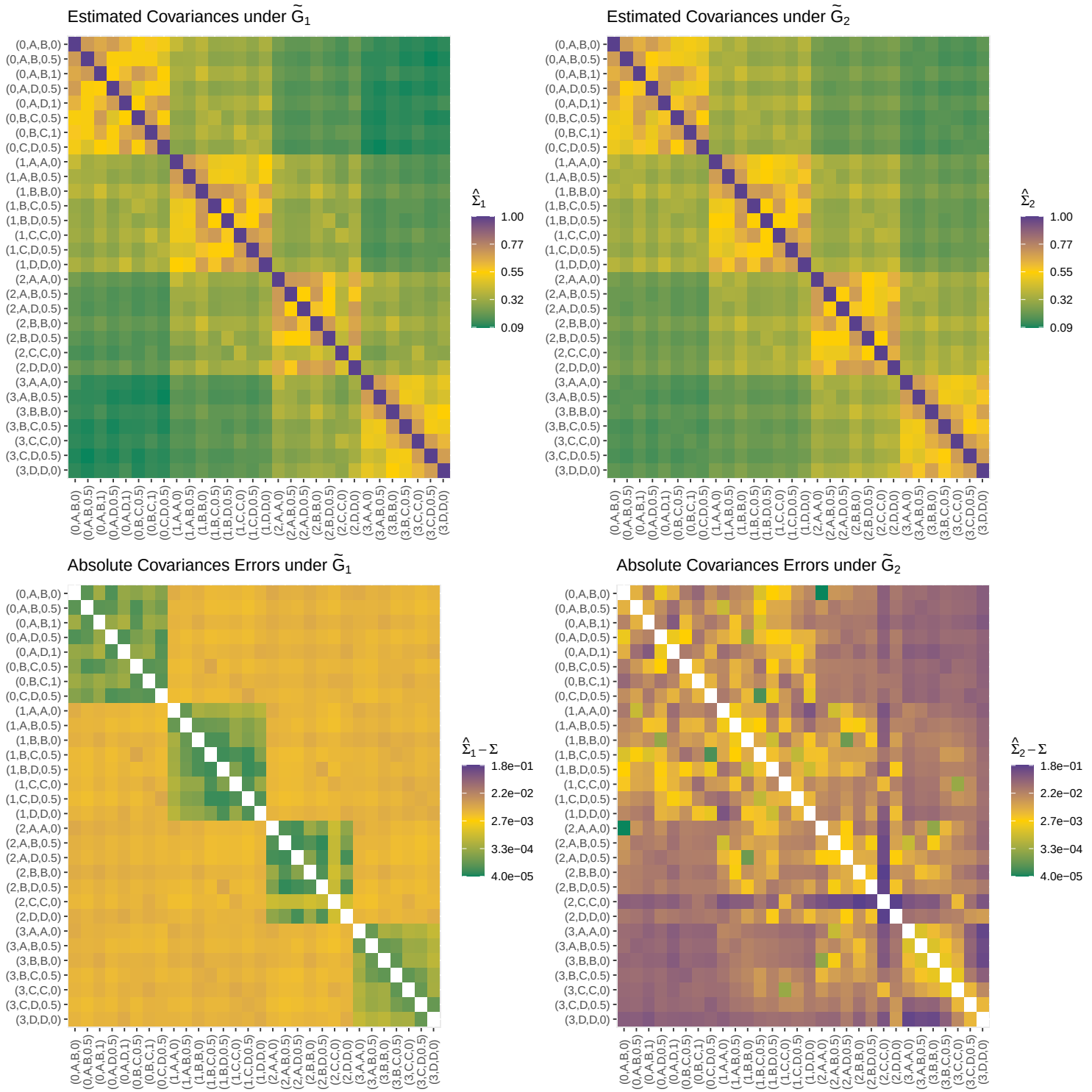


Figure 3.24: **3rd Example with function** $\psi_1(x) = e^{-x}$**.** Estimated covariance matrices $\hat{\boldsymbol{\Sigma}}$ (top) and covariances errors $\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

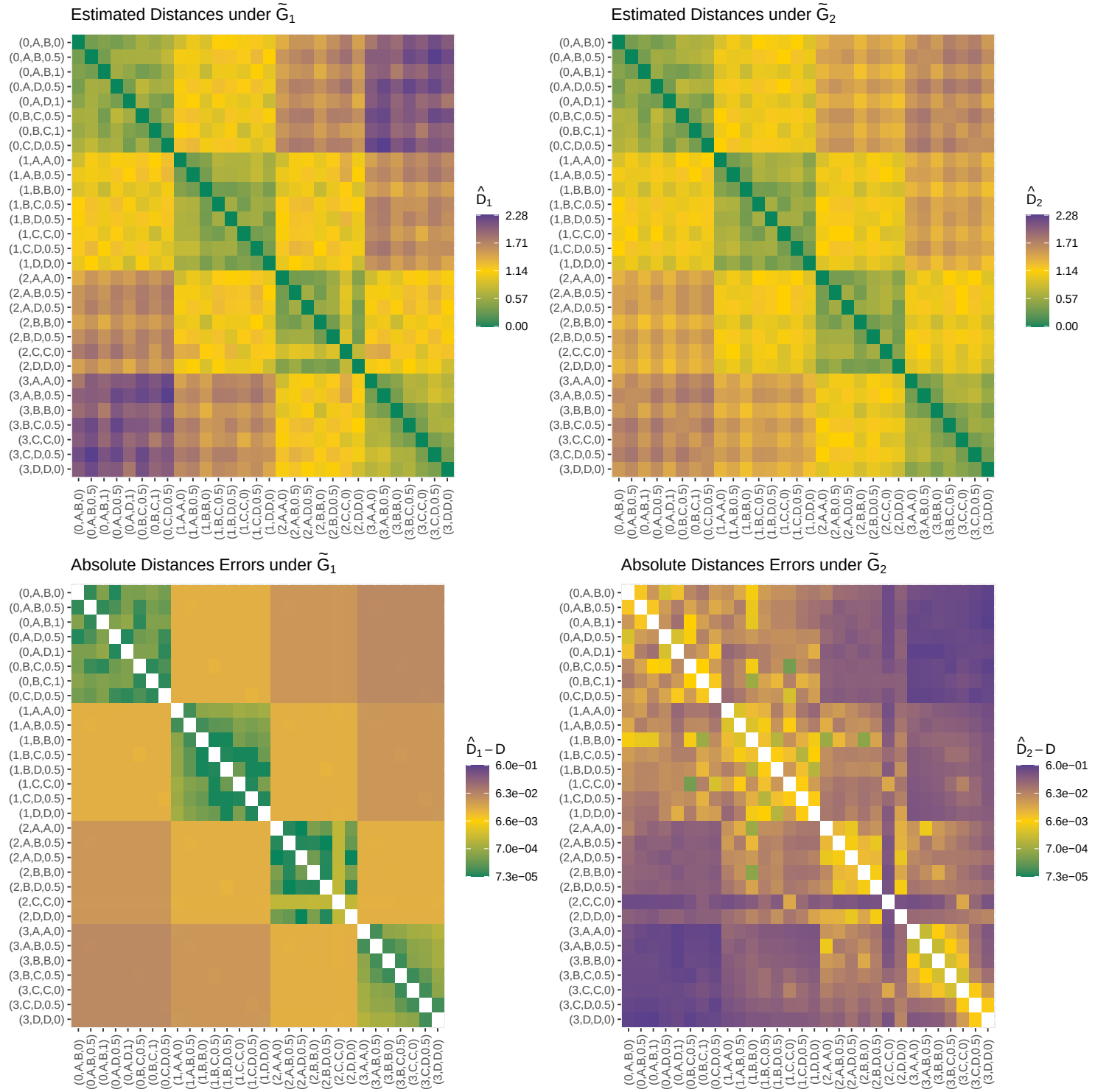


Figure 3.25: **3rd Example with function** $\psi_2(x) = \frac{1}{1+x}$**.** Estimated distance matrices $\hat{\mathbf{D}}$ (top) and distance errors $\hat{\mathbf{D}} - \mathbf{D}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

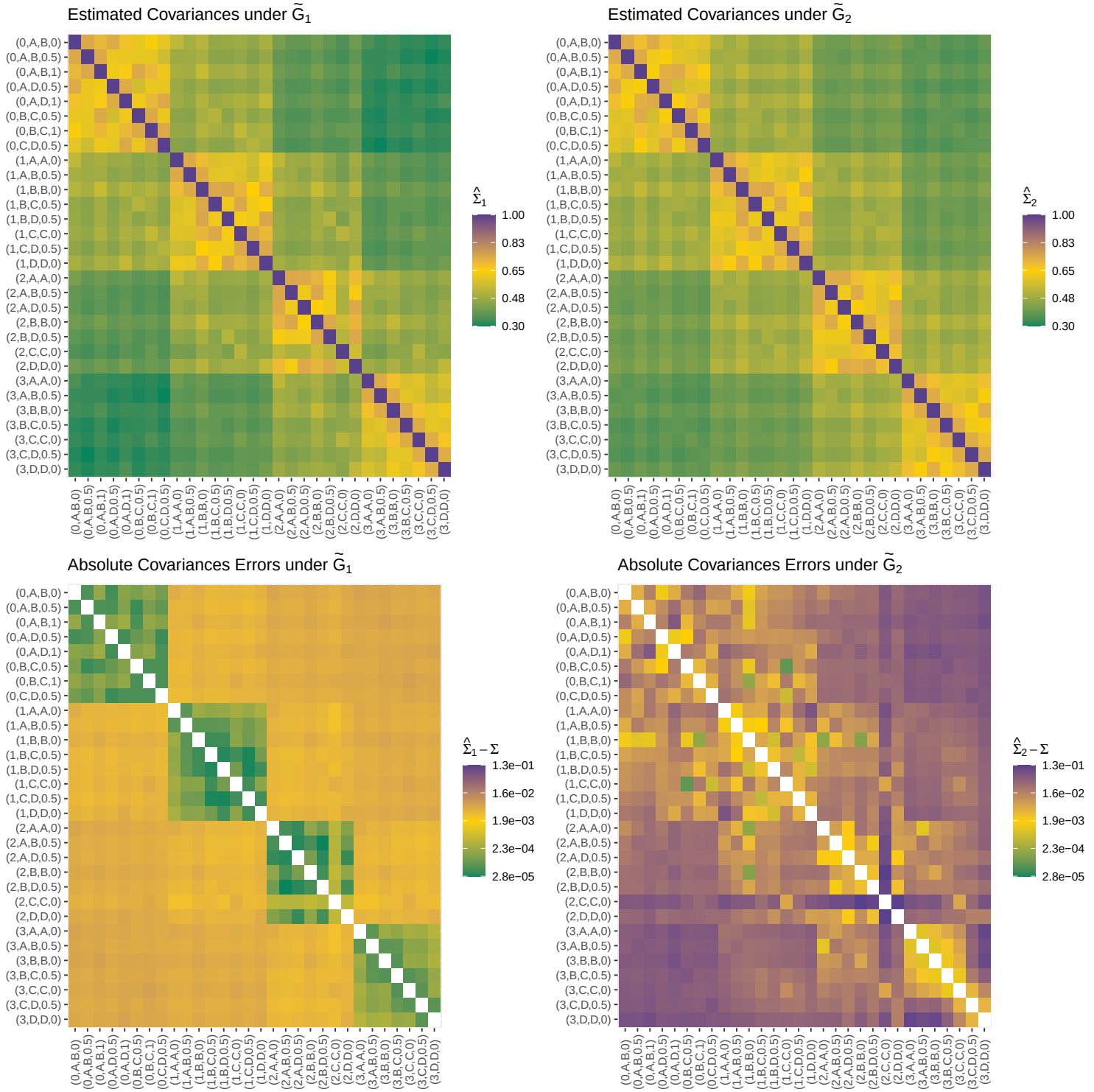


Figure 3.26: **3rd Example with function** $\psi_2(x) = \frac{1}{1+x}$**.** Estimated covariance matrices $\hat{\boldsymbol{\Sigma}}$ (top) and covariances errors $\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ (bottom) under the two graphs $\widetilde{\boldsymbol{G}}_1$ (left) and $\widetilde{\boldsymbol{G}}_2$ (right). Each row/column represents a spatio-temporal point, denoted by $(t, \underline{e}, \overline{e}, \delta(u))$, accordingly to Definition 3.7.

Chapter 4

Vector-Valued Gaussian Processes and their Kernels on a Class of Metric Graphs[1]

4.1 Introduction

4.1.1 Context: Random Fields on Networks

Networks analysis has become ubiquitous in several branches of theoretical and applied sciences, including mathematics, statistics, machine learning, artificial intelligence, with applications connected to the data planet. Networks can be used, either, to represent connections between random variables, or to provide a topological structure where a given process (called a random field) is observed. Our paper deals with the second situation, under the condition that the observed process is a vector-valued Gaussian random field that is second-order isotropic in the sense that the matrix-valued covariance kernel depends on distance between the points in the network.

Random fields can be continuously (Anderes et al., 2020) or discretely indexed over a network; for the second case, the common nomenclature is that of point processes on networks (Moradi and Mateu, 2020; Baddeley et al., 2021; Rakshit et al., 2017). For both cases, defining a random field over a network is a challenging task.

The electrical engineering and machine learning communities have been very active on the subject of networks, and the reader is referred to Ortega et al. (2018) and Chami et al. (2022) for a comprehensive review.

For this manuscript, the network defines a topological structure representing the domain of definition of a given random field. This fact is exploited by Anderes et al. (2020): endowing the network with a metric provides a metric space that is then (quasi) embeddable over Hilbert spaces. This is the crux of the argument that allows to build isotropic covariance functions over networks, where isotropy is

[1]This chapter is a version of the paper Filosi et al. (2025b), available on arXiv:2501.10208.

understood with respect to two alternative metrics (the geodesic and the resistance metrics). A relevant remark is that networks endowed with metrics can be suitably represented through metric graphs. This approach is adopted by Bolin and Lindgren (2011), Anderes et al. (2020) and Bolin et al. (2024). We shall make use of this fact subsequently. Scalar random fields that evolve temporally over graphs have been challenged by Porcu et al. (2023), Filosi et al. (2025a) (see Chapter 3) and Tang and Zimmerman (2024).

The way a metric graph can be specified is certainly not unique. Anderes et al. (2020) work over graphs with Euclidean edges — called generalised networks in Porcu et al. (2023) — which extend linear networks to non-linear edges. Further, the random field defined over such structures can have realisations over any point over the edges, and not only in the nodes. Roughly, these are 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 fact that linear networks are overly limited to provide a suitable topological structure to random fields is well understood in the machine learning literature (Alsheikh et al., 2014; Hamilton et al., 2017; Borovitskiy et al., 2023), as well as in spatial statistics (Cressie et al., 2006; Ver Hoef et al., 2006; Peterson et al., 2013, 2007; Montembeault et al., 2012; Xiao et al., 2017; Perry and Wolfe, 2013; Deng et al., 2014; Baddeley et al., 2017).

4.1.2 Challenge

Our paper is unique according to the state of the art. The literature on vector-valued random fields on (linear or generalised) networks is elusive. This is not surprising: kernels on metric graphs are a very recent subject, and so far the main efforts have been focused to the case of scalar-valued random fields.

A subset of the graphs used in this Chapter is represented by the so-called Euclidean trees with a given number of leaves, which are substantially linear networks. For such a case, since a linear network is embedded into the two-dimensional Euclidean space, any matrix-valued covariance functions depending on the ℓ_1 norm (the so-called Manhattan distance) can be used as a covariance function for a vector-valued field on a linear network. The reader is referred to Zastavnyi (2000) for relevant results in terms of isometric embeddings. Unfortunately, there are two problems related to this fact:

1. To the knowledge of the authors, there are no models for matrix-valued covariances depending on the ℓ_1 distance;

2. Such constructions are no longer valid for the case of generalised networks.

Another relevant aspect is related to what we term *cross metrics* in the paper. To illustrate the principle, consider two random fields — call them Z_1 and Z_2. Since the network does not necessarily represent a geographic space, there is no reason

to believe that the distance between $Z_1(u)$ and $Z_2(v)$, for two different points u, v on the network, should be the same as the distance between $Z_1(u)$ and $Z_1(v)$ for instance. We take this aspect into account and devote attention to matrix-valued metrics.

4.1.3 Our Constructions: Road Map

Our constructions are based on the following steps. We provide here a concise description. Careful notation and terminology are established in Section 4.2.
The Road Map

1. The purpose of the work is to construct a p-variate Gaussian random field, denoted $\boldsymbol{Z} := \{\boldsymbol{Z}(u) = (Z_1(u), \ldots, Z_p(u))^\top,\ u \in G\}$ throughout and defined over a metric graph, G, with a given matrix-valued covariance function that is isotropic, that is it depends on the *distance* between the points of the graph. Hence, the topological space G, endowed with a proper matrix-valued (semi) metric $\mathbf{D} : G \times G \to [0, +\infty)^{p\times p}$, becomes a (semi) metric space $(G, \mathbf{D})$.

2. Such a (semi) metric space can be (quasi) embedded on a suitable Hilbert space, which in turns allows to resort classical machinery for covariance functions on metric spaces (Schoenberg, 1942).

3. Hence, a substantial part of the job is to build the multivariate metric $\mathbf{D}$.

4. We then take advantage of isometric embedding arguments as in Anderes et al. (2020) to claim that, for a certain class of continuous mappings $\psi : [0, +\infty) \to (0, +\infty)$, the element-wise composition

$$\mathbf{K}(u, v) := \psi\left(\mathbf{D}(u, v)\right), \qquad u, v \in G, \tag{4.1}$$

   provides a valid matrix-valued covariance function.

5. We then generalise this construction by considering a matrix-valued class of continuous mappings $\boldsymbol{\Psi} : [0, +\infty)^{p\times p} \to [0, +\infty)^{p\times p}$ having elements Ψ_{ij}, such that the element-wise composition

$$\mathbf{K}(u, v) := \boldsymbol{\Psi}\left(\mathbf{D}(u, v)\right) = \left[\Psi_{ij}(D_{ij}(u, v))\right]_{i,j=1}^{p}, \qquad u, v \in G, \tag{4.2}$$

   is a valid matrix-valued covariance function.

In turn, the construction of the multivariate metric $\mathbf{D}$ requires a collection of steps, which entails several technical challenges:

(a) The vector-valued random field $\boldsymbol{Z}$ is built through the sum of two random fields, $\boldsymbol{Z}_V$ and $\boldsymbol{Z}_E$, where the subscripts V and E stand, respectively, for the sets containing vertices and the edges associated with the metric graph, G. Both random fields are continuously indexed over the graph;

(b) The random field $\boldsymbol{Z}_V$ at any point of an edge, $e \in E$, is constructed through interpolation of the random field defined at the extremes of each edge (which are obviously vertices in V);

(c) The multivariate metric is then checked for desiderata. Specifically, that the marginal metrics (those in the diagonal of $\mathbf{D}$) respect the ingenious univariate condition provided by Anderes et al. (2020). Further, the cross-elements in $\mathbf{D}$ are required to be homogeneous, in the sense that $D_{ij} = D_{i'j'}$ for $i \neq j$ and $i' \neq j'$, for D_{ij} a generic element of $\mathbf{D}$.

For a subclass termed *Euclidean trees*, we provide multivariate kernels with compact support depending on the so-called Manhattan distance.

The structure of the paper is the following. Section 4.2 provides the necessary background and notation. Results are provided in Sections 4.3 and 4.4. Section 4.5 concludes the paper. As proofs are rather technical and lengthy, they are deferred to Section 4.6.

Finally, to guide the reader into the structure of the results of this manuscript, we provide a road-map in Figure 4.1.

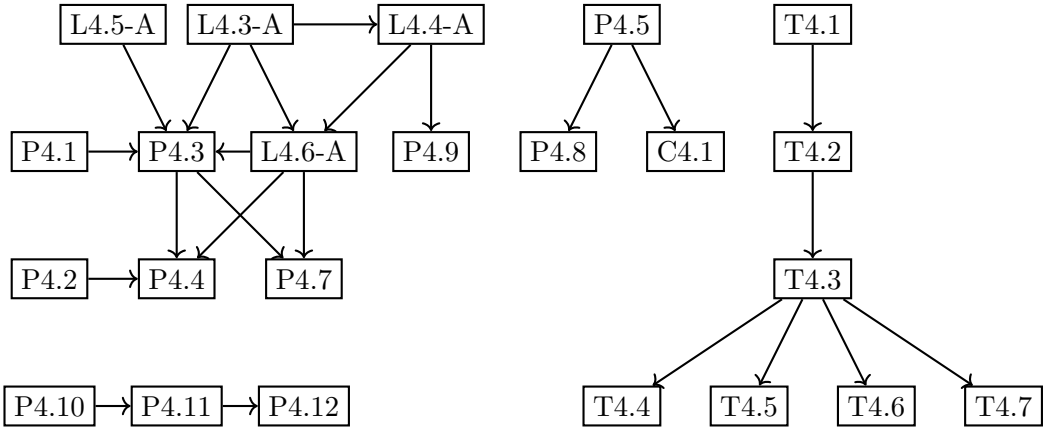


Figure 4.1: Road-map of the main results in this manuscript. L, P, T and C stand for Lemma, Proposition, Theorem and Corollary, respectively. The suffix -A indicates that the result is stated at the end of the chapter, in Section 4.6. The missing results have either no direct connection with the above results (L4.1, P4.6) or are essential for many others (L4.2).

4.2 Mathematical Background

4.2.1 Notation Handout

Throughout, p is an integer greater than 1 (the case $p = 1$ has been treated by Anderes et al. (2020)). Bold lowercase letters denote (deterministic) vectors (*e.g.*

$\boldsymbol{\delta}$). Bold uppercase letters denote (deterministic) matrices (*e.g.* $\mathbf{A}$). Bold uppercase italic letters denote random vectors (*e.g.* $\boldsymbol{Z}$). $\mathbf{A}^\top$, $\boldsymbol{\delta}^\top$ and $\boldsymbol{Z}^\top$ are the transposes of $\mathbf{A}$, $\boldsymbol{\delta}$ and $\boldsymbol{Z}$, while $\mathbf{A}^-$ stands for the Moore-Penrose inverse of $\mathbf{A}$. Non-bold letters represent scalar (real-valued) quantities or non-numerical objects (sets, graphs, vertices, edges, or points). The (i,j)-th entry of $\mathbf{A}$ is denoted A_{ij}, while the i-th components of $\boldsymbol{\delta}$ and $\boldsymbol{Z}$ are δ_i and Z_i, respectively. The all-ones vector of size n, all-ones matrix of size $n \times n$, and identity matrix of size $n \times n$ are denoted as $\mathbf{1}_n$, $\mathbf{1}_{n\times n}$ and $\mathbf{I}_n$, respectively. Finally, we define $\mathbf{J}_n := \operatorname{diag}[0, 0, \ldots, 0, 1]^\top$ ($n \times n$ diagonal matrix with only one non-zero entry).

Continuity, differentiation, integration and scalar- or matrix-valued functions (except power functions) involving matrices are understood as being performed element-wise. Concerning the power functions, $\mathbf{A}^2$ is $\mathbf{A}$ multiplied by itself, $\sqrt{\mathbf{A}} = \mathbf{A}^{1/2}$ is the principal square root of $\mathbf{A}$, and $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$ in the matrix sense. Products (*e.g.* $\mathbf{AB}$ or $\mathbf{A}\boldsymbol{\delta}$) and Kronecker products (*e.g.* $\mathbf{A} \otimes \mathbf{B}$) are also understood in the matrix sense, not as element-wise operations.

4.2.2 Gaussian Random Gields on Graphs with Euclidean Edges

The paper deals with a vector-valued Gaussian random field $\boldsymbol{Z}$ defined on a metric graph as exposed below. We assume a random field with zero-mean vector and with a matrix-valued covariance matrix $\mathbf{K} : G \times G \to \mathbb{R}^{p\times p}$ that is isotropic in the sense there exists a pair $(\boldsymbol{\Psi}, \mathbf{D})$, with $\boldsymbol{\Psi} : [0, +\infty)^{p\times p} \to \mathbb{R}^{p\times p}$ and $\mathbf{D} : G \times G \to [0, +\infty)^{p\times p}$, both continuous, such that Equation (4.2) is satisfied. We work with a specific class of metric graphs, termed graphs *with Euclidean edges* as being proposed by Anderes et al. (2020) to generalise linear networks. The main idea underlying this construction is to associate every edge of the graph to a segment of the real line. As a consequence, it is possible to consider both the vertices and the points over the edges as actual points of the graph, where distances can be computed and random fields can be defined. As this is natural on linear networks as well, we stress that graphs with Euclidean edges do not have any restriction on the topology, *i.e.* the edges' lengths are free to vary and so are the connections between the vertices. Here we rephrase their original definition.

Definition 4.1 (Graph with Euclidean edges)**.** The type of topology considered by Anderes et al. (2020) is called graph *with Euclidean edges*, denoted with a triple $G = (V, E, \{\varphi_e\}_{e\in E})$ throughout, where the elements are blended in the following way:

(a) (V, E) has a graph structure, *i.e.* V is the set of vertices and $E \subset V \times V$ accounts for the edges. We assume that this graph is simple and connected, *i.e.* V is finite, the graph has not repeated edges or edges that join a vertex to itself and every pair of vertices is connected by a path.

(b) Each edge $e \in E$ is provided with a *length* $\ell(e) > 0$ and a *weight* $w(e) := 1/\ell(e)$; furthermore, it is associated with a unique abstract set, also denoted e, such

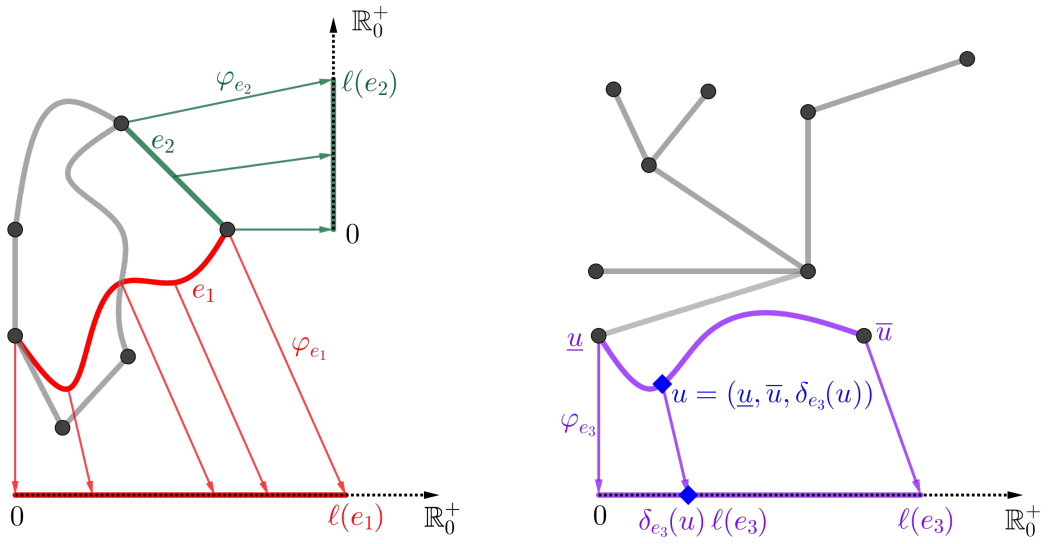


Figure 4.2: Left: a graph with Euclidean edges, where the bijections φ_{e_1} and φ_{e_2} have been highlighted. Right: a Euclidean tree with 5 leaves where the role of $\underline{u}$, $\overline{u}$ and $\delta_{e_3}(u)$ have been stressed. Adaptation of Filosi et al. (2025a, Figure 1) (see Chapter 3).

that V and all the edge sets are mutually disjoint.

(c) Let v_1 and v_2 be vertices connected by $e \in E$. Then, φ_e is a continuous and bijective mapping defined on $e \cup \{v_1, v_2\}$, such that φ_e maps e onto an open interval $(0, \ell(e)) \subset \mathbb{R}$ and $\{v_1, v_2\}$ onto $\{0, \ell(e)\}$.

Henceforth, we shall assume the existence of a total order relation on the set of vertices V and that every edge is represented through the ordered pair (v_1, v_2), where $v_1 < v_2$. By abuse of notation, we write $u \in G$ to denote a point on the graph: it can be either a vertex or a point on an edge. Each point $u \in G$ will be identified with the triple $(\underline{u}, \overline{u}, \delta)$, where: $\underline{u} < \overline{u}$ are the endpoints of the edge e containing u and $\delta \in [0, 1]$ is the relative distance of u from $\underline{u}$; formally: $\delta = \delta(u) := w(e)\varphi_e(u)$. Notice that, whenever u is a vertex, it is always possible to write $u = (\underline{u}, \overline{u}, \delta)$ for $\delta \in \{0, 1\}$, though the choice of $\underline{u}$ and $\overline{u}$ may not be unique. Finally, we use the notation $e(u)$, where $e : G \to E$, to denote the edge that contains u. If $u \in V$, then $e(u)$ is any of the edges that have an endpoint in u. It is in order to note that our definition slightly deviates from the original one in Anderes et al. (2020): we do not require any distance consistency property, as we do not use the geodesic metric.

A graph with Euclidean edges is called an *Euclidean tree* if it has a tree structure, *i.e.* given any two points $u_1, u_2 \in G$, there is exactly one path that connects u_1 to u_2. A vertex of a Euclidean tree is called a *leaf* if it is connected to only one other vertex. Figure 4.2 aims to clarify the notions introduced so far.

4.2.3 Classes of Matrices Used in the Paper

An $n \times n$ real matrix $\mathbf{A}$ is said to be *positive semidefinite* if and only if $\boldsymbol{c}^\top \mathbf{A} \boldsymbol{c} \geq 0$ for all $\boldsymbol{c} \in \mathbb{R}^n$, and *conditionally negative semidefinite* if and only if $\boldsymbol{c}^\top \mathbf{A} \boldsymbol{c} \leq 0$ for all $\boldsymbol{c} \in \mathbb{R}^n$ such that ${\mathbf{1}_n}^\top \boldsymbol{c} = 0$.

An $n \times n$ matrix $\mathbf{L}$ is called *quasi-Laplacian* if and only if it is symmetric, positive semidefinite, and has exactly one null eigenvalue, with corresponding eigenvector $\mathbf{1}_n$. A quasi-Laplacian matrix $\mathbf{L}$ is called *Laplacian* if and only if it has non-positive off-diagonal entries. Notice that Laplacian matrices arise naturally from weighted graphs: there is a bijection between the set of (positively) weighted, simple and undirected graphs and the set of Laplacian matrices (Devriendt, 2022, Subsection 2.1). Henceforth, we will consider the Laplacian matrix of a given graph with Euclidean edges, with entries defined as Devriendt (2022, Equation (1)):

$$L_{ij} = \begin{cases} \sum_{v \in V} w((v, v_i)) \text{ if } i = j \\ -w((v_i, v_j)) \text{ if } i \neq j, \end{cases} \tag{4.3}$$

with $w((v_i, v_j))$ the weight of the edge joining vertices v_i and v_j if these vertices are connected, 0 otherwise.

Quasi-Laplacian matrices have been introduced as a generalisation of Laplacian matrices, as they will often appear in this manuscript. Their interpretation is slightly blurred, since, thinking about the bijection between Laplacian matrices and weighted graphs, it seems natural to interpret quasi-Laplacian matrices as Laplacian matrices of graphs with possibly negative weights. Nevertheless, as one may expect, quasi-Laplacian matrices share several properties of Laplacian matrices formally stated below.

Proposition 4.1. *Let* $\mathbf{L}$ *be a quasi-Laplacian matrix partitioned as follows:*

$$\mathbf{L} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{C} \end{bmatrix}, \tag{4.4}$$

where $\mathbf{A}$ *and* $\mathbf{C}$ *are square matrices. Then the following statements hold.*

1. $\mathbf{B} \neq \mathbf{0}$.
2. $\mathbf{A}$ *and* $\mathbf{C}$ *are strictly positive definite.*
3. *Both* $\mathbf{L}/\mathbf{C} := \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\top$ *and* $\mathbf{L}/\mathbf{A} := \mathbf{C} - \mathbf{B}^\top\mathbf{A}^{-1}\mathbf{B}$ *are quasi-Laplacian matrices, i.e. , the set of quasi-Laplacian matrices is closed under the Schur complement operation.*
4. $\mathbf{L}^-$ *is a quasi-Laplacian matrix.*

Proof of Proposition 4.1. Throughout this proof, we assume that $\mathbf{L} \in \mathbb{R}^{n \times n}$, $\mathbf{A} \in \mathbb{R}^{n_1 \times n_1}$ and $\mathbf{C} \in \mathbb{R}^{n_2 \times n_2}$, with $n_1, n_2 > 0$ and $n = n_1 + n_2$.

1. Assume $\mathbf{B} = 0$. Then $n-1 = \text{rank}(\mathbf{L}) = \text{rank}(\mathbf{A})+\text{rank}(\mathbf{C})$. But since $\mathbf{L}\mathbf{1}_n = \mathbf{0}$, it follows that $\mathbf{A}\mathbf{1}_{n_1} = \mathbf{0}$ and $\mathbf{C}\mathbf{1}_{n_2} = \mathbf{0}$. This implies that $\text{rank}(\mathbf{A}) \leq n_1 - 1$ and $\text{rank}(\mathbf{C}) \leq n_2 - 1$. Contradiction.

2. We will show only that $\mathbf{C}$ strictly positive definite, as the proof for the other case relies on the same argument. Since $\mathbf{C}$ is a principal submatrix of $\mathbf{L}$, it is positive semidefinite. As a consequence, it is sufficient to show that it is invertible. Assume, by contradiction, that $\mathbf{C}$ is singular. This is equivalent to assume that $\exists \boldsymbol{y} \neq \mathbf{0}$ such that $\mathbf{C}\boldsymbol{y} = \mathbf{0}$. Since $\mathbf{L}$ is positive semidefinite we have, for each $\boldsymbol{x} \in \mathbb{R}^{n_1}$ and for each $\gamma \in \mathbb{R}$:

$$\begin{aligned} 0 &\leq \begin{bmatrix} \boldsymbol{x} \\ \gamma\boldsymbol{y} \end{bmatrix}^\top \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{C} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \gamma\boldsymbol{y} \end{bmatrix} \\ &= \boldsymbol{x}^\top\mathbf{A}\boldsymbol{x} + \boldsymbol{x}^\top\mathbf{B}(\gamma\boldsymbol{y}) + (\gamma\boldsymbol{y})^\top\mathbf{B}^\top\boldsymbol{x} + (\gamma\boldsymbol{y})^\top\mathbf{C}(\gamma\boldsymbol{y}) = \boldsymbol{x}^\top\mathbf{A}\boldsymbol{x} + 2\gamma\boldsymbol{x}^\top\mathbf{B}\boldsymbol{y}. \end{aligned}$$

   If $\boldsymbol{x}^\top\mathbf{B}\boldsymbol{y} \neq 0$, we can find a suitable γ that contradicts the previous inequality. Hence $\boldsymbol{x}^\top\mathbf{B}\boldsymbol{y} = 0$ and therefore, being $\boldsymbol{x}$ arbitrary, $\mathbf{B}\boldsymbol{y} = \mathbf{0}$. Consider now the vector $\boldsymbol{z} := (0, ..., 0, \boldsymbol{y}^\top)^\top$: we have $\mathbf{L}\boldsymbol{z} = \mathbf{0}$, but clearly $\boldsymbol{z}$ does not lie in the null space of $\mathbf{L}$, which is $\{\lambda\mathbf{1}_n : \lambda \in \mathbb{R}\}$. Contradiction.

3. Clearly $\mathbf{L}/\mathbf{C}$ is symmetric. Furthermore, the Guttman rank additivity formula (Zhang, 2005, Equation 0.9.2) states that $\text{rank}(\mathbf{L}) = \text{rank}(\mathbf{C})+\text{rank}(\mathbf{L}/\mathbf{C})$. Here we have $\text{rank}(\mathbf{L}) = n - 1$ and $\text{rank}(\mathbf{C}) = n_2$, as a consequence $\text{rank}(\mathbf{L}/\mathbf{C}) = n_1 - 1$. Furthermore, from Zhang (2005, Theorem 1.12), being $\mathbf{C}$ strictly positive definite and L positive semidefinite, we have that $\mathbf{L}/\mathbf{C}$ is positive semidefinite as well. The last thing that is left to be shown is that the null eigenvector of $\mathbf{L}/\mathbf{C}$ is $\mathbf{1}_{n_1}$. Since $\mathbf{1}_n$ is the null eigenvector of $\mathbf{L}$, we get

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \mathbf{1}_n = \mathbf{0} \qquad \text{and} \qquad \begin{bmatrix} \mathbf{B}^\top & \mathbf{C} \end{bmatrix} \mathbf{1}_n = \mathbf{0},$$

   videlicet $\mathbf{A}\mathbf{1}_{n_1} + \mathbf{B}\mathbf{1}_{n_2} = \mathbf{0}$ and $\mathbf{B}^\top\mathbf{1}_{n_1} + \mathbf{C}\mathbf{1}_{n_2} = \mathbf{0}$. Using these relations, the following chain of equivalences concludes the proof:

$$\begin{aligned} &(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\top)\mathbf{1}_{n_1} = \mathbf{0} \iff \mathbf{A}\mathbf{1}_{n_1} = \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\top\mathbf{1}_{n_1} \\ &\iff -\mathbf{B}\mathbf{1}_{n_2} = \mathbf{B}\mathbf{C}^{-1}(-\mathbf{C}\mathbf{1}_{n_2}) \iff -\mathbf{B}\mathbf{1}_{n_2} = -\mathbf{B}\mathbf{1}_{n_2}. \end{aligned}$$

4. Consider the eigendecomposition of $\mathbf{L}$:

$$\mathbf{L} = \mathbf{W} \begin{bmatrix} \boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \mathbf{W}^\top, \tag{4.5}$$

   with $\boldsymbol{\Delta}$ a $(n-1) \times (n-1)$ diagonal matrix with positive diagonal entries. Since $\mathbf{W}$ is an orthogonal matrix, we have

$$\mathbf{L}^- = \left(\mathbf{W} \begin{bmatrix} \boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \mathbf{W}^\top \right)^- = (\mathbf{W}^\top)^- \begin{bmatrix} \boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}^- \mathbf{W}^- = \mathbf{W} \begin{bmatrix} \boldsymbol{\Delta}^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \mathbf{W}^\top. \tag{4.6}$$

Therefore $\mathbf{L}^-$ is symmetric, is positive semidefinite and with exactly one null eigenvalue, corresponding to the same null-space eigenvector of $\mathbf{L}$.

□

4.2.4 Metrics Used in the Paper

The work by Anderes et al. (2020) shows that there is no unique way to define a metric over a graph with Euclidean edges. While the geodesic distance is physically intuitive, Anderes et al. (2020) prove that it has very limited use in terms of available covariance functions. Further, when considering the geodesic distance not all graphs with Euclidean edges become *permissible* (Anderes et al., 2020), and a collection of technical restrictions is required to make things work. Alternatively, one can use the resistance metric, being a generalisation of the electric distance used for electric circuits (Klein and Randic, 1993). To provide a description of the resistance metric, some further background is needed. Let X be a set and $Z : X \to \mathbb{R}$ a square-integrable random field. The *variogram* of Z is defined via

$$\begin{aligned} \gamma_Z : X \times X &\to \mathbb{R} \\ (x_1, x_2) &\mapsto \gamma_Z(x_1, x_2) := \mathbb{V}\mathrm{ar}\left(Z(x_1) - Z(x_2)\right). \end{aligned}$$

Clearly, a variogram is always symmetric and non-negative valued. In addition, it is conditionally negative semidefinite, *i.e.* it satisfies (Chilès and Delfiner, 2012)

$$\sum_{i,j=1}^{n} c_i c_j \gamma(x_i, x_j) \leq 0 \tag{4.7}$$

whenever $\sum_{i=1}^{n} c_i = 0$, for any $x_1, \ldots, x_n \in X$.

Anderes et al. (2020) define the resistance metric, denoting d_R throughout, as the variogram of a random field Z on the graph with Euclidean edges G, that is constructed *ad hoc* in order to be as much as possibly parenthetical to a Brownian bridge on the graph. Technicalities are deferred to their paper.

The concept of variogram can be generalised to the case of a zero-mean square-integrable vector-valued random field $\boldsymbol{Z}$ on a non-empty set X, through the following lines. The *pseudo-variogram* of $\boldsymbol{Z}$ is defined through the identity

$$\begin{aligned} \boldsymbol{\Gamma}_{\boldsymbol{Z}} : X \times X &\to [0, +\infty)^{p \times p} \\ (x_1, x_2) &\mapsto \boldsymbol{\Gamma}_{\boldsymbol{Z}}(x_1, x_2) := \Big[\mathbb{V}\mathrm{ar}\left(Z_i(x_1) - Z_j(x_2)\right) \Big]_{i,j=1}^{p}. \end{aligned} \tag{4.8}$$

Straightforward calculations allow to rewrite the above as

$$\begin{aligned} \boldsymbol{\Gamma}_{\boldsymbol{Z}}(x_1, x_2) = (\operatorname{diag} \mathbb{V}\mathrm{ar}\, \boldsymbol{Z}(x_1))\ \mathbf{1}_p{}^\top + \mathbf{1}_p\ (\operatorname{diag} \mathbb{V}\mathrm{ar}\, \boldsymbol{Z}(x_2))^\top \\ - \mathbb{C}\mathrm{ov}\left(\boldsymbol{Z}(x_1), \boldsymbol{Z}(x_2)\right) - \mathbb{C}\mathrm{ov}\left(\boldsymbol{Z}(x_2), \boldsymbol{Z}(x_1)\right), \end{aligned} \tag{4.9}$$

where $\mathbb{V}\mathrm{ar}\,\boldsymbol{Z}(x_1)$ is the $p \times p$ collocated variance-covariance matrix of $\boldsymbol{Z}(x_1)$, $\mathrm{diag}\,\mathbb{V}\mathrm{ar}\,\boldsymbol{Z}(x_1)$ is the $p \times 1$ vector containing the main diagonal of $\mathbb{V}\mathrm{ar}\,\boldsymbol{Z}(x_1)$ and $\mathbb{C}\mathrm{ov}\left(\boldsymbol{Z}(x_1), \boldsymbol{Z}(x_2)\right) := \mathbb{E}\left(\boldsymbol{Z}(x_1)\boldsymbol{Z}(x_2)^\top\right)$. The pseudo-variogram satisfies the following properties (Dörr and Schlather, 2023):

- for any $x_1, x_2 \in X$,
$$\boldsymbol{\Gamma}_{\boldsymbol{Z}}(x_1, x_2) = \boldsymbol{\Gamma}_{\boldsymbol{Z}}(x_2, x_1)^\top; \tag{4.10}$$
- for any $x \in X$, the diagonal entries of $\boldsymbol{\Gamma}_{\boldsymbol{Z}}(x, x)$ are zero;
- $\boldsymbol{\Gamma}_{\boldsymbol{Z}}$ is $\mathbf{1}_p$-conditionally negative semidefinite, *i.e.* for all $n \in \mathbb{N}^+$, $x_1, \ldots, x_n \in X$ and $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n \in \mathbb{R}^p$ such that ${\mathbf{1}_p}^\top \sum_{i=1}^n \boldsymbol{c}_i = 0$, it holds
$$\sum_{i,j=1}^{n} \boldsymbol{c}_i^\top \boldsymbol{\Gamma}_{\boldsymbol{Z}}(x_i, x_j)\boldsymbol{c}_j \leq 0. \tag{4.11}$$

In the following, we define the multivariate distance $\mathbf{D} : X \times X \to [0, +\infty)^{p \times p}$ via $\mathbf{D}(x_1, x_2) := \boldsymbol{\Gamma}_{\mathbf{Z}}(x_1, x_2)$. It is in order to notice that at this point the matrix-valued distance $\mathbf{D}$ is *not* symmetric, *i.e.* in general $D_{ij}(x_1, x_2) \neq D_{ji}(x_1, x_2)$. This may pave the way for new asymmetric metrics.

4.2.5 Classes of Functions and Schoenberg Characterisations

This section follows closely Zastavnyi (2023). A function $k : X \times X \to \mathbb{R}$ is positive semidefinite if, for all $n \in \{1, 2, \ldots\}$, $x_1, \ldots, x_n \in X$ and $c_1, \ldots, c_n \in \mathbb{R}$, one has

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0. \tag{4.12}$$

In addition, k is a strictly positive definite function if it is positive semidefinite and

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) = 0 \implies c_1 = \cdots = c_n = 0. \tag{4.13}$$

The extension to the matrix-valued case reads as follows. A matrix-valued function $\mathbf{K} : X \times X \to \mathbb{R}^{p \times p}$ is positive semidefinite if, for all $n \in \{1, 2, \ldots\}$, $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n \in \mathbb{R}^p$ and $x_1, \ldots, x_n \in X$, it holds:

$$\sum_{i,j=1}^{n} \boldsymbol{c}_i^\top \mathbf{K}(x_i, x_j)\boldsymbol{c}_j \geq 0. \tag{4.14}$$

In addition, $\mathbf{K}$ is strictly positive definite if it is positive semidefinite and the condition

$$\sum_{i,j=1}^{n} \boldsymbol{c}_i^\top \mathbf{K}(x_i, x_j)\boldsymbol{c}_j = 0 \tag{4.15}$$

implies that there exists a pair $i \neq j$ such that $x_i = x_j$ or $\boldsymbol{c}_1 = \cdots = \boldsymbol{c}_n = 0$. A relevant result for our developments is coming from Proposition 1 in Zastavnyi (2023), which is formally stated for a neater exposition.

Proposition 4.2. *Let* $\mathbf{K} : X \times X \to \mathbb{R}^{p\times p}$. *Then, the following are equivalent:*

- $\mathbf{K}$ *is positive semidefinite;*
- *for all* $n \in \{1, 2, \ldots\}$ *and for all* $x_1, \ldots, x_n \in X$, *the* $np \times np$ *block matrix* $\mathbf{C} := [C_{ij}]_{i,j=1}^{np} := [\mathbf{K}(x_i, x_j)]_{i,j=1}^{n}$ *is positive semidefinite, i.e. ,* $\boldsymbol{c}^\top \mathbf{C}\boldsymbol{c} \geq 0$ *for any* $\boldsymbol{c} \in \mathbb{R}^{np}$.

A function $\psi : [0, +\infty) \to [0, +\infty)$ is called *completely monotone* if it is continuous on $[0, +\infty)$, infinitely differentiable on $(0, +\infty)$ and for each $i \in \mathbb{N}$ it holds

$$(-1)^i \, \psi^{(i)}(x) \geq 0,$$

where $\psi^{(i)}$ denotes the i^{th} derivative of ψ and $\psi^{(0)} := \psi$. A function $\psi : [0, +\infty) \to [0, +\infty)$ is completely monotone if and only if it is the Laplace transform of a (unique) finite Borel measure μ on $[0, +\infty)$, *i.e.*

$$\psi(x) = \int_0^{+\infty} \exp(-tx) \, \mu(\mathrm{d}t), \quad x \in [0, +\infty). \tag{4.16}$$

A function $g : [0, +\infty) \to [0, +\infty)$ is called a *Bernstein function* if g in infinitely differentiable on $[0, +\infty)$ and $g^{(1)}$ is a completely monotone function.

Our expository material is now completed by reporting Theorem 5 of Zastavnyi (2023).

Theorem 4.1. *Let* $\psi : [0, +\infty) \to (0, +\infty)$ *be a non-constant completely monotone function,* $g : [0, +\infty) \to [0, +\infty)$ *be a Bernstein function and let* $\boldsymbol{\Gamma}_{\boldsymbol{Z}} : X \times X \to [0, +\infty)^{p\times p}$, *for a non-empty set* X. *If, for all* $x_1, x_2 \in X$, $\boldsymbol{\Gamma}_{\boldsymbol{Z}}^\top(x_1, x_2) = \boldsymbol{\Gamma}_{\boldsymbol{Z}}(x_2, x_1)$ *and* $\boldsymbol{\Gamma}_{\boldsymbol{Z}}$ *is* $\mathbf{1}_p$*-conditionally negative semidefinite, then, for all* $\xi > 0$,

$$(x_1, x_2) \mapsto \psi\Big(\xi \, g\left(\boldsymbol{\Gamma}_{\boldsymbol{Z}}(x_1, x_2)\right)\Big), \qquad x_1, x_2 \in X,$$

where ψ *and* g *are applied element-wise, is a positive semidefinite function on* X.

4.3 General Results

The following notation will ease the exposition throughout:

- n is the number of vertices of the graph G (cardinality of V),
- $\mathbf{L}$ is the Laplacian matrix of G as per Equation (4.3),
- $\boldsymbol{\Sigma} := \mathbf{L}^-$ is the Moore-Penrose inverse of $\mathbf{L}$,

- $\mathbf{\Lambda}$ and $\mathbf{W}$ are the eigenvalues and eigenvectors matrices of $\mathbf{\Sigma}$, *i.e.* $\mathbf{\Sigma} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^\top$,
- for two matrices $\mathbf{A}, \mathbf{B}$ of the same size and for a positive integer p, we define

$$\mathcal{M}_p(\mathbf{A}, \mathbf{B}) := \mathcal{M}_p \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} := \mathbf{I}_p \otimes \mathbf{A} + (\mathbf{1}_{p\times p} - \mathbf{I}_p) \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A} & \mathbf{B} & \dots & \mathbf{B} \\ \mathbf{B} & \mathbf{A} & \dots & \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B} & \mathbf{B} & \dots & \mathbf{A} \end{bmatrix}. \tag{4.17}$$

The construction provided in this section extends, to the matrix-valued case, that in Anderes et al. (2020). Specifically, we shall consider a vector-valued random field, $\boldsymbol{Z}$, that is obtained through the identity

$$\boldsymbol{Z}(u) := \boldsymbol{Z}_V(u) + \boldsymbol{Z}_E(u), \qquad u \in G. \tag{4.18}$$

In turn, the random fields $\boldsymbol{Z}_V$ and $\boldsymbol{Z}_E$ obeys to different construction principles. Specifically, $\boldsymbol{Z}_V$ is a multivariate Gaussian random vector on the vertices and is linearly interpolated on the edges, whilst $\boldsymbol{Z}_E$ is zero on the vertices and adds some variability on the edges. We note that Anderes et al. (2020) consider this construction for the case $p = 1$ and this entails the use of a specific class of variograms that are then used to build the resistance metric, d_R. It is not surprising that our construction will rely on the pseudo-variogram $\mathbf{\Gamma}_{\boldsymbol{Z}}$ for the generalisation to the p-variate setting with $p > 1$. While these constructions are mathematically involved, a simple description is provided here, with technicalities and proofs deferred to Appendix 4.6.

4.3.1 Construction for the Vertices

This section provides a construction for the random field $\boldsymbol{Z}_V$ as per the identity (4.18). While the search for metrics is probably unlimited in terms of alternatives, our spectrum is suitably restricted by providing two *desiderata*, denoted $\mathbb{D}$ throughout.

$\mathbb{D}1$**:** The construction (4.18) needs to provide a pseudo-variogram $\mathbf{\Gamma}_{\boldsymbol{Z}}$ for which the diagonal entries should obey to the variogram construction as in Anderes et al. (2020).

$\mathbb{D}2$**:** The multivariate random field $\boldsymbol{Z}_V$ on V should enjoy a homogeneous and intuitive conditional independence structure. More precisely, we set the block-precision matrix $\mathbf{\Theta}$ of $\boldsymbol{Z}_V$ to have the following structure:

$$\mathbf{\Theta} = \mathcal{M}_p \begin{pmatrix} \mathbf{Q} \\ -\alpha \mathbf{I}_n \end{pmatrix}, \tag{4.19}$$

where α is a positive real value and $\mathbf{Q}$ is a suitable $n \times n$ matrix that will be defined next. This means that, for $i \neq j$ and $v_1 \neq v_2$, $Z_{V,i}(v_1)$ and $Z_{V,j}(v_2)$ are conditionally independent given everything else.

Define the $n \times n$ matrices:

$$\mathbf{Q} := \frac{1}{2}\left(\mathbf{L} + \alpha(p-2)\mathbf{I}_n + \sqrt{\mathbf{L}^2 - 2\alpha(p-2)\mathbf{L} + \alpha^2 p^2 \mathbf{I}_n}\right) \tag{4.20}$$

$$= \frac{1}{2}\mathbf{W}\left(\mathbf{\Lambda}^- + \alpha(p-2)\mathbf{I}_n + \sqrt{(\mathbf{\Lambda}^-)^2 - 2\alpha(p-2)\mathbf{\Lambda}^- + \alpha^2 p^2 \mathbf{I}_n}\right)\mathbf{W}^\top, \tag{4.21}$$

$$\mathbf{X} := \frac{1}{\alpha(p-1)}\left(\mathbf{\Sigma}\mathbf{Q} - \mathbf{I}_n\right) \tag{4.22}$$

$$= \frac{\mathbf{W}\left(\alpha(p-2)\mathbf{\Lambda} + \sqrt{\mathbf{I}_n - \mathbf{J}_n - 2\alpha(p-2)\mathbf{\Lambda} + \alpha^2 p^2 \mathbf{\Lambda}^2} - \mathbf{I}_n - \mathbf{J}_n\right)\mathbf{W}^\top}{2\alpha(p-1)}, \tag{4.23}$$

where:

- $\mathbf{I}_n$ is the $n \times n$ identity matrix,
- $\mathbf{J}_n := \operatorname{diag}\left[0, 0, \ldots, 0, 1\right]^\top \in \mathbb{R}^{n \times n}$.

Notice that in Equations (4.21) and (4.23) the principal square root coincides with the element-wise square root, being its arguments diagonal matrices. These equations have been established from Equations (4.20) and (4.22) by using the fact that $\mathbf{W}$ is an orthogonal matrix and that $\mathbf{\Lambda}\mathbf{\Lambda}^- = \mathbf{I}_n - \mathbf{J}_n$ insofar as the first $p-1$ diagonal entries of $\mathbf{\Lambda}$ are positive and the last diagonal entry is zero. In addition, we define the two positive constants

$$k_1 := \frac{p-1}{\alpha n p^2} \qquad k_2 := \frac{p^2 - p + 1}{\alpha n p^2 (p-1)},$$

and the $n \times n$ matrices

$$\tilde{\mathbf{\Sigma}} := \mathbf{\Sigma} + k_1 \mathbf{1}_{n \times n} \qquad \tilde{\mathbf{X}} := \mathbf{X} + k_2 \mathbf{1}_{n \times n}.$$

Although these definitions may appear peculiar, they are the only solution to the *desiderata* $\mathbb{D}$ previously stated and are crucial to define the random field $\boldsymbol{Z}_V$. We start with the vertices, where $\boldsymbol{Z}_V$ is assumed a zero-mean p-variate Gaussian random vector $\boldsymbol{Z}_V\big|_V : V \to \mathbb{R}^p$ having covariance matrix-valued function

$$\mathbb{C}\mathrm{ov}\left(\boldsymbol{Z}_V(v_1), \boldsymbol{Z}_V(v_2)\right) := \mathcal{M}_p\begin{pmatrix}\tilde{\mathbf{\Sigma}}[v_1, v_2] \\ \tilde{\mathbf{X}}[v_1, v_2]\end{pmatrix} \in \mathbb{R}^{p \times p}. \tag{4.24}$$

Checking the positive semi-definiteness of (4.24) relies on a result of independent interest which implies that both $\mathbf{\Theta}$ and $\mathbf{\Theta}^-$ are positive semidefinite matrices. We state this formally below.

Proposition 4.3. *The matrix $\mathbf{\Theta}$ defined in Equation (4.19) and its Moore-Penrose inverse $\mathbf{\Theta}^-$ are quasi-Laplacian.*

Proof of Proposition 4.3. We first show that $\boldsymbol{\Theta}$ is quasi-Laplacian. To this end, we need to show that $\boldsymbol{\Theta}$ is symmetric, positive semidefinite, $\boldsymbol{\Theta}\mathbf{1}_{np} = \mathbf{0}$, and $\boldsymbol{\Theta}$ has only one null eigenvalue.

1. Since $\mathbf{Q}$ and $\alpha\mathbf{I}_n$ are symmetric and $\boldsymbol{\Theta} := \mathcal{M}_p(\mathbf{Q}, -\alpha\mathbf{I}_n)$, $\boldsymbol{\Theta}$ is symmetric.

2. First recall that $\boldsymbol{\Theta} = \mathbf{I}_p \otimes (\mathbf{Q} + \alpha\mathbf{I}_n) + \mathbf{1}_{p\times p} \otimes (-\alpha\mathbf{I}_n)$: we aim to apply Lemma 4.5 to the matrix $\boldsymbol{\Theta}$ with $\mathbf{X} := \mathbf{I}_p \otimes (\mathbf{Q} + \alpha\mathbf{I}_n)$ and $\mathbf{Y} := \mathbf{1}_{p\times p} \otimes (-\alpha\mathbf{I}_n)$. In the following, we write $\boldsymbol{\lambda}(\mathbf{A})$ to denote the vector of eigenvalues of $\mathbf{A}$. In addition, the inequalities apply to all elements: we write $\boldsymbol{\lambda}(\mathbf{A}) \geq c$ as a shortcut of $\forall i \in \{1, \dots, n\}$, $\lambda_i(\mathbf{A}) \geq c$, for an $n \times n$ matrix $\mathbf{A}$.

$$\begin{aligned}\boldsymbol{\lambda}(\mathbf{X}) &= \boldsymbol{\lambda}(\mathbf{I}_p \otimes (\mathbf{Q} + \alpha\mathbf{I}_n)) = \boldsymbol{\lambda}(\mathbf{I}_p) \otimes \boldsymbol{\lambda}(\mathbf{Q} + \alpha\mathbf{I}_n) = \mathbf{1}_p \otimes (\boldsymbol{\lambda}(\mathbf{Q}) + \alpha) \\ &\geq \lambda_n(\mathbf{Q}) + \alpha = \alpha(p-1) + \alpha = \alpha p \\ \boldsymbol{\lambda}(\mathbf{Y}) &= \boldsymbol{\lambda}(\mathbf{1}_{p\times p} \otimes (-\alpha\mathbf{I}_n)) = \begin{bmatrix} p \\ 0 \\ \vdots \\ 0 \end{bmatrix} \otimes \begin{bmatrix} -\alpha \\ \vdots \\ -\alpha \end{bmatrix} \geq -\alpha p.\end{aligned}$$

   Therefore $\boldsymbol{\Theta}$ is positive semidefinite.

3. To show $\boldsymbol{\Theta}\mathbf{1}_{np} = \mathbf{0}$, given the special structure of $\boldsymbol{\Theta}$, it is sufficient to show that

$$\begin{bmatrix} \mathbf{Q} & -\alpha\mathbf{I}_n & \dots & -\alpha\mathbf{I}_n \end{bmatrix} \mathbf{1}_{np} = \mathbf{0},$$

   i.e. $\mathbf{Q}\mathbf{1}_n = \alpha(p-1)\mathbf{1}_n$, which has already been proved in Lemma 4.3.

4. We now need to show that $\boldsymbol{\Theta}$ has no other null eigenvalues, *i.e.* the only solution to

$$\begin{bmatrix} \mathbf{Q} & -\alpha\mathbf{I}_n & \dots & -\alpha\mathbf{I}_n \\ -\alpha\mathbf{I}_n & \mathbf{Q} & \dots & -\alpha\mathbf{I}_n \\ \vdots & \vdots & \ddots & \vdots \\ -\alpha\mathbf{I}_n & -\alpha\mathbf{I}_n & \dots & \mathbf{Q} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_n \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \tag{4.25}$$

   is (up to a multiplicative constant) $\boldsymbol{v}_1 = \dots = \boldsymbol{v}_n = \mathbf{1}_n$. By reading Equation (4.25) by blocks, we get the following: $\forall i \in \{1, \dots, n\}$

$$\mathbf{Q}\boldsymbol{v}_i - \alpha \sum_{k \neq i} \boldsymbol{v}_k = \mathbf{0}.$$

   Taking now the difference for $i = i$ and $i = j$, we get

$$(\mathbf{Q} + \alpha\mathbf{I}_n)(\boldsymbol{v}_i - \boldsymbol{v}_j) = \mathbf{0}$$

and, since $\mathbf{Q}+\alpha\mathbf{I}_n$ is strictly positive definite directly by Lemma 4.3, we obtain that $\boldsymbol{v}_i = \boldsymbol{v}_j =: \boldsymbol{v}$ for all i, j. The problem thus becomes

$$(\mathbf{Q} - \alpha(p-1)\mathbf{I}_n)\boldsymbol{v} = \mathbf{0}.$$

To conclude it is sufficient to notice that the eigenvalues of $\mathbf{Q} - \alpha(p-1)\mathbf{I}_n$ are $\lambda_i(\mathbf{Q}) - \alpha(p-1)$ and, by Lemma 4.3, all the eigenvalues are not less than $\alpha(p-1)$ and exactly one achieve such a value. As a consequence, the kernel of $\mathbf{Q} - \alpha(p-1)\mathbf{I}_n$ has dimension one and, since we already know that $(\mathbf{Q} - \alpha(p-1)\mathbf{I}_n)\mathbf{1}_n = \mathbf{0}$, the only possibility is $\boldsymbol{v} = \mathbf{1}_n$.

That $\boldsymbol{\Theta}^-$ is also a quasi-Laplacian matrix comes from Lemma 4.6 and the closure of the set of quasi-Laplacian matrices under Moore-Penrose inversion (Proposition 4.1).

□

Proposition 4.4. *The function defined in (4.24) is positive semidefinite.*

Proof of Proposition 4.4. It is sufficient to apply Proposition 4.2 with $X := V$. We show that the $np \times np$ matrix $[\mathbf{K}(v_i, v_j)]_{i,j=1}^{np}$ is positive semidefinite. First, notice that $[\mathbf{K}(v_i, v_j)]_{i,j=1}^{np} = \mathcal{M}_p\left(\tilde{\boldsymbol{\Sigma}}, \tilde{\mathbf{X}}\right) = \tilde{\mathbf{K}}$. Since $\tilde{\mathbf{K}} \in \mathbb{R}^{np\times np}$ is positive semidefinite (see Lemma 4.6 and Proposition 4.3), each matrix $\mathbf{C}$ as defined in Proposition 4.2, being a principal submatrix of $\tilde{\mathbf{K}}$, is positive semidefinite as well. This proves the Proposition. □

Once $\boldsymbol{Z}_V$ is defined over the vertices, the corresponding values over the edges are attained through a sheer linear interpolation. Namely, for $u = (\underline{u}, \overline{u}, \delta(u)) \in G$, we set

$$\boldsymbol{Z}_V(u) := (1 - \delta(u))\boldsymbol{Z}_V(\underline{u}) + \delta(u)\boldsymbol{Z}_V(\overline{u}). \tag{4.26}$$

4.3.2 Construction for the Edges

We define $\boldsymbol{Z}_E$ as a p-variate zero-mean Gaussian random field independent of $\boldsymbol{Z}_V$ whose covariance matrix-valued function is:

$$\mathbb{C}\mathrm{ov}\left(\boldsymbol{Z}_E(u_1), \boldsymbol{Z}_E(u_2)\right) := \mathbb{1}\left(e_1 = e_2\right)\ell(e_1)\left(\delta_1 \wedge \delta_2 - \delta_1\delta_2\right)\mathcal{M}_p\begin{pmatrix}1\\ \beta\end{pmatrix} \in [0, +\infty)^{p\times p}, \tag{4.27}$$

for $e_i := e(u_i)$, $\delta_i := \delta(u_i)$, $i \in \{1, 2\}$ and where $\beta \in [0, 1]$ is a fixed parameter. Notice that $\boldsymbol{Z}_E\big|_{e_1}$ is independent from $\boldsymbol{Z}_E\big|_{e_2}$ whenever $e_1 \neq e_2$, in addition $\ell(e_1)\left(\delta_1 \wedge \delta_2 - \delta_1\delta_2\right)$ is the covariance function of a standard Brownian bridge on $[0, \ell(e_1)]$, whilst $\mathcal{M}_p\left(1, \beta\right)$ is positive semidefinite for $\beta \in [0, 1]$. This ensures that (4.27) is positive semidefinite, as the element-wise product of positive semidefinite functions is itself positive semidefinite.

4.3.3 Compendium

The following result comes directly from the above definitions.

Proposition 4.5. *Let $\mathbf{K}_{\boldsymbol{Z}_V}$, $\mathbf{K}_{\boldsymbol{Z}_E}$ be the covariance mappings associated respectively with the random fields $\boldsymbol{Z}_V$ and $\boldsymbol{Z}_E$. Then, it is true that*

$$\mathbf{K}_{\boldsymbol{Z}_V}(u_1, u_2) = \mathcal{M}_p \begin{pmatrix} \boldsymbol{\delta}_1^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{\delta}_2 \\ \boldsymbol{\delta}_1^\top \tilde{\mathbf{X}} \boldsymbol{\delta}_2 \end{pmatrix} \in \mathbb{R}^{p \times p}, \tag{4.28}$$

for $u_i = (\underline{u}_i, \overline{u}_i, \delta_i) \in G$ ($i \in \{1, 2\}$), and where $\boldsymbol{\delta}_i$ is the n-dimensional vector whose entries are

$$\delta_i := \begin{cases} 1 - \delta_i & \text{if } v = \underline{u}_i \\ \delta_i & \text{if } v = \overline{u}_i \\ 0 & \text{otherwise} \end{cases}, \quad v \in V. \tag{4.29}$$

Further, it is true that

$$\mathbf{K}_{\boldsymbol{Z}_E}(u_1, u_2) = \mathcal{M}_p \begin{pmatrix} \mathbb{1}(e_1 = e_2)\, \ell(e_1)\, (\delta_1 \wedge \delta_2 - \delta_1 \delta_2) \\ \beta \mathbb{1}(e_1 = e_2)\, \ell(e_1)\, (\delta_1 \wedge \delta_2 - \delta_1 \delta_2) \end{pmatrix}.$$

As a direct implication of Proposition 4.5 in concert with the fact that $\boldsymbol{Z}_V$ and $\boldsymbol{Z}_E$ are independent, we obtain the following.

Corollary 4.1. *It is true that*

$$\begin{aligned} \mathbf{K}_{\boldsymbol{Z}}(u_1, u_2) &= \mathbf{K}_{\boldsymbol{Z}_V}(u_1, u_2) + \mathbf{K}_{\boldsymbol{Z}_E}(u_1, u_2) \\ &= \mathcal{M}_p \begin{pmatrix} \boldsymbol{\delta}_1^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{\delta}_2 + \mathbb{1}(e_1 = e_2)\, \ell(e_1)\, (\delta_1 \wedge \delta_2 - \delta_1 \delta_2) \\ \boldsymbol{\delta}_1^\top \tilde{\mathbf{X}} \boldsymbol{\delta}_2 + \beta \mathbb{1}(e_1 = e_2)\, \ell(e_1)\, (\delta_1 \wedge \delta_2 - \delta_1 \delta_2) \end{pmatrix}. \end{aligned} \tag{4.30}$$

We are now able to derive the matrix-valued metric associated with the random field $\boldsymbol{Z}$ in (4.18). To do so, we let $\mathbf{D} : G \times G \to [0, +\infty)^{p \times p}$ be the pseudo-variogram of $\boldsymbol{Z}$, *i.e.*

$$\mathbf{D}(u_1, u_2) := \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_1, u_2) = \left[\mathbb{V}\mathrm{ar}\left(Z_i(u_1) - Z_j(u_2)\right)\right]_{i,j=1}^{p}. \tag{4.31}$$

A straightforward application of Equation (4.9) in concert with tedious calculations provides the following fact.

Proposition 4.6. *It is true that*

$$\begin{aligned} \mathbf{D}(u_1, u_2) &= \mathcal{M}_p \begin{pmatrix} (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \tilde{\boldsymbol{\Sigma}} (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) \\ \boldsymbol{\delta}_1^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{\delta}_2 - 2 \boldsymbol{\delta}_1^\top \tilde{\mathbf{X}} \boldsymbol{\delta}_2 \end{pmatrix} \\ &+ \mathcal{M}_p \begin{pmatrix} \ell(e_1)\delta_1(1 - \delta_1) + \ell(e_2)\delta_2(1 - \delta_2) - 2\mathbb{1}(e_1 = e_2)\, \ell(e_1)\, (\delta_1 \wedge \delta_2 - \delta_1 \delta_2) \\ \ell(e_1)\delta_1(1 - \delta_1) + \ell(e_2)\delta_2(1 - \delta_2) - 2\beta\mathbb{1}(e_1 = e_2)\, \ell(e_1)\, (\delta_1 \wedge \delta_2 - \delta_1 \delta_2) \end{pmatrix}, \end{aligned} \tag{4.32}$$

for $u_1, u_2 \in G$, with notation as in Proposition 4.5.

Notice that the former line coincides with the contribution of the random field $\boldsymbol{Z}_V$, whilst the latter is the contribution of $\boldsymbol{Z}_E$.

We describe below some properties of the metric $\mathbf{D}$ provided through our construction.

Proposition 4.7. *The matrix-valued expression (4.31) satisfies properties similar to the (real-valued) quasi-metrics, i.e. it satisfies:*

1. $D_{ij}(u_1, u_2) \geq 0$;
2. $D_{ij}(u_1, u_2) = 0 \iff u_1 = u_2$ *and* $i = j$;
3. $\mathbf{D}(u_1, u_2) = \mathbf{D}(u_1, u_2)^\top = \mathbf{D}(u_2, u_1) = \mathbf{D}(u_2, u_1)^\top$ *for any* $u_1, u_2 \in G$;
4. *The mapping* $(u_1, u_2) \mapsto \mathbf{D}(u_1, u_0)\mathbf{A} + \mathbf{A}^\top \mathbf{D}(u_0, u_2) - \mathbf{D}(u_1, u_2) - \mathbf{A}^\top \mathbf{D}(u_0, u_0)\mathbf{A}$ *is positive semidefinite for any* $u_0 \in G$ *and any* $p \times p$ *matrix* $\mathbf{A}$ *such that* $\mathbf{A}^\top \mathbf{1}_p = \mathbf{1}_p$;
5. *The matrices* $\mathbf{D}(u_1, u_2) - \mathbf{D}(u_0, u_0)$ *and* $2\mathbf{D}(u_1, u_0) + 2\mathbf{D}(u_2, u_0) - \mathbf{D}(u_1, u_2) - 3\mathbf{D}(u_0, u_0)$ *are symmetric and positive semidefinite for any* $u_0, u_1, u_2 \in G$.

Note the resemblance between the fourth property in Proposition 4.7 and the second statement in Zastavnyi (2023, Proposition 4).

Proposition 4.8. *Let* G *be a graph with Euclidean edges and let* $\mathbf{D} : G \times G \to [0, +\infty)^{p \times p}$ *as defined in (4.31) and having elements* D_{ij}, $i, j = 1, \ldots, p$. *Then, the following properties hold:*

1. *for all* $i \in \{1, \ldots, p\}$, *the distance* D_{ii} *coincides with the resistance distance defined by Anderes et al. (2020), i.e.* $D_{ii}(u_1, u_2) = d_R(u_1, u_2)$;
2. *the distance* $\mathbf{D}$ *is matrix-homogeneous, i.e. for every* $i \neq j$ *and* $i' \neq j'$, *we have* $D_{ij}(u_1, u_2) = D_{i'j'}(u_1, u_2)$ *and* $D_{ii}(u_1, u_2) = D_{i'i'}(u_1, u_2)$.

Proof of Proposition 4.8. 1. We split the computation of the distances (the one proposed here, $\mathbf{D}$, and the one defined by Anderes et al. (2020), d_R) in two: the contribution (*i.e.* the variogram) of the $\boldsymbol{Z}_V$ process and the one of $\boldsymbol{Z}_E$, denoted via the superscripts V and E respectively. We show that $\mathbf{D}^V_{ii}(u_1, u_2) = d^V_R(u_1, u_2)$ and $\mathbf{D}^E_{ii}(u_1, u_2) = d^E_R(u_1, u_2)$. The latter comes straightforwardly by comparing the diagonal entries of Equation (4.27), *i.e.* $\mathbb{1}(e_1 = e_2)\,\ell(e_1)\,(\delta_1 \wedge \delta_2 - \delta_1\delta_2)$, with the kernel R_e in Anderes et al. (2020, Equation 15).

Thus, let us show that $\mathbf{D}^V_{ii}(u_1, u_2) = d^V_R(u_1, u_2)$. Let $u_1, u_2 \in G$ and $k_{V,ii}$ be the i-th diagonal entry of $\mathbf{K}_{\boldsymbol{Z}_V}$. Then, on account of Proposition 4.5:

$$\begin{aligned}\mathbf{D}^V_{ii}(u_1, u_2) &= k_{V,ii}(u_1, u_1) + k_{V,ii}(u_2, u_2) - 2k_{V,ii}(u_1, u_2) \\ &= \boldsymbol{\delta}_1^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{\delta}_2 - 2\boldsymbol{\delta}_1^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{\delta}_2\end{aligned}$$

$$
\begin{aligned}
&= (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top (\boldsymbol{\Sigma} + k_1 \mathbf{1}_{n\times n}) (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) \\
&= (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \boldsymbol{\Sigma} (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) + k_1 (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \mathbf{1}_{n\times n} (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) \\
&= (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \boldsymbol{\Sigma} (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2),
\end{aligned}
$$

where in the last step we have used $\boldsymbol{\delta}_1^\top \mathbf{1}_n = \boldsymbol{\delta}_2^\top \mathbf{1}_n = 1$ owing to Equation (4.29).

It is left to show that this quantity coincides with the resistance counterpart of Anderes et al. (2020). They define their covariance matrix, here denoted by $\boldsymbol{\Sigma}_\mathbf{A}$, as the (proper) inverse of a modified Laplacian matrix $\mathbf{L}$, where a one has been added on an arbitrary diagonal entry. In formulae: $\boldsymbol{\Sigma}_\mathbf{A} := (\mathbf{L} + \boldsymbol{e}_j \boldsymbol{e}_j^\top)^{-1}$, where $\boldsymbol{e}_j$ is the j^{th} vector of the canonical basis of $\mathbb{R}^n$. It follows directly from Anderes et al. (2020, Equation 13) that $d_R^V = (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \boldsymbol{\Sigma}_\mathbf{A} (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)$.

Let us show that, for a given quasi-Laplacian matrix $\mathbf{L}$ and for any $\boldsymbol{x} \in \mathbb{R}^n$ such that ${\mathbf{1}_n}^\top \boldsymbol{x} \neq 0$, it holds:

$$
(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \left(\mathbf{L} + \boldsymbol{x}\,\boldsymbol{x}^\top\right)^- (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) = (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \mathbf{L}^- (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2). \tag{4.33}
$$

In order to express the Moore-Penrose inverse of the rank-1 update $\mathbf{L} + \boldsymbol{x}\,\boldsymbol{x}^\top$, we exploit Theorem 1 of Meyer (1973) with $\mathbf{A} := \mathbf{L}$ and $\boldsymbol{c} := \boldsymbol{d} := \boldsymbol{x}$. Firstly, let us show that the hypotheses of the theorem are satisfied: we have to prove that $\boldsymbol{x} \notin R(\mathbf{L})$, where R denotes the range or column space. $R(\mathbf{L})$ is perpendicular to the kernel of $\mathbf{L}$, that is (recall that $\mathbf{L}$ is a quasi-Laplacian matrix) $\ker(\mathbf{L}) = \{h\,\mathbf{1}_n : h \in \mathbb{R}\}$. Now, $\boldsymbol{x} \in R(\mathbf{L}) \iff \boldsymbol{x} \perp \ker(\mathbf{L}) \iff {\mathbf{1}_n}^\top \boldsymbol{x} = 0$, yet ${\mathbf{1}_n}^\top \boldsymbol{x} \neq 0$ by hypothesis. Therefore $\boldsymbol{x} \notin R(\mathbf{L})$, *i.e.* we can apply the above-mentioned theorem. In the following computations, we have adopted the notation used by Meyer (1973, Section 2), *i.e.* :

$$
\begin{aligned}
&\boldsymbol{k} := \mathbf{L}^- \boldsymbol{x} && \boldsymbol{h} := \boldsymbol{x}^\top \mathbf{L}^- \\
&\boldsymbol{u} := \left(\mathbf{I}_n - \mathbf{L}\mathbf{L}^-\right) \boldsymbol{x} = \frac{1}{n} \mathbf{1}_{n\times n} \boldsymbol{x} && \boldsymbol{v} := \boldsymbol{x}^\top \left(\mathbf{I}_n - \mathbf{L}^-\mathbf{L}\right) = \frac{1}{n} \boldsymbol{x}^\top \mathbf{1}_{n\times n} \\
&\beta := 1 + \boldsymbol{x}^\top \mathbf{L}^- \boldsymbol{x}.
\end{aligned}
$$

In addition, we have $\|\boldsymbol{u}\|^2 = \|\boldsymbol{v}\|^2 = \boldsymbol{u}^\top \boldsymbol{u} = \frac{1}{n^2} {\mathbf{1}_n}^\top ({\mathbf{1}_n}^\top \boldsymbol{x}) ({\mathbf{1}_n}^\top \boldsymbol{x}) \mathbf{1}_n = \frac{1}{n} ({\mathbf{1}_n}^\top \boldsymbol{x})^2 \neq 0$. Let us finally apply Equation (3.1) of Meyer (1973):

$$
\begin{aligned}
\left(\mathbf{L} + \boldsymbol{x}\boldsymbol{x}^\top\right)^- &= \mathbf{L}^- - \boldsymbol{k}\boldsymbol{u}^- - \boldsymbol{v}^- \boldsymbol{h} + \beta \boldsymbol{v}^- \boldsymbol{u}^- \\
&= \mathbf{L}^- - \frac{\boldsymbol{k}\boldsymbol{u}^\top}{\|\boldsymbol{u}\|^2} - \frac{\boldsymbol{v}^\top \boldsymbol{h}}{\|\boldsymbol{v}\|^2} + \beta \frac{\boldsymbol{v}^\top \boldsymbol{u}^\top}{\|\boldsymbol{v}\|^2 \|\boldsymbol{u}\|^2} \\
&= \mathbf{L}^- - \frac{\mathbf{L}^- \boldsymbol{x} {\mathbf{1}_n}^\top}{{\mathbf{1}_n}^\top \boldsymbol{x}} - \frac{\mathbf{1}_n \boldsymbol{x}^\top \mathbf{L}^-}{{\mathbf{1}_n}^\top \boldsymbol{x}} + \left(1 + \boldsymbol{x}^\top \mathbf{L}^- \boldsymbol{x}\right) \frac{\mathbf{1}_{n\times n}}{\left({\mathbf{1}_n}^\top \boldsymbol{x}\right)^2}.
\end{aligned}
$$

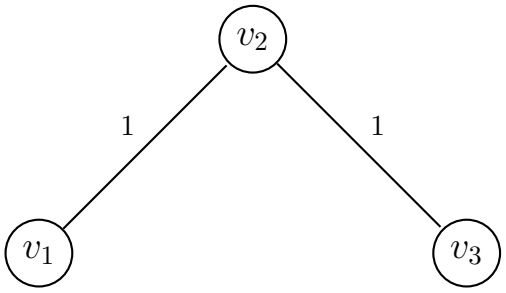


Figure 4.3: An example of graph for which the distance $D_{11}(v_1, v_3) \not\leq D_{12}(v_1, v_3)$, for $p \geq 3$ and α wisely chosen.

To show (4.33), it is now sufficient to prove that

$$(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \left(\left(1 + \boldsymbol{x}^\top \mathbf{L}^- \boldsymbol{x}\right) \frac{\mathbf{1}_{n\times n}}{\left(\mathbf{1}_n{}^\top \boldsymbol{x}\right)^2} - \frac{\mathbf{L}^- \boldsymbol{x} \mathbf{1}_n{}^\top}{\mathbf{1}_n{}^\top \boldsymbol{x}} - \frac{\mathbf{1}_n \boldsymbol{x}^\top \mathbf{L}^-}{\mathbf{1}_n{}^\top \boldsymbol{x}} \right) (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) = 0.$$

It is possible to see this by noticing that $\mathbf{1}_n{}^\top (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) = \mathbf{1}_n{}^\top \boldsymbol{\delta}_1 - \mathbf{1}_n{}^\top \boldsymbol{\delta}_2 = 1 - 1 = 0$ and similarly $(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \mathbf{1}_n = 0$, therefore each term of the above equation equals 0.

As a consequence,

$$\begin{aligned} d_R^V(u_1, u_2) &= (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \boldsymbol{\Sigma}_{\mathbf{A}} (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) \\ &= (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \left(\mathbf{L} + \boldsymbol{e}_j \boldsymbol{e}_j^\top \right)^- (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) \\ &= (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \mathbf{L}^- (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) = D_{ii}^V(u_1, u_2). \end{aligned}$$

2. The proof of this point is straightforward from Equation (4.32): notice that $D_{ij}(u_1, u_2)$, for $i \neq j$, is the off-diagonal element of $\mathbf{D}(u_1, u_2)$, whilst $D_{ii}(u_1, u_2)$ is the main diagonal one.

□

Albeit counter-intuitive, the distance within the same variable is not always less than the same distance among variables. In formulae, $D_{ii}(u_1, u_2) \not\leq D_{ij}(u_1, u_2)$, in general. The following example shows this fact.

Example 4.1. Let $p \geq 3$ and $\alpha > 0$. Consider the graph depicted in Figure 4.3 with $V := \{v_1, v_2, v_3\}$ and whose Laplacian matrix $\mathbf{L}$ and $\boldsymbol{\Sigma} = \mathbf{L}^-$ are

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \frac{1}{9} \begin{bmatrix} 5 & -1 & -4 \\ -1 & 2 & -1 \\ -4 & -1 & 5 \end{bmatrix}.$$

In addition, the eigendecomposition $\mathbf{\Sigma} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^\top$ is

$$\mathbf{W} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}, \qquad \mathbf{\Lambda} = \begin{bmatrix} 1 & & \\ & \frac{1}{3} & \\ & & 0 \end{bmatrix}.$$

Let us consider the distance between v_1 and v_3 for the same variable and for different variables, *i.e.* $D_{11}(v_1, v_3)$ and $D_{12}(v_1, v_3)$. Since both v_1 and v_3 are vertices, the contribution of $\boldsymbol{Z}_E$ is null, therefore, by Equation (4.32), we have:

$$\begin{aligned} D_{11}(v_1, v_3) &= (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_3)^\top \tilde{\mathbf{\Sigma}} (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_3) \\ &= [1, 0, -1] \tilde{\mathbf{\Sigma}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \\ &= \tilde{\Sigma}_{11} + \tilde{\Sigma}_{33} - 2\tilde{\Sigma}_{13} \\ &= \Sigma_{11} + \Sigma_{33} - 2\Sigma_{13} = 2, \\ D_{12}(v_1, v_3) &= \boldsymbol{\delta}_1^\top \tilde{\mathbf{\Sigma}} \boldsymbol{\delta}_1 + \boldsymbol{\delta}_3^\top \tilde{\mathbf{\Sigma}} \boldsymbol{\delta}_3 - 2\boldsymbol{\delta}_1^\top \tilde{\mathbf{X}} \boldsymbol{\delta}_3 \\ &= \tilde{\Sigma}_{11} + \tilde{\Sigma}_{33} - 2\tilde{X}_{13} \\ &= \Sigma_{11} + \Sigma_{33} + 2k_1 - 2X_{13} - 2k_2 \\ &= \frac{10}{9} - 2 \cdot \frac{1}{3\alpha p(p-1)} - 2X_{13}. \end{aligned}$$

Now it is possible to compute X_{13} by means of Equation (4.22):

$$\begin{aligned} X_{13} &= \frac{1}{2\alpha(p-1)} W_{1\star} \left(\alpha(p-2)\mathbf{\Lambda} + \sqrt{\mathbf{I}_n - \mathbf{J}_n - 2\alpha(p-2)\mathbf{\Lambda} + \alpha^2 p^2 \mathbf{\Lambda}^2} - \mathbf{I}_n - \mathbf{J}_n \right) W_{\star 3}^\top \\ &= \frac{1}{2\alpha(p-1)} \sum_{i=1}^{3} W_{1i} W_{3i} \left(\alpha(p-2)\mathbf{\Lambda} + \sqrt{\mathbf{I}_n - \mathbf{J}_n - 2\alpha(p-2)\mathbf{\Lambda} + \alpha^2 p^2 \mathbf{\Lambda}^2} - \mathbf{I}_n - \mathbf{J}_n \right)_{ii} \\ &= \frac{1}{2\alpha(p-1)} \Bigg(-\frac{1}{2} \left(\alpha(p-2) \cdot 1 + \sqrt{1 - 2\alpha(p-2) \cdot 1 + \alpha^2 p^2 \cdot 1^2} - 1 \right) \\ &\quad + \frac{1}{6} \left(\alpha(p-2) \cdot \frac{1}{3} + \sqrt{1 - 2\alpha(p-2) \cdot \frac{1}{3} + \alpha^2 p^2 \cdot \frac{1}{3^2}} - 1 \right) \\ &\quad + \frac{1}{3} \left(\alpha(p-2) \cdot 0 + \sqrt{0 - 2\alpha(p-2) \cdot 0 + \alpha^2 p^2 \cdot 0^2} - 2 \right) \Bigg) \\ &= \frac{1}{2\alpha(p-1)} \Bigg(-\frac{1}{2} \sqrt{1 - 2\alpha(p-2) + \alpha^2 p^2} \\ &\quad + \frac{1}{6} \sqrt{1 - \frac{2}{3}\alpha(p-2) + \frac{\alpha^2 p^2}{9}} - \frac{4}{9}\alpha(p-2) - \frac{1}{3} \Bigg). \end{aligned}$$

p	α	$\mathbf{Q}$	$\mathbf{X}$	$D_{ij}(u_1,u_2)$ for $i \neq j$
$+\infty$	fixed	$\alpha(p-1)\mathbf{I}_n$	$\boldsymbol{\Sigma}$	$d_R(u_1,u_2)$
fixed	$+\infty$	$\alpha(p-1)\mathbf{I}_n$	$\boldsymbol{\Sigma}$	$d_R(u_1,u_2)$
fixed	0^+	$\mathbf{L}$	$-\frac{1}{\alpha n(p-1)}\mathbf{1}_{n\times n}$	$\frac{2}{\alpha n p}$
$+\infty$	$\frac{1}{p\tau}$	$\frac{\mathbf{I}_n}{\tau}$	$\boldsymbol{\Sigma}-\tau\mathbf{I}_n$	$d_R(u_1,u_2)+2\tau\boldsymbol{\delta}_1^\top\boldsymbol{\delta}_2$
$+\infty$	$\frac{1}{p\tau}$	$\mathbf{L}+\frac{1}{n\tau}\mathbf{1}_{n\times n}$	$-\frac{\tau}{n}\mathbf{1}_{n\times n}$	$d_R(u_1,u_2)+\frac{2\tau}{n}+2\boldsymbol{\delta}_1^\top\mathbf{X}\boldsymbol{\delta}_2$

Table 4.1: Asymptotic results with respect to p and α. In the 4^{th} row, $0 < \tau \leq \lambda_{n-1}$, with λ_{n-1} being the smallest positive eigenvalue of $\boldsymbol{\Sigma}$, while in the last row, $\tau \geq \lambda_1$. In the last column, the results involving $d_R(u_1,u_2)$ hold only for $\beta \to 1$ (the other result holds for every $\beta \in [0,1]$)

With $p = 3$, the maximum discrepancy $d(\alpha) := D_{11}(v_1,v_3) - D_{12}(v_1,v_3)$ is reached at $\alpha^\star = 1.32092$. In such a case, $X_{13} = -0.48598$ and thus:

$$D_{12}(v_1,v_3) = \frac{10}{9} - 2\cdot\frac{1}{18\cdot 1.32092} - 2\cdot(-0.48598) = 1.99896 < 2 = D_{11}(v_1,v_3),$$

that is, $d(\alpha^\star) = 0.00103$. Notice that for $\alpha := \hat{\alpha} = 0.97222$ and for $\alpha \to +\infty$, we achieve $d(\hat{\alpha}) = 0$, and for $0 < \alpha < \hat{\alpha}$ we have $d(\alpha) < 0$. One can obtain larger values for $d(\alpha)$ by taking $\alpha := \frac{1}{p}$ and letting $p \to +\infty$. In this setting, we have $X_{13} \to -\frac{1}{3}$ and $d(\alpha) \to \frac{2}{9}$.

The following result illustrates some asymptotic properties of the proposed metric.

Proposition 4.9. *Let G be a graph with Euclidean edges and let $\mathbf{Q}$ and $\mathbf{X}$ as defined in Equations (4.20)-(4.23). Then, the asymptotic results indicated in Table 4.1 hold.*

Theorem 4.2. *Let G be a graph with Euclidean edges and $\mathbf{D} : G\times G \to [0,+\infty)^{p\times p}$ the distance defined at (4.31). In addition, let $\psi : [0,+\infty) \to (0,+\infty)$ be a non-constant completely monotone function and $g : [0,+\infty) \to [0,+\infty)$ a Bernstein function. Then, for every $\xi > 0$, the mapping $\mathbf{K}$ defined as*

$$(u_1,u_2) \mapsto \mathbf{K}(u_1,u_2) = \psi\left(\xi\, g\left(\mathbf{D}(u_1,u_2)\right)\right), \tag{4.34}$$

where ψ and g are applied element-wise, is a valid covariance function. If, in addition, $\psi(0) = 1$, then (4.34) is a valid correlation function.

One of the main drawbacks of this construction is that the distance is homogeneous (see Proposition 4.8). Consequently, not only do the resulting matrix-valued covariance function (4.34) assigns the same marginal covariance for all the variables, *i.e.* $\psi(\xi g(D_{ii}(u_1,u_2))) = \psi(\xi g(d_R(u_1,u_2)))$ for any $u_1,u_2 \in G$ and any $i \in \{1,\ldots,p\}$, but also the cross-covariance between two different variables is independent from the

variables themselves, *i.e.* $\psi(\xi\, g(D_{ij}(u_1,u_2)))$ does not depend on i,j as long as $i \neq j$. Such a restriction can be easily circumvented. Let $[\rho_{ij}]_{i,j=1}^{p}$ be a collocated correlation matrix, *e.g.*, $\rho_{ii}=1$ and $-1 \leq \rho_{ij} \leq 1$ when $i \neq j$, and the matrix being symmetric and positive semidefinite. Let $\boldsymbol{\sigma} = (\sigma_1,\ldots,\sigma_p)^\top$ with $\sigma_i > 0$ for $i=1,\ldots,p$. Let $\mathbf{K}$ be the mapping in (4.34) and define the mapping $\widetilde{\mathbf{K}} : G \times G \to \mathbb{R}^{p\times p}$ with entries

$$\widetilde{K}_{ij}(u_1,u_2) := \sigma_i\sigma_j\rho_{ij}K_{ij}(u_1,u_2), \quad i,j=1,\ldots,p, \quad u_1,u_2 \in G.$$

Then, by straight application of the Schur product theorem, one gets that $\widetilde{\mathbf{K}}$ is a positive semidefinite matrix-valued function. However, all the marginal covariances (diagonal entries of $\widetilde{\mathbf{K}}$) are proportional, and the same happens for all the cross-covariances (off-diagonal entries of $\widetilde{\mathbf{K}}$), which is still restrictive. The results following throughout allow to circumvent such a limitation.

4.3.4 Generalisations

An interesting generalisation is provided below.

Theorem 4.3. *Let $\theta \in (0,1]$. Let $\mathbf{F} : [0,+\infty) \to \mathbb{R}^{p\times p}$ be a Borel measure such that, for any fixed $\xi \geq 0$, the matrix $\mathbf{F}(\mathrm{d}\xi)$ is symmetric and positive semidefinite. Let $\mathbf{D}$ as in Equation (4.31) and let $\boldsymbol{\Psi}(\cdot) : [0,+\infty) \to \mathbb{R}^{p\times p}$ be a continuous mapping determined through*

$$\boldsymbol{\Psi}(\mathbf{D}) := \Big[\int_0^{+\infty} \exp(-\xi D_{ij}^{\theta})\, F_{ij}(\mathrm{d}\xi)\Big]_{i,j=1}^{p}, \tag{4.35}$$

where $F_{ij}(\mathrm{d}\cdot)$ and D_{ij} are the (i,j)-th entries of $\mathbf{F}(\mathrm{d}\cdot)$ and $\mathbf{D}$, respectively. Then, $\boldsymbol{\Psi}$ is the covariance function of a Gaussian random field $\boldsymbol{Z}$ constructed according to the addition principle (4.18).

As a first application of this result, we adapt a construction that has been originally provided by Porcu et al. (2018) in Euclidean spaces and working with the classical Euclidean norm. The Shkarofsky-Gneiting family of functions $\mathcal{SG}_{\alpha,\beta,\nu}(\cdot)$ is defined on $[0,+\infty)$ through

$$\mathcal{SG}_{\alpha,\beta,\nu}(t) = \left(1+\frac{t}{\beta}\right)^{-\nu/2} \frac{\mathcal{K}_\nu\left(\frac{1}{\sqrt{\alpha}}\sqrt{\beta+t}\right)}{\mathcal{K}_\nu\left(\sqrt{\frac{\beta}{\alpha}}\right)}, \qquad t \geq 0, \tag{4.36}$$

where $\mathcal{K}_\nu$ stands for the modified Bessel function of the second kind. Arguments therein show that $\mathcal{SG}$ is the Laplace transform of a positive and bounded function. The family is very interesting as it admits the special limit cases

$$\mathcal{SG}_{\alpha,0,-\nu}(t) = \mathcal{M}_{\alpha,\nu}(t) := \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{t}{\alpha}\right)^{\nu/2}\mathcal{K}_\nu\left(\sqrt{\frac{t}{\alpha}}\right), \qquad t \geq 0,$$

for $\alpha, \nu > 0$, and

$$\mathcal{SG}_{0,\beta,\nu}(t) = \mathcal{C}_{\beta,\nu}(t) := \left(1 + \frac{t}{\beta}\right)^{-\nu}, \qquad t \geq 0,$$

for $\beta, \nu > 0$. We consider the case $p = 2$ and define the bivariate Shkarofsky-Gneiting mapping $\boldsymbol{\mathcal{SG}}_2 : [0, +\infty)^{2\times 2} \to \mathbb{R}^{2\times 2}$ through

$$\boldsymbol{\mathcal{SG}}_2(\mathbf{D}) = \begin{bmatrix} \sigma_1^2\, \mathcal{SG}_{\alpha_1,\beta_1,\nu_1}(D_{11}^{\theta}) & \sigma_1\sigma_2\rho\, \mathcal{SG}_{\alpha_{12},\beta_{12},\nu_{12}}(D_{12}^{\theta}) \\ \\ \sigma_1\sigma_2\rho\, \mathcal{SG}_{\alpha_{12},\beta_{12},\nu_{12}}(D_{12}^{\theta}) & \sigma_2^2\, \mathcal{SG}_{\alpha_2,\beta_2,\nu_2}(D_{22}^{\theta}) \end{bmatrix}, \tag{4.37}$$

where $\theta \in (0, 1]$, $(\sigma_1, \sigma_2, \alpha_1, \alpha_2, \alpha_{12}, \beta_1, \beta_2, \beta_{12}) \in (0, +\infty)^8$ and $(\rho, \nu_1, \nu_2, \nu_{12}) \in \mathbb{R}^4$. The following result is a straightforward combination of Theorem 4.3 with the arguments in Theorem 1 of Porcu et al. (2018), so that the proof is omitted.

Theorem 4.4. *Let $\boldsymbol{\mathcal{SG}}_2$ be the matrix-valued function defined by Equation (4.37). If either*

A. PARSIMONIOUS $\boldsymbol{\mathcal{SG}}_2$**:** $\alpha_{12} = \frac{1}{2}\frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2)}$, $\beta_{12} = \frac{1}{2}(\beta_1 + \beta_2)$, $\nu_{12} = \frac{1}{2}(\nu_1 + \nu_2)$, *and*

$$|\rho| \leq \sqrt{\frac{A(\alpha_1, \beta_1, \nu_1)A(\alpha_2, \beta_2, \nu_2)}{A^2(\alpha_{12}, \beta_{12}, \nu_{12})}},$$

with $A(\cdot, \cdot, \cdot)$ defined as

$$A(\alpha, \beta, \nu) = \frac{2^{\nu-1}\,(\alpha\beta)^{\nu/2}}{\mathcal{K}_{\nu}\left(\sqrt{\frac{\beta}{\alpha}}\right)},$$

or

B. FULL $\boldsymbol{\mathcal{SG}}_2$**:** $\alpha_{12} < \frac{1}{2}\frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2)}$, $\beta_{12} > \frac{1}{2}(\beta_1 + \beta_2)$, $\nu_{12} \neq \frac{1}{2}(\nu_1 + \nu_2)$, *and*

$$\begin{aligned} |\rho|^2 \leq\ & \frac{A(\alpha_1, \beta_1, \nu_1)A(\alpha_2, \beta_2, \nu_2)}{A^2(\alpha_{12}, \beta_{12}, \nu_{12})} \\ & \cdot B\left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} - \frac{1}{2\alpha_{12}}, \beta_1 + \beta_2 - 2\beta_{12}, \nu_1 + \nu_2 - 2\nu_{12}\right), \end{aligned}$$

with B defined as

$$B(\alpha, \beta, \nu) = \left(\frac{\left(\nu - \sqrt{\alpha\beta + \nu^2}\right)}{2\beta}\right)^{\nu} \exp\left(\frac{\alpha\beta + \nu(-\nu - \sqrt{\alpha\beta + \nu^2})}{-\nu + \sqrt{\alpha\beta + \nu^2}}\right),$$

then $\boldsymbol{\mathcal{SG}}_2$ is positive semidefinite on any graph with Euclidean edges.

As a second application to a full multivariate setting ($p \geq 2$), consider the mappings $\boldsymbol{\mathcal{M}}_p$, $\boldsymbol{\mathcal{C}}_p$ and $\boldsymbol{\mathcal{SG}}_p$ defined on $[0, +\infty)^{p \times p}$ by

$$\boldsymbol{\mathcal{M}}_p(\mathbf{D}) = \left[\sigma_{ij}\, \mathcal{M}_{\alpha_{ij},\nu_{ij}}\left(D_{ij}^{\theta}\right)\right]_{i,j=1}^{p}, \tag{4.38}$$

$$\boldsymbol{\mathcal{C}}_p(\mathbf{D}) = \left[\sigma_{ij}\, \mathcal{C}_{\alpha_{ij},\nu}\left(D_{ij}^{\theta}\right)\right]_{i,j=1}^{p}, \tag{4.39}$$

$$\boldsymbol{\mathcal{SG}}_p(\mathbf{D}) = \left[\sigma_{ij}\, \mathcal{SG}_{\alpha_{ij},\beta_{ij},\nu}\left(D_{ij}^{\theta}\right)\right]_{i,j=1}^{p}, \tag{4.40}$$

with $[\sigma_{ij}]_{i,j=1}^{p}$, $[\nu_{ij}]_{i,j=1}^{p}$, $[\alpha_{ij}]_{i,j=1}^{p}$ and $[\beta_{ij}]_{i,j=1}^{p}$ being real symmetric matrices, ν a real number and $\theta \in (0, 1]$. Then the following statement holds.

Theorem 4.5. *Let $\boldsymbol{\mathcal{M}}_p$ defined by Equation (4.38). If either*

A. *$[\nu_{ij}]_{i,j=1}^{p}$ and $\left[\nu_{ij}\alpha_{ij}\right]_{i,j=1}^{p}$ are conditionally negative semidefinite and*

$$\left[\frac{\sigma_{ij}\nu_{ij}^{\nu_{ij}}\exp(-\nu_{ij})}{\Gamma(\nu_{ij})}\right]_{i,j=1}^{p}$$

is positive semidefinite

or

B. *there exists $\beta > 0$ such that $[\nu_{ij}]_{i,j=1}^{p}$ and $[\alpha_{ij}^{-1} - \beta\nu_{ij}]_{i,j=1}^{p}$ are conditionally negative semidefinite and $\left[\sigma_{ij}(\beta\alpha_{ij})^{-\nu_{ij}}\exp(-\nu_{ij})\right]_{i,j=1}^{p}$ is positive semidefinite,*

then $\boldsymbol{\mathcal{M}}_p$ is positive semidefinite on any graph with Euclidean edges.

Theorem 4.6. *Let $\boldsymbol{\mathcal{C}}_p$ defined by Equation (4.39) with $\nu > 0$. If $[\beta_{ij}]_{i,j=1}^{p}$ is conditionally negative semidefinite and $\left[\sigma_{ij}\beta_{ij}^{\nu}\right]_{i,j=1}^{p}$ is positive semidefinite, then $\boldsymbol{\mathcal{C}}_p$ is positive semidefinite on any graph with Euclidean edges.*

Theorem 4.7. *Let $\boldsymbol{\mathcal{SG}}_p$ defined by Equation (4.40). If $[\alpha_{ij}^{-1}]_{i,j=1}^{p}$ and $[\beta_{ij}]_{i,j=1}^{p}$ are conditionally negative semidefinite and $\left[\sigma_{ij}(\alpha_{ij}\beta_{ij})^{\nu/2}/\mathcal{K}_{\nu}\left(\sqrt{\frac{\alpha_{ij}}{\beta_{ij}}}\right)\right]_{i,j=1}^{p}$ is positive semidefinite, then $\boldsymbol{\mathcal{SG}}_p$ is positive semidefinite on any graph with Euclidean edges.*

4.4 Euclidean Trees

Euclidean trees with a number m of leaves can be embedded on the m'-dimensional Euclidean space with $m' := \lceil m/2 \rceil$ (Anderes et al., 2020) endowed with the ℓ_1 distance: for a positive integer m' and two points $x, y \in \mathbb{R}^{m'}$, the ℓ_1 or Manhattan-city block distance d_M can be defined as $d_M(x, y) = \sum_{k=1}^{m'} |x_k - y_k|$. According to this embedding, for every $m' \geq 2$, every matrix-valued covariance function depending

on d_M can be used as an isotropic covariance function for a Euclidean tree The surprising fact is that, to the best of our knowledge, the literature has no models of multivariate covariance functions depending on such a metric. There is a fact that is even more surprising. Call Φ_m^p the class of continuous mappings $\mathbf{K} : [0, +\infty) \to \mathbb{R}^{p\times p}$ such that the composition $\mathbf{K}(d_M(\cdot,\cdot))$ is positive semidefinite on a Euclidean tree with m leaves. A characterisation of this class is missing in the literature and we provide it below.

Proposition 4.10. *Let m, p be two positive integers. Let $m' := \lceil m/2 \rceil$. Let $\mathbf{K} : [0, +\infty) \to \mathbb{R}^{p\times p}$ be continuous with $\mathbf{K}(0) = \mathbf{1}_{p\times p}$. Then, $\mathbf{K}$ belongs to the class Φ_m^p if and only if*

$$\mathbf{K}(t) = \int_0^{+\infty} \omega_{m'}(rt)\, \mathbf{F}(\mathrm{d}r), \qquad t \geq 0, \tag{4.41}$$

where $\mathbf{F}$ is a bounded symmetric matrix-valued measure such that $\mathbf{F}(\mathrm{d}r)$ is positive semidefinite for all $r > 0$, and where

$$\omega_{m'}(t) = \frac{\Gamma(m'/2)}{\sqrt{\pi}\Gamma((m'-1)/2)} \int_1^{+\infty} \Omega_{m'}(\sqrt{v}t) v^{-m'/2} (v-1)^{(m'-3)/2} \mathrm{d}v, \tag{4.42}$$

with $\Omega_{m'}$ being the characteristic function of a random vector that is uniformly distributed over the unit sphere $\mathbb{S}^{m'-1}$ embedded in $\mathbb{R}^{m'}$.

Some comments are in order. For $p = 1$, the result has been proved by Cambanis et al. (1983). The following strict inclusion relations hold:

$$\Phi_1^p \supset \Phi_2^p \supset \cdots \supset \bigcap_m \Phi_m^p =: \Phi_\infty^p.$$

A convergence argument from Schoenberg (1942) applies *mutatis mutandis* to assert (no proof needed) that $\mathbf{K} \in \Phi_\infty^p$ if and only if

$$\mathbf{K}(t) = \int_0^{+\infty} \exp(-rt)\mathbf{F}(\mathrm{d}r), \qquad t \geq 0, \tag{4.43}$$

with $\mathbf{F}$ as in Proposition 4.10. Closed-form expressions for the inner kernel $\omega_{m'}$ have been available thanks to Gneiting (1998). In particular, we have

$$\omega_2(t) = -\frac{2}{\pi}\mathrm{si}(t) \quad \text{and} \quad \omega_3(t) = \frac{1}{2}\left(\frac{\sin t}{t} + \cos t + t\,\mathrm{si}(t)\right), \quad t \geq 0,$$

where si denotes the sine integral function.

There exists a wealth of multivariate covariance models that depend on the Euclidean distance, denoted $\|\cdot\|_m$, in $\mathbb{R}^m$. Call Ψ_m^p the class of continuous mappings $\mathbf{H} : [0, +\infty) \to \mathbb{R}^{p\times p}$ with $\mathbf{H}(0) = \mathbf{1}_{p\times p}$ and $\mathbf{H}(\|\cdot\|_m)$ being positive semidefinite in $\mathbb{R}^m$. The following result proves something extremely useful even for the scalar-case, that apparently was overlooked by Anderes et al. (2020).

Proposition 4.11. *Let m, p be positive integers. Let $m' := \lceil m/2 \rceil$. Let $\mathbf{H}$ be a member of the class $\Psi^p_{2m'-1}$. Let the mapping $\mathbf{K}$ be defined through*

$$\mathbf{K}(t) = I^{(m'-1)}\mathbf{H}(t), \qquad t \geq 0, \tag{4.44}$$

where $I^{(m'-1)}$ is the $(m'-1)$ iterated application of the operator I, defined as

$$(If)(t) := \frac{\int_t^{+\infty} f(u)\mathrm{d}u}{\int_0^{+\infty} f(u)\mathrm{d}u}.$$

Then, $\mathbf{K}$ belongs to the class Φ^p_m.

Propositions 4.10 and 4.11 provide a recipe to build new members of the class Φ^p_m. We start by univariate models, as described in Table 4.2, where we use the shortcuts Ψ_m and Φ_m for Ψ^1_m and Φ^1_m, respectively.

Conditions	Member of $\Psi_{2m'-1}$	Member of Φ_m
$\nu \geq m'$	$\psi_\nu(t) = (1-t)^\nu_+$	$\phi_\nu(t) = (1-t)^{\nu+m'-1}_+$
$\nu \geq m'+1$	$\psi_\nu(t) = (1-t)^\nu_+ (1+\nu t)$	$\phi_\nu(t) = \frac{1}{m'}(1-t)^{\nu+m'-1}_+ (m'+\nu t)$

Table 4.2: Elements of the class $\Psi_{2m'-1}$ and corresponding element of the class Φ_m after iterative application of the operator I as defined through Proposition 4.11. The first column provides the parametric restriction ensuring the corresponding element in the second column to be in $\Psi_{2m'-1}$.

We can now use Proposition 4.11 in concert with Table 4.2 and the following Lemma, for which a proof is not provided as it follows the same path as Theorem 4.3.

Lemma 4.1. *Let m, p be positive integers and $\phi : [0, +\infty) \to \mathbb{R}$ be a member of the class Φ_m. Let $\mathbf{F}$ be a matrix-valued measure as in Proposition 4.10. Then, the scale mixture*

$$\mathbf{K}(t) = \int_0^{+\infty} \phi(rt)\,\mathbf{F}(\mathrm{d}r), \qquad t \geq 0, \tag{4.45}$$

provides an element of the class Φ^p_m.

Using the arguments in Theorem 1 of Daley et al. (2015) in concert with Lemma 4.1, one can prove that the multivariate model $\mathbf{K}(t) = [K_{ij}(t)]^p_{i,j=1}$ with

$$K_{ij}(t) = \sigma_i \sigma_{ij} \rho_{ij} \phi_{\nu_{ij}}\left(\frac{t}{b_{ij}}\right), \qquad t \geq 0,$$

is a member of the class Φ_m^p provided the constraints in Theorem 1 of Daley et al. (2015) hold. Here, ϕ_ν can be any of the entries in Table 4.2.

We finish this section by providing a direct construction that is based on the kernel ω_m as being defined through Equation (4.42).

Proposition 4.12. *Let m, p be positive integers. Let $m' := \lceil m/2 \rceil$. Let $[\rho_{ij}]_{i,j=1}^p$, $[b_{ij}]_{i,j=1}^p$ and $[\nu_{ij}]_{i,j=1}^p$ be real symmetric matrices such that $b_{ij} > 0$ and $\nu_{ij} > 2m' - 1$ for $i, j = 1, \ldots, p$. For $r > 0$, define $\mathbf{A}(r)$ with entries*

$$A_{ij}(r) = \frac{\rho_{ij}\Gamma(\frac{\nu_{ij}}{2})}{\Gamma(\frac{\nu_{ij}}{2} - m' + \frac{1}{2})\Gamma(m' - \frac{1}{2})} \left(1 - b_{ij}^2 r^2\right)_+^{\frac{\nu_{ij}}{2} - m' - \frac{1}{2}}, \quad i, j = 1, \ldots, p. \quad (4.46)$$

Then, the mapping $\mathbf{K} : [0, +\infty) \to \mathbb{R}^{p \times p}$ with entries

$$K_{ij}(x) = \frac{\rho_{ij}}{b_{ij}^{2m'-1}} \, \omega_{\nu_{ij}}\left(\frac{x}{b_{ij}}\right), \qquad x \geq 0, \quad i, j = 1, \ldots, p, \quad (4.47)$$

belongs to Φ_m^p provided the matrix $\mathbf{A}(r)$ as defined in (4.46) is positive semidefinite for any $r > 0$.

Sufficient conditions ensuring $\mathbf{A}(r)$ to be positive semidefinite for every fixed $r \geq 0$ are provided by Emery and Porcu (2023, Proposition 1). The kernel (4.47) is especially flexible as it can be used to generate new families of multivariate kernels by using the scale mixture approaches highlighted above.

We conclude this section with an illustration to the bivariate setting ($p = 2$) for a Euclidean tree with $m = 4$ leaves. The entries of the covariance function are given by $K_{11}(d_R(u, v)) = (1 - d_R(u, v)/4)_+^3$, $K_{22}(d_R(u, v)) = (1 - d_R(u, v))_+^3$, and $K_{12}(d_R(u, v)) = 0.6 \times (1 - d_R(u, v)/2.5)_+^3$, for all $u, v \in G$. Here, the collocated correlation coefficient between the two random field components is 0.6.

Figure 4.4 displays the decay of this covariance function in terms of the resistance metric measured from a reference location. It is evident that the first component of the random field has a larger correlation range, as reflected in the left panel: high (low) values of this variable tend to be surrounded by a wider extent of high (low) values. Figure 4.5 shows a realisation of a bivariate Gaussian random field over 750 sites on the tree. The Cholesky decomposition of the covariance matrix was used to perform this simulation.

4.5 Conclusions

This paper has shown the intricacies related to the construction to multivariate kernels over generalised frameworks that are represented through a broad class of metric graphs. One might argue that other classes of metric graphs are preferable in certain situations. However, to our knowledge, the only alternative has been proposed by Bolin et al. (2024) and coupled with the substantially different approach of stochastic partial differential equations. It is extremely challenging to attempt for

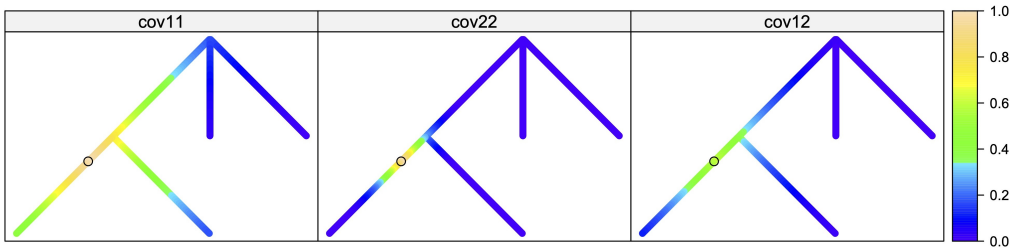


Figure 4.4: Marginal and cross covariance functions in terms of the resistance metric, over a tree with $m = 4$ leaves, measured from a reference location (black circle).

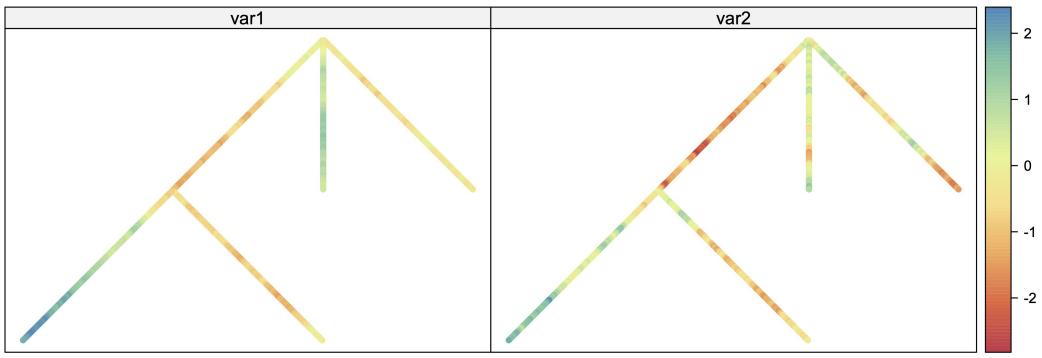


Figure 4.5: Realisation of a positively correlated bivariate random field on a tree with $m = 4$ leaves, with a covariance function of Askey type.

a comparison between these approaches. One important advantage provided by Bolin et al. (2024) for scalar-valued random fields is that an example of a once mean-square differentiable random field is provided. Yet, this is limited to a specific example, while the approach by Anderes et al. (2020) — hence, our approach — allows to embrace a wealth of examples. Being the first contribution related to vector-valued random fields on metric graphs, this paper does not have a competitor to compare with.

A next intuitive step to this research is represented by vector-valued space-time random fields. The machinery provided in this Chapter allows for a fairly general building block to start with. Space-time models might also be the building blocks to non-stationary models for networks.

The impact of this research is apparent. The works by Baddeley et al. (2017) and Moradi and Mateu (2020) are a clear indication of the importance of this work to modelling point processes over networks. There is a fervent activity related to vector-valued processes in the machine learning community, with the reader referred to Borovitskiy et al. (2023) for details.

4.6 Technical Results and Proofs

This section contains proofs and technical results. To help the reader, Figure 4.1 provides the road-map of the main results of this manuscript.

Lemma 4.2. *Let* $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ *be* $n \times n$ *matrices and let* $p \in \{1, 2, \ldots\}$. *Then*

$$\mathcal{M}_p \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathcal{M}_p \begin{pmatrix} \mathbf{C} \\ \mathbf{D} \end{pmatrix} = \mathcal{M}_p \begin{pmatrix} \mathbf{AC} + (p-1)\mathbf{BD} \\ \mathbf{AD} + \mathbf{BC} + (p-2)\mathbf{BD} \end{pmatrix}. \tag{4.48}$$

Proof of Lemma 4.2. By the standard properties of the Kronecker product,

$$\begin{aligned}
&(\mathbf{I}_p \otimes \mathbf{A} + (\mathbf{1}_{p\times p} - \mathbf{I}_p) \otimes \mathbf{B})\,(\mathbf{I}_p \otimes \mathbf{C} + (\mathbf{1}_{p\times p} - \mathbf{I}_p) \otimes \mathbf{D}) = \\
&= (\mathbf{I}_p \otimes \mathbf{A})(\mathbf{I}_p \otimes \mathbf{C}) + (\mathbf{I}_p \otimes \mathbf{A})\,((\mathbf{1}_{p\times p} - \mathbf{I}_p) \otimes \mathbf{D}) \\
&\quad + ((\mathbf{1}_{p\times p} - \mathbf{I}_p) \otimes \mathbf{B})\,(\mathbf{I}_p \otimes \mathbf{C}) + ((\mathbf{1}_{p\times p} - \mathbf{I}_p) \otimes \mathbf{B})\,((\mathbf{1}_{p\times p} - \mathbf{I}_p) \otimes \mathbf{D}) \\
&= (\mathbf{I}_p\mathbf{I}_p) \otimes (\mathbf{AC}) + (\mathbf{I}_p(\mathbf{1}_{p\times p} - \mathbf{I}_p)) \otimes (\mathbf{AD}) \\
&\quad + ((\mathbf{1}_{p\times p} - \mathbf{I}_p)\mathbf{I}_p) \otimes (\mathbf{BC}) + ((\mathbf{1}_{p\times p} - \mathbf{I}_p)(\mathbf{1}_{p\times p} - \mathbf{I}_p)) \otimes (\mathbf{BD}) \\
&= \mathbf{I}_p \otimes (\mathbf{AC}) + (\mathbf{1}_{p\times p} - \mathbf{I}_p) \otimes (\mathbf{AD} + \mathbf{BC}) + (p\mathbf{1}_{p\times p} - 2 \cdot \mathbf{1}_{p\times p} + \mathbf{I}_p) \otimes (\mathbf{BD}) \\
&= \mathbf{I}_p \otimes (\mathbf{AC}) + (\mathbf{1}_{p\times p} - \mathbf{I}_p) \otimes (\mathbf{AD} + \mathbf{BC}) \\
&\quad + \big((p-2)(\mathbf{1}_{p\times p} - \mathbf{I}_p) + (p-1)\mathbf{I}_p\big) \otimes (\mathbf{BD}) \\
&= \mathbf{I}_p \otimes (\mathbf{AC} + (p-1)\mathbf{BD}) + (\mathbf{1}_{p\times p} - \mathbf{I}_p) \otimes (\mathbf{AD} + \mathbf{BC} + (p-2)\mathbf{BD}).
\end{aligned}$$

□

Lemma 4.3. *Let* $\mathbf{L}$ *be an* $n \times n$ *quasi-Laplacian matrix and* $\boldsymbol{\Sigma} = \mathbf{L}^{-}$. *Let* $\boldsymbol{\Sigma} =: \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^\top$ *be the eigendecomposition of* $\boldsymbol{\Sigma}$, *with* $\boldsymbol{\Lambda} = \operatorname{diag}\left(\left[\lambda_1, \ldots, \lambda_n\right]\right)$ *and* $\lambda_1 \geq$

$\cdots \geq \lambda_{n-1} > \lambda_n = 0$. *Let* $\mathbf{Q}$ *be defined as in Subsection 4.3.1. Then, the eigenvalues* $\lambda_n(\mathbf{Q}) < \lambda_1(\mathbf{Q}) \leq \cdots \leq \lambda_{n-1}(\mathbf{Q})$ *of* $\mathbf{Q}$ *are:*

$$\lambda_i(\mathbf{Q}) = \begin{cases} \frac{1}{2}\left(\lambda_i^{-1} + \sqrt{\lambda_i^{-2} - 2\alpha(p-2)\lambda_i^{-1} + \alpha^2 p^2} + \alpha(p-2)\right) & i \neq n \\ \alpha(p-1) & i = n. \end{cases}$$

In particular the smallest eigenvalue of $\mathbf{Q}$, $\lambda_n(\mathbf{Q})$, *is strictly positive, thus* $\mathbf{Q}$ *is strictly positive definite. In addition,* $\mathbf{Q}\mathbf{1}_n = \alpha(p-1)\mathbf{1}_n$.

Proof of Lemma 4.3. Recall the eigendecomposition of matrix $\mathbf{Q}$ given in Equation (4.21): it is clear that the eigenvalues of $\mathbf{Q}$ are the elements on the diagonal of the matrix

$$\frac{1}{2}\left(\mathbf{\Lambda}^- + \sqrt{(\mathbf{\Lambda}^-)^2 - 2\alpha(p-2)\mathbf{\Lambda}^- + \alpha^2 p^2 \mathbf{I}_n} + \alpha(p-2)\mathbf{I}_n\right),$$

which are $\{f(\lambda^-) : \lambda \in \{\lambda_1, \ldots, \lambda_n\}\}$, where $\lambda_1, \ldots, \lambda_n$ are the diagonal entries of $\mathbf{\Lambda}$ (positive, except for λ_n that is null) and

$$f(\lambda^-) := \frac{1}{2}\left(\lambda^- + \sqrt{(\lambda^-)^2 - 2\alpha(p-2)\lambda^- + \alpha^2 p^2} + \alpha(p-2)\right).$$

Here, we recall that, for a scalar $\lambda \in \mathbb{R}$, $\lambda^- := \lambda^{-1}$ if $\lambda \neq 0$ and $\lambda^- := 0$ if $\lambda = 0$. It is straightforward to show that $\lambda^- \mapsto f(\lambda^-)$ is a strictly increasing function on $\lambda^- \geq 0$. As a consequence, since $\lambda \mapsto \lambda^-$ is strictly decreasing on $\lambda > 0$, $\lambda \mapsto f(\lambda^-)$ is strictly decreasing on $\lambda > 0$. Therefore, $\lambda_1 \geq \cdots \geq \lambda_{n-1}$ are mapped in $\lambda_1(\mathbf{Q}) = f(\lambda_1^{-1}) \leq \cdots \leq \lambda_{n-1}(\mathbf{Q}) = f(\lambda_{n-1}^{-1})$. For the case $\lambda_n = 0$, we have $\lambda_n(\mathbf{Q}) = f(\lambda_n^-) = f(0) = \frac{1}{2}(\alpha p + \alpha(p-2)) = \alpha(p-1)$. In addition, since $\lambda_n = 0 < \lambda_1(\mathbf{Q})$, we have $\alpha(p-1) = \lambda_n(\mathbf{Q}) < \lambda_1(\mathbf{Q})$.

Finally, to show that $\mathbf{Q}\mathbf{1}_n = \alpha(p-1)\mathbf{1}_n$, it is sufficient to notice from the eigendecomposition of $\mathbf{Q}$ given in Equation (4.21) that $\mathbf{1}_n$, being (proportional to) the last column of $\mathbf{W}$, is an eigenvector of $\mathbf{Q}$. And by the above result it is clear that the corresponding eigenvalue is $\lambda_n(\mathbf{Q}) = \alpha(p-1)$. This concludes the proof. □

Lemma 4.4. *Let* $\mathbf{L}$ *be an* $n \times n$ *quasi-Laplacian matrix,* $\mathbf{\Sigma} = \mathbf{L}^-$ *and* $\alpha > 0$. *Furthermore, let* $\mathbf{Q}$ *and* $\mathbf{X}$ *as defined in Subsection 4.3.1. Then the following holds:*

1. $\mathbf{\Sigma}$, $\mathbf{L}$, $\mathbf{Q}$ *and* $\mathbf{X}$ *commute;*
2. $\mathbf{\Sigma}\mathbf{L} = \mathbf{L}\mathbf{\Sigma} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_{n\times n}$;
3. $\mathbf{Q}^2 - \alpha^2(p-1)\mathbf{I}_n - \mathbf{Q}\mathbf{L} - \alpha(p-2)\mathbf{Q} + \alpha(p-2)\mathbf{L} = \mathbf{0}$;
4. $\mathbf{\Sigma}\mathbf{Q}^2 - \alpha^2(p-1)\mathbf{\Sigma} - \mathbf{Q} - \alpha(p-2)\mathbf{\Sigma}\mathbf{Q} + \alpha(p-2)\mathbf{I}_n + \frac{\alpha}{n}\mathbf{1}_{n\times n} = \mathbf{0}$.

Proof of Lemma 4.4. 1. It is sufficient to notice that all the matrices share the same eigenvector matrix $\mathbf{W}$, *i.e.* for any $\mathbf{A}, \mathbf{B} \in \{\mathbf{\Sigma}, \mathbf{L}, \mathbf{Q}, \mathbf{X}\}$, there are two

diagonal matrices $\mathbf{\Lambda_A}, \mathbf{\Lambda_B}$ such that $\mathbf{A} = \mathbf{W}\mathbf{\Lambda_A}\mathbf{W}^\top$ and $\mathbf{B} = \mathbf{W}\mathbf{\Lambda_B}\mathbf{W}^\top$. Therefore

$$\begin{aligned}\mathbf{AB} &= \mathbf{W}\mathbf{\Lambda_A}\mathbf{W}^\top\mathbf{W}\mathbf{\Lambda_B}\mathbf{W}^\top = \mathbf{W}\mathbf{\Lambda_A}\mathbf{\Lambda_B}\mathbf{W}^\top = \mathbf{W}\mathbf{\Lambda_B}\mathbf{\Lambda_A}\mathbf{W}^\top \\ &= \mathbf{W}\mathbf{\Lambda_B}\mathbf{W}^\top\mathbf{W}\mathbf{\Lambda_A}\mathbf{W}^\top = \mathbf{BA}.\end{aligned}$$

2. By exploiting the diagonalisations $\mathbf{\Sigma} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^\top$ (Lemma 4.3) and $\mathbf{L} = \mathbf{W}\mathbf{\Lambda}^-\mathbf{W}^\top$ (Eqs. (4.5) and (4.6)), we get:

$$\mathbf{\Sigma L} = \mathbf{W}\mathbf{\Lambda}\mathbf{\Lambda}^-\mathbf{W}^\top = \mathbf{W}(\mathbf{I}_n - \mathbf{J}_n)\mathbf{W}^\top = \mathbf{I}_n - \mathbf{W}\mathbf{J}_n\mathbf{W}^\top.$$

Given that $\mathbf{L}$ is a quasi-Laplacian matrix, its eigenvector matrix $\mathbf{W}$ can be written as

$$\mathbf{W} = \begin{bmatrix}\mathbf{U} & c\mathbf{1}_n\end{bmatrix},$$

where $\mathbf{U} \in \mathbb{R}^{n\times(n-1)}$ and $nc^2 = 1$. Therefore

$$\mathbf{W}\mathbf{J}_n\mathbf{W}^\top = \begin{bmatrix}\mathbf{U} & c\mathbf{1}_n\end{bmatrix}\begin{bmatrix}\mathbf{0}_{(n-1)\times(n-1)} & \mathbf{0}_{(n-1)\times 1} \\ \mathbf{0}_{1\times(n-1)} & 1\end{bmatrix}\begin{bmatrix}\mathbf{U}^\top \\ c\mathbf{1}_n{}^\top\end{bmatrix} = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n{}^\top, \quad (4.49)$$

and this concludes the proof.

3. From the definition (4.20) of $\mathbf{Q}$, one has:

$$2\mathbf{Q} - \mathbf{L} - \alpha(p-2)\mathbf{I}_n = \sqrt{\mathbf{L}^2 - 2\alpha(p-2)\mathbf{L} + \alpha^2p^2\mathbf{I}_n}.$$

Now, taking the square of both sides (recall that $\mathbf{Q}$ and $\mathbf{L}$ commute):

$$\begin{aligned}&4\mathbf{Q}^2 + \mathbf{L}^2 + \alpha^2(p-2)^2\mathbf{I}_n - 4\mathbf{QL} - 4\alpha(p-2)\mathbf{Q} + 2\alpha(p-2)\mathbf{L} \\ &= \mathbf{L}^2 - 2\alpha(p-2)\mathbf{L} + \alpha^2p^2\mathbf{I}_n.\end{aligned}$$

By grouping the common terms and dividing everything by 4 we get the desired result.

4. To show the second assertion, it is sufficient to multiply the first equation by $\mathbf{\Sigma} = \mathbf{L}^-$:

$$\begin{aligned}&\mathbf{\Sigma}\mathbf{Q}^2 - \alpha^2(p-1)\mathbf{\Sigma} - \mathbf{Q}(\mathbf{\Sigma L}) - \alpha(p-2)\mathbf{\Sigma Q} + \alpha(p-2)(\mathbf{\Sigma L}) = \mathbf{0} \\ &\mathbf{\Sigma}\mathbf{Q}^2 - \alpha^2(p-1)\mathbf{\Sigma} - \mathbf{Q} + \frac{1}{n}\lambda_n(\mathbf{Q})\mathbf{1}_{n\times n} - \alpha(p-2)\mathbf{\Sigma Q} + \alpha(p-2)\mathbf{I}_n \\ &\quad - \frac{\alpha(p-2)}{n}\mathbf{1}_{n\times n} = \mathbf{0},\end{aligned}$$

where we have used $\mathbf{\Sigma L} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_{n\times n}$ and $\mathbf{Q}\mathbf{1}_n = \lambda_n(\mathbf{Q})\mathbf{1}_n = \alpha(p-1)\mathbf{1}_n$ (Lemma 4.3).

□

Lemma 4.5. *Let $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n\times n}$ be symmetric matrices. For $i \in \{1, \dots, n\}$, let $\lambda_i(\mathbf{X})$ and $\lambda_i(\mathbf{Y})$ denote the i^{th} eigenvalues of $\mathbf{X}$ and $\mathbf{Y}$, respectively. Let $c \in \mathbb{R}$ such that $\lambda_i(\mathbf{X}) \geq c$ and $\lambda_i(\mathbf{Y}) \geq -c$ for all $i \in \{1, \dots, n\}$. Then $\mathbf{X} + \mathbf{Y}$ is positive semidefinite.*

Proof of Lemma 4.5. Let $\mathbf{X} = \mathbf{W}_1\boldsymbol{\Delta}_1\mathbf{W}_1^\top$ and $\mathbf{Y} = \mathbf{W}_2\boldsymbol{\Delta}_2\mathbf{W}_2^\top$ be the eigendecompositions of $\mathbf{X}$ and $\mathbf{Y}$ and let $\boldsymbol{v} \in \mathbb{R}^n$. In the following lines, we have defined $\boldsymbol{a} := \mathbf{W}_1^\top\boldsymbol{v}$ and $\boldsymbol{b} := \mathbf{W}_2^\top\boldsymbol{v}$.

$$\begin{aligned}\boldsymbol{v}^\top(\mathbf{X}+\mathbf{Y})\boldsymbol{v} &= \boldsymbol{v}^\top\mathbf{W}_1\Delta_1\mathbf{W}_1^\top\boldsymbol{v} + \boldsymbol{v}\mathbf{W}_2\boldsymbol{\Delta}_2\mathbf{W}_2^\top\boldsymbol{v} = \boldsymbol{a}^\top\boldsymbol{\Delta}_1\boldsymbol{a} + \boldsymbol{b}^\top\boldsymbol{\Delta}_2\boldsymbol{b} \\ &= \sum_{i=1}^n \lambda_i(\mathbf{X})a_i^2 + \sum_{i=1}^n \lambda_i(\mathbf{Y})b_i^2 \geq c\sum_{i=1}^n a_i^2 - c\sum_{i=1}^n b_i^2 \\ &= c(\boldsymbol{a}^\top\boldsymbol{a} - \boldsymbol{b}^\top\boldsymbol{b}) = c(\boldsymbol{v}^\top\boldsymbol{v} - \boldsymbol{v}^\top\boldsymbol{v}) = 0.\end{aligned}$$

Videlicet, $\mathbf{X} + \mathbf{Y}$ is positive semidefinite. □

Lemma 4.6. *Let $\mathbf{L}$ an $n \times n$ Laplacian matrix, let $\boldsymbol{\Sigma} := \mathbf{L}^-$ its Moore-Penrose inverse and let $\alpha > 0$. Let $\boldsymbol{\Theta}$, $\mathbf{Q}$, $\mathbf{X}$, k_1, k_2, $\tilde{\boldsymbol{\Sigma}}$ and $\tilde{\mathbf{X}}$ as described in Subsection 4.3.1. Finally let:*

$$\tilde{\mathbf{K}} := \mathcal{M}_p\begin{pmatrix}\tilde{\boldsymbol{\Sigma}} \\ \tilde{\mathbf{X}}\end{pmatrix}. \tag{4.50}$$

Then $\tilde{\mathbf{K}} = \boldsymbol{\Theta}^-$.

Proof of Lemma 4.6. In order to show that $\tilde{\mathbf{K}} = \boldsymbol{\Theta}^-$, by definition of the Moore-Penrose inverse, we need to prove

1. $\boldsymbol{\Theta}\tilde{\mathbf{K}}$ and $\tilde{\mathbf{K}}\boldsymbol{\Theta}$ are symmetric,
2. $\boldsymbol{\Theta}\tilde{\mathbf{K}}\boldsymbol{\Theta} = \boldsymbol{\Theta}$,
3. $\tilde{\mathbf{K}}\boldsymbol{\Theta}\tilde{\mathbf{K}} = \tilde{\mathbf{K}}$.

To show symmetry, recall from Lemma 4.4 that all the matrices $\mathbf{Q}, \alpha\mathbf{I}_n, \boldsymbol{\Sigma}, \mathbf{X}, \mathbf{1}_{n\times n}$ commute. As a consequence, $\forall \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D} \in \{\mathbf{Q}, \alpha\mathbf{I}_n, \boldsymbol{\Sigma}, \mathbf{X}, \mathbf{1}_{n\times n}\}$, it holds (Lemma 4.2):

$$\begin{aligned}&\mathcal{M}_p\begin{pmatrix}\mathbf{A}\\ \mathbf{B}\end{pmatrix}\mathcal{M}_p\begin{pmatrix}\mathbf{C}\\ \mathbf{D}\end{pmatrix} = \mathcal{M}_p\begin{pmatrix}\mathbf{AC} + (p-1)\mathbf{BD}\\ \mathbf{AD} + \mathbf{BC} + (p-2)\mathbf{BD}\end{pmatrix} \\ &= \mathcal{M}_p\begin{pmatrix}\mathbf{CA} + (p-1)\mathbf{DB}\\ \mathbf{DA} + \mathbf{CB} + (p-2)\mathbf{DB}\end{pmatrix} = \mathcal{M}_p\begin{pmatrix}\mathbf{C}\\ \mathbf{D}\end{pmatrix}\mathcal{M}_p\begin{pmatrix}\mathbf{A}\\ \mathbf{B}\end{pmatrix}.\end{aligned}$$

In addition, $\mathcal{M}_p(\mathbf{A}, \mathbf{B})^\top = \mathcal{M}_p(\mathbf{A}^\top, \mathbf{B}^\top) = \mathcal{M}_p(\mathbf{A}, \mathbf{B})$, as $\mathbf{A}, \mathbf{B}$ are symmetric. Therefore,

$$\begin{aligned}&\left(\boldsymbol{\Theta}\tilde{\mathbf{K}}\right)^\top = \tilde{\mathbf{K}}^\top\boldsymbol{\Theta}^\top = \mathcal{M}_p\begin{pmatrix}\boldsymbol{\Sigma} + k_1\mathbf{1}_{n\times n}\\ \mathbf{X} + k_2\mathbf{1}_{n\times n}\end{pmatrix}^\top\mathcal{M}_p\begin{pmatrix}\mathbf{Q}\\ -\alpha\mathbf{I}_n\end{pmatrix}^\top \\ &= \mathcal{M}_p\begin{pmatrix}\boldsymbol{\Sigma}^\top + k_1{\mathbf{1}_{n\times n}}^\top\\ \mathbf{X}^\top + k_2{\mathbf{1}_{n\times n}}^\top\end{pmatrix}\mathcal{M}_p\begin{pmatrix}\mathbf{Q}^\top\\ -\alpha\mathbf{I}_n\end{pmatrix}\end{aligned}$$

$$
\begin{aligned}
&= \mathcal{M}_p\begin{pmatrix}\boldsymbol{\Sigma}\\ \mathbf{X}\end{pmatrix}\mathcal{M}_p\begin{pmatrix}\mathbf{Q}\\ -\alpha\mathbf{I}_n\end{pmatrix} + \mathcal{M}_p\begin{pmatrix}k_1\mathbf{1}_{n\times n}\\ k_2\mathbf{1}_{n\times n}\end{pmatrix}\mathcal{M}_p\begin{pmatrix}\mathbf{Q}\\ -\alpha\mathbf{I}_n\end{pmatrix}\\
&= \mathcal{M}_p\begin{pmatrix}\mathbf{Q}\\ -\alpha\mathbf{I}_n\end{pmatrix}\mathcal{M}_p\begin{pmatrix}\boldsymbol{\Sigma}\\ \mathbf{X}\end{pmatrix} + \mathcal{M}_p\begin{pmatrix}\mathbf{Q}\\ -\alpha\mathbf{I}_n\end{pmatrix}\mathcal{M}_p\begin{pmatrix}k_1\mathbf{1}_{n\times n}\\ k_2\mathbf{1}_{n\times n}\end{pmatrix}\\
&= \mathcal{M}_p\begin{pmatrix}\mathbf{Q}\\ -\alpha\mathbf{I}_n\end{pmatrix}\mathcal{M}_p\begin{pmatrix}\boldsymbol{\Sigma}+k_1\mathbf{1}_{n\times n}\\ \mathbf{X}+k_2\mathbf{1}_{n\times n}\end{pmatrix} = \boldsymbol{\Theta}\tilde{\mathbf{K}}.
\end{aligned}
$$

The same arguments show that $\left(\tilde{\mathbf{K}}\boldsymbol{\Theta}\right)^\top = \tilde{\mathbf{K}}\boldsymbol{\Theta}$. To prove the other two relations, we start with following equalities. In the next lines, we used Lemmas 4.2, 4.3 and 4.4.

$$
\begin{aligned}
\boldsymbol{\Theta}\mathbf{K} &= \mathcal{M}_p\begin{pmatrix}\mathbf{Q}\\ -\alpha\mathbf{I}_n\end{pmatrix}\mathcal{M}_p\begin{pmatrix}\boldsymbol{\Sigma}\\ \mathbf{X}\end{pmatrix} = \mathcal{M}_p\begin{pmatrix}\mathbf{Q}\boldsymbol{\Sigma}-\alpha(p-1)\mathbf{X}\\ \mathbf{Q}\mathbf{X}-\alpha\boldsymbol{\Sigma}-\alpha(p-2)\mathbf{X}\end{pmatrix}\\
&= \mathcal{M}_p\begin{pmatrix}\mathbf{Q}\boldsymbol{\Sigma}-\alpha(p-1)\frac{1}{-\alpha(p-1)}(\mathbf{I}_n-\boldsymbol{\Sigma}\mathbf{Q})\\ \frac{1}{-\alpha(p-1)}(\mathbf{Q}-\boldsymbol{\Sigma}\mathbf{Q}^2)-\alpha\boldsymbol{\Sigma}-\alpha(p-2)\mathbf{X}\end{pmatrix} = \mathcal{M}_p\begin{pmatrix}\mathbf{I}_n\\ -\frac{1}{n(p-1)}\mathbf{1}_{n\times n}\end{pmatrix}\\
\boldsymbol{\Theta}\mathbf{K}\boldsymbol{\Theta} &= \mathcal{M}_p\begin{pmatrix}\mathbf{I}_n\\ -\frac{1}{n(p-1)}\mathbf{1}_{n\times n}\end{pmatrix}\mathcal{M}_p\begin{pmatrix}\mathbf{Q}\\ -\alpha\mathbf{I}_n\end{pmatrix}\\
&= \mathcal{M}_p\begin{pmatrix}\mathbf{Q}+\alpha(p-1)\frac{1}{n(p-1)}\mathbf{1}_{n\times n}\\ -\alpha\mathbf{I}_n-\frac{1}{n(p-1)}\mathbf{Q}\mathbf{1}_{n\times n}+\alpha(p-2)\frac{1}{n(p-1)}\mathbf{1}_{n\times n}\end{pmatrix}\\
&= \mathcal{M}_p\begin{pmatrix}\mathbf{Q}+\frac{\alpha}{n}\mathbf{1}_{n\times n}\\ -\alpha\mathbf{I}_n-\frac{\alpha}{n(p-1)}\mathbf{1}_{n\times n}\end{pmatrix}\\
\boldsymbol{\Theta}\,\mathcal{M}_p\begin{pmatrix}k_1\mathbf{1}_{n\times n}\\ k_2\mathbf{1}_{n\times n}\end{pmatrix} &= \mathcal{M}_p\begin{pmatrix}k_1\mathbf{Q}\mathbf{1}_{n\times n}-\alpha(p-1)k_2\mathbf{1}_{n\times n}\\ k_2\mathbf{Q}\mathbf{1}_{n\times n}-\alpha k_1\mathbf{1}_{n\times n}-\alpha k_2(p-2)\mathbf{1}_{n\times n}\end{pmatrix}\\
&= \mathcal{M}_p\begin{pmatrix}-\alpha(p-1)(k_2-k_1)\mathbf{1}_{n\times n}\\ -\alpha(k_1-k_2)\mathbf{1}_{n\times n}\end{pmatrix}\\
\boldsymbol{\Theta}\,\mathcal{M}_p\begin{pmatrix}k_1\mathbf{1}_{n\times n}\\ k_2\mathbf{1}_{n\times n}\end{pmatrix}\boldsymbol{\Theta} &= \mathcal{M}_p\begin{pmatrix}-\alpha(p-1)(k_2-k_1)\mathbf{1}_{n\times n}\\ -\alpha(k_1-k_2)\mathbf{1}_{n\times n}\end{pmatrix}\mathcal{M}_p\begin{pmatrix}\mathbf{Q}\\ -\alpha\mathbf{I}_n\end{pmatrix}\\
&= \mathcal{M}_p\begin{pmatrix}-\alpha(p-1)(k_2-k_1)\mathbf{Q}\mathbf{1}_{n\times n}+\alpha^2(p-1)(k_1-k_2)\mathbf{1}_{n\times n}\\ \alpha^2(p-1)(k_2-k_1)\mathbf{1}_{n\times n}-\alpha(k_1-k_2)\mathbf{Q}\mathbf{1}_{n\times n}+\alpha^2(p-2)(k_1-k_2)\mathbf{1}_{n\times n}\end{pmatrix}\\
&= \mathcal{M}_p\begin{pmatrix}\alpha^2p(p-1)(k_1-k_2)\mathbf{1}_{n\times n}\\ -\alpha^2p(k_1-k_2)\mathbf{1}_{n\times n}\end{pmatrix}
\end{aligned}
$$

As a consequence:

$$
\begin{aligned}
\boldsymbol{\Theta}\tilde{\mathbf{K}}\boldsymbol{\Theta} &= \boldsymbol{\Theta}\left(\mathbf{K}+\mathcal{M}_p\begin{pmatrix}k_1\mathbf{1}_{n\times n}\\ k_2\mathbf{1}_{n\times n}\end{pmatrix}\right)\boldsymbol{\Theta}\\
&= \mathcal{M}_p\begin{pmatrix}\mathbf{Q}+\frac{\alpha}{n}\mathbf{1}_{n\times n}\\ -\alpha\mathbf{I}_n-\frac{\alpha}{n(p-1)}\mathbf{1}_{n\times n}\end{pmatrix}+\mathcal{M}_p\begin{pmatrix}\alpha^2p(p-1)(k_1-k_2)\mathbf{1}_{n\times n}\\ -\alpha^2p(k_1-k_2)\mathbf{1}_{n\times n}\end{pmatrix}
\end{aligned}
$$

$$= \mathcal{M}_p \begin{pmatrix} \mathbf{Q} + \frac{\alpha}{n}\mathbf{1}_{n\times n} + \alpha^2 p(p-1)(k_1-k_2)\mathbf{1}_{n\times n} \\ -\alpha\mathbf{I}_n - \frac{\alpha}{n(p-1)}\mathbf{1}_{n\times n} - \alpha^2 p(k_1-k_2)\mathbf{1}_{n\times n} \end{pmatrix} = \mathcal{M}_p \begin{pmatrix} \mathbf{Q} \\ -\alpha\mathbf{I}_n \end{pmatrix} = \mathbf{\Theta}.$$

Similarly we prove that $\tilde{\mathbf{K}}\mathbf{\Theta}\tilde{\mathbf{K}} = \tilde{\mathbf{K}}$:

$$\mathbf{K}\mathbf{\Theta}\mathbf{K} = \mathcal{M}_p \begin{pmatrix} \mathbf{\Sigma} \\ \mathbf{X} \end{pmatrix} \mathcal{M}_p \begin{pmatrix} \mathbf{I}_n \\ -\frac{1}{n(p-1)}\mathbf{1}_{n\times n} \end{pmatrix} = \mathcal{M}_p \begin{pmatrix} \mathbf{\Sigma} - \frac{1}{n}\mathbf{X}\mathbf{1}_{n\times n} \\ -\frac{1}{n(p-1)}\mathbf{\Sigma}\mathbf{1}_{n\times n} + \mathbf{X} - \frac{p-2}{n(p-2)}\mathbf{X}\mathbf{1}_{n\times n} \end{pmatrix}$$

$$= \mathcal{M}_p \begin{pmatrix} \mathbf{\Sigma} + \frac{1}{\alpha n(p-1)}\mathbf{1}_{n\times n} \\ \mathbf{X} + \frac{p-2}{\alpha n(p-1)^2}\mathbf{1}_{n\times n} \end{pmatrix}$$

$$\mathcal{M}_p \begin{pmatrix} k_1\mathbf{1}_{n\times n} \\ k_2\mathbf{1}_{n\times n} \end{pmatrix} \mathbf{\Theta}\mathbf{K} = \mathbf{K}\mathbf{\Theta}\mathcal{M}_p \begin{pmatrix} k_1\mathbf{1}_{n\times n} \\ k_2\mathbf{1}_{n\times n} \end{pmatrix}$$

$$= \mathcal{M}_p \begin{pmatrix} k_1\mathbf{1}_{n\times n} \\ k_2\mathbf{1}_{n\times n} \end{pmatrix} \mathcal{M}_p \begin{pmatrix} \mathbf{I}_n \\ -\frac{1}{n(p-1)}\mathbf{1}_{n\times n} \end{pmatrix}$$

$$= \mathcal{M}_p \begin{pmatrix} k_1\mathbf{1}_{n\times n} - \frac{1}{n}k_2\mathbf{1}_{n\times n}\mathbf{1}_{n\times n} \\ -\frac{k_1}{n(p-1)}\mathbf{1}_{n\times n}\mathbf{1}_{n\times n} + k_2\mathbf{1}_{n\times n} - \frac{k_2(p-2)}{n(p-1)}\mathbf{1}_{n\times n}\mathbf{1}_{n\times n} \end{pmatrix}$$

$$= \mathcal{M}_p \begin{pmatrix} (k_1-k_2)\mathbf{1}_{n\times n} \\ -\frac{1}{p-1}(k_1-k_2)\mathbf{1}_{n\times n} \end{pmatrix}$$

$$\mathcal{M}_p \begin{pmatrix} k_1\mathbf{1}_{n\times n} \\ k_2\mathbf{1}_{n\times n} \end{pmatrix} \mathbf{\Theta}\mathcal{M}_p \begin{pmatrix} k_1\mathbf{1}_{n\times n} \\ k_2\mathbf{1}_{n\times n} \end{pmatrix} = \mathcal{M}_p \begin{pmatrix} k_1\mathbf{1}_{n\times n} \\ k_2\mathbf{1}_{n\times n} \end{pmatrix} \mathcal{M}_p \begin{pmatrix} -\alpha(p-1)(k_2-k_1)\mathbf{1}_{n\times n} \\ -\alpha(k_1-k_2)\mathbf{1}_{n\times n} \end{pmatrix}$$

$$= \mathcal{M}_p \begin{pmatrix} -\alpha(p-1)k_1(k_2-k_1)n\mathbf{1}_{n\times n} - (p-1)\alpha k_2(k_1-k_2)n\mathbf{1}_{n\times n} \\ \big(-\alpha k_1(k_1-k_2)n - \alpha(p-1)k_2(k_2-k_1)n - \alpha(p-2)k_2(k_1-k_2)n\big)\mathbf{1}_{n\times n} \end{pmatrix}$$

$$= \mathcal{M}_p \begin{pmatrix} \alpha n(p-1)(k_1-k_2)^2\mathbf{1}_{n\times n} \\ -\alpha n(k_1-k_2)^2\mathbf{1}_{n\times n} \end{pmatrix}$$

Therefore,

$$\tilde{\mathbf{K}}\mathbf{\Theta}\tilde{\mathbf{K}} = \mathbf{K}\mathbf{\Theta}\mathbf{K} + 2\,\mathcal{M}_p \begin{pmatrix} k_1\mathbf{1}_{n\times n} \\ k_2\mathbf{1}_{n\times n} \end{pmatrix} \mathbf{\Theta}\mathbf{K} + \mathcal{M}_p \begin{pmatrix} k_1\mathbf{1}_{n\times n} \\ k_2\mathbf{1}_{n\times n} \end{pmatrix} \mathbf{\Theta}\mathcal{M}_p \begin{pmatrix} k_1\mathbf{1}_{n\times n} \\ k_2\mathbf{1}_{n\times n} \end{pmatrix} = \tilde{\mathbf{K}}.$$

□

Proof of Proposition 4.7. 1. The proof of this point is straightforward, as the pseudo variogram only takes non-negative values.

2. The condition $u_1 = u_2$ and $i = j$ is sufficient to have $D_{ij}(u_1, u_2) = 0$ due to the fact that the diagonal entries of $\mathbf{\Gamma}_{\boldsymbol{Z}}(u_1, u_1)$ are zero. Let us show that it is also necessary. From the definition of the pseudo-variogram as per Equation (4.8), $D_{ij}(u_1, u_2) = 0$ if and only if $Z_i(u_1) = Z_j(u_2)$. Since $\boldsymbol{Z} = \boldsymbol{Z}_V + \boldsymbol{Z}_E$ with $\boldsymbol{Z}_V$ and $\boldsymbol{Z}_E$ being mutually independent, a necessary condition is that $Z_{V,i}(u_1) = Z_{V,j}(u_2)$. Given the correlation structure (4.24) at the vertices,

the variance-covariance matrix $\tilde{\mathbf{K}}$ of $\boldsymbol{Z}_V\big|_V$ is quasi-Laplacian (Lemma 4.6 and Proposition 4.3) and thus it has only one null eigenvalue associated with the vector $\mathbf{1}_{np}$. That is, the only non-trivial linear combination of $\boldsymbol{Z}_V\big|_V$ having a zero variance is, up to a constant factor, ${\mathbf{1}_{np}}^\top \boldsymbol{Z}_V\big|_V$. Therefore, $Z_{V,i}(u_1) = Z_{V,j}(u_2)$ for $u_1, u_2 \in V$ only if $i = j$ and $u_1 = u_2$. Given the linear interpolation on the edges (Equation (4.26)), the same holds true for any two points u_1 and u_2 belonging to G.

3. The claim stems from Equations (4.10) and (4.32).

4. Let $u_0, u_1, \ldots, u_n \in G$ and $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n \in \mathbb{R}^p$. Let $\boldsymbol{c}_0 = -\mathbf{A}\sum_{i=1}^n \boldsymbol{c}_i$, where $\mathbf{A}$ is a $p \times p$ matrix such that $\mathbf{A}^\top \mathbf{1}_p = \mathbf{1}_p$. Then, from Equation (4.11), one has the following.

$$
\begin{aligned}
0 &\geq \sum_{i,j=0}^{n} \boldsymbol{c}_i^\top \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_i, u_j)\boldsymbol{c}_j \\
&= \sum_{i,j=1}^{n} \boldsymbol{c}_i^\top \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_i, u_j)\boldsymbol{c}_j + \sum_{i=1}^{n} \boldsymbol{c}_i^\top \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_i, u_0)\boldsymbol{c}_0 + \sum_{j=1}^{n} \boldsymbol{c}_0^\top \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_j)\boldsymbol{c}_j \\
&\quad + \boldsymbol{c}_0^\top \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_0)\boldsymbol{c}_0 \\
&= \sum_{i,j=1}^{n} \boldsymbol{c}_i^\top \left[\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_i, u_j) - \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_i, u_0)\mathbf{A} - \mathbf{A}^\top \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_j) + \mathbf{A}^\top \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_0)\mathbf{A}\right] \boldsymbol{c}_j
\end{aligned}
$$

That is, the mapping

$$(u, v) \mapsto -\left(\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u, v) - \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u, u_0)\mathbf{A} - \mathbf{A}^\top \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, v) + \mathbf{A}^\top \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_0)\mathbf{A}\right)$$

is positive semidefinite (Equation 4.14).

5. Taking $\boldsymbol{c}_1 = \ldots = \boldsymbol{c}_n := \boldsymbol{c}$ and $\mathbf{A} = \mathbf{I}_p$, one gets

$$0 \geq \boldsymbol{c}^\top \sum_{i,j=1}^{n} \left[\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_i, u_j) - \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_i, u_0) - \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_j) + \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_0)\right] \boldsymbol{c}.$$

Accounting for the fact that $\boldsymbol{\Gamma}_{\boldsymbol{Z}} = \boldsymbol{\Gamma}_{\boldsymbol{Z}}^\top$ and $\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_0) = \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_1, u_1) = \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_2, u_2)$, one obtains, for $n = 1$ and $u_0 = u_2$:

$$0 \geq \boldsymbol{c}^\top \left[\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_0) - \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_1, u_2)\right] \boldsymbol{c},$$

and for $n = 2$:

$$0 \geq \boldsymbol{c}^\top \left[\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_1, u_2) + 3\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_0) - 2\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_1, u_0) - 2\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_2, u_0)\right] \boldsymbol{c},$$

i.e. , $\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_1, u_2) - \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_0)$ and $2\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_1, u_0) + 2\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_2, u_0) - \boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_1, u_2) - 3\boldsymbol{\Gamma}_{\boldsymbol{Z}}(u_0, u_0)$ are positive semidefinite matrices.

□

Proof of Proposition 4.9. We prove each row of the table, in turn. In the following computations, $\sim_a$ stands for "asymptotically equivalent to".

1. Fix $\alpha > 0$ and let $p \to +\infty$. Considering Equations (4.21) and (4.23), we have:

$$\begin{aligned} &\boldsymbol{\Lambda}^- + \alpha(p-2)\mathbf{I}_n + \sqrt{(\boldsymbol{\Lambda}^-)^2 - 2\alpha(p-2)\boldsymbol{\Lambda}^- + \alpha^2 p^2 \mathbf{I}_n} && (4.51)\\ &\quad \sim_a \alpha(p-2)\mathbf{I}_n + \alpha p \mathbf{I}_n = 2\alpha(p-1)\mathbf{I}_n \\ &\alpha(p-2)\boldsymbol{\Lambda} + \sqrt{\mathbf{I}_n - \mathbf{J}_n - 2\alpha(p-2)\boldsymbol{\Lambda} + \alpha^2 p^2 \boldsymbol{\Lambda}^2} - \mathbf{I}_n - \mathbf{J}_n && (4.52)\\ &\quad \sim_a \alpha(p-2)\boldsymbol{\Lambda} + \alpha p \boldsymbol{\Lambda} = 2\alpha(p-1)\boldsymbol{\Lambda}. \end{aligned}$$

   As a consequence, $\mathbf{Q} \sim_a \alpha(p-1)\mathbf{I}_n$ and $\mathbf{X} \to \boldsymbol{\Sigma}$. Furthermore, $k_1, k_2 \to 0$ and thus $\tilde{\boldsymbol{\Sigma}}, \tilde{\mathbf{X}} \to \boldsymbol{\Sigma}$ as $p \to +\infty$. From this last relation, we obtain that the process $\boldsymbol{Z}_V$ satisfies $\mathbb{C}\mathrm{or}\,(Z_{V,i}, Z_{V,j}) \to 1$ for all $i, j \in \{1, \dots, p\}$. Throughout the whole proof, the expression for the distance changes only via the first line of Equation (4.32), reported below for an neat exposition.

$$\begin{aligned} &\boldsymbol{\delta}_1^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{\delta}_2 - 2\boldsymbol{\delta}_1^\top \tilde{\mathbf{X}} \boldsymbol{\delta}_2 && (4.53)\\ &= \boldsymbol{\delta}_1^\top \boldsymbol{\Sigma} \boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^\top \boldsymbol{\Sigma} \boldsymbol{\delta}_2 - 2\boldsymbol{\delta}_1^\top \mathbf{X} \boldsymbol{\delta}_2 + 2k_1 - 2k_2. \end{aligned}$$

   In this case,

$$(4.53) \to (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \tilde{\boldsymbol{\Sigma}} (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)$$

   If we add the hypothesis $\beta \to 1$, $\mathbb{C}\mathrm{or}\,(Z_{E,i}, Z_{E,j}) \to 1$ as well. This implies that all the $p \times p$ entries of $\mathbf{D}(u_1, u_2)$ converge to the same term, that is $d_R(u_1, u_2)$.

2. Fix $p \geq 2$ and let $\alpha \to +\infty$. The same asymptotic properties (4.51) and (4.52) hold (notice that p and α always appear with the same exponent in each monomial). Therefore, the same reasoning applies.

3. Fix $p \geq 2$ and let $\alpha \to 0^+$. From Equation (4.20), it is patent that $\mathbf{Q} \to \frac{1}{2}(\mathbf{L} + \sqrt{\mathbf{L}^2}) = \mathbf{L}$; whilst from Equation (4.23), we have:

$$\mathbf{X} = \frac{\mathbf{W}\left((p-2)\boldsymbol{\Lambda} + \frac{1}{\alpha}\sqrt{\mathbf{I}_n - \mathbf{J}_n - 2\alpha(p-2)\boldsymbol{\Lambda} + \alpha^2 p^2 \boldsymbol{\Lambda}} - \frac{1}{\alpha}\mathbf{I}_n - \frac{1}{\alpha}\mathbf{J}_n\right)\mathbf{W}^\top}{2(p-1)}.$$

   Consider the diagonal matrix in the big brackets: it is possible to expand the square root term by means of the Taylor expansion $\sqrt{a - b\cdot\alpha + c\cdot\alpha^2} = \sqrt{a} - \frac{b}{2\sqrt{a}}\alpha + o(\alpha)$ as $\alpha \to 0^+$: the i^{th} element on the main diagonal of $\sqrt{\mathbf{I}_n - \mathbf{J}_n - 2\alpha(p-2)\boldsymbol{\Lambda} + \alpha^2 p^2 \boldsymbol{\Lambda}}$ is therefore:

$$\begin{cases} \sqrt{1} - \dfrac{2(p-2)\lambda_i}{2\sqrt{1}}\,\alpha + o(\alpha) = 1 - (p-2)\lambda_i\,\alpha + o(\alpha) & \text{if } i \neq n \\ 0 & \text{if } i = n. \end{cases}$$

As a consequence, we have:

$$\begin{aligned}(p-2)\mathbf{\Lambda} + \frac{1}{\alpha}\sqrt{\mathbf{I}_n - \mathbf{J}_n - 2\alpha(p-2)\mathbf{\Lambda} + \alpha^2 p^2 \mathbf{\Lambda}} - \frac{1}{\alpha}\mathbf{I}_n - \frac{1}{\alpha}\mathbf{J}_n \\ \sim_a (p-2)\mathbf{\Lambda} + \frac{1}{\alpha}(\mathbf{I}_n - \mathbf{J}_n) - (p-2)\mathbf{\Lambda} - \frac{1}{\alpha}\mathbf{I}_n - \frac{1}{\alpha}\mathbf{J}_n \\ = -\frac{2}{\alpha}\mathbf{J}_n,\end{aligned}$$

that is:

$$\mathbf{X} \sim \frac{1}{2(p-1)}\mathbf{W}\left(-\frac{2}{\alpha}\mathbf{J}_n\right)\mathbf{W}^\top = -\frac{1}{\alpha(p-1)}\frac{1}{n}\mathbf{1}_{n\times n},$$

where in the last step we have used from the proof of Lemma 4.4 that $\mathbf{W}\mathbf{J}_n\mathbf{W}^\top = \frac{1}{n}\mathbf{1}_{n\times n}$ (Equation (4.49)). In this case

$$\begin{aligned}(4.53) &\sim_a \boldsymbol{\delta}_1^\top \mathbf{\Sigma}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^\top \mathbf{\Sigma}\boldsymbol{\delta}_2 - 2\boldsymbol{\delta}_1^\top\left(-\frac{1}{\alpha n(p-1)}\mathbf{1}_{n\times n}\right)\boldsymbol{\delta}_2 - 2\frac{1}{\alpha np(p-1)} \\ &= \boldsymbol{\delta}_1^\top \mathbf{\Sigma}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^\top \mathbf{\Sigma}\boldsymbol{\delta}_2 + \frac{2}{\alpha np}.\end{aligned}$$

Clearly, the contribution of the process $\boldsymbol{Z}_E$ in Equation (4.32) and the above term $\boldsymbol{\delta}_1^\top \mathbf{\Sigma}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^\top \mathbf{\Sigma}\boldsymbol{\delta}_2$ do not depend on α. As a consequence, $D_{ij}(u_1, u_2) \sim_a \frac{2}{\alpha np}$ for $i \neq j$ as $\alpha \to 0^+$.

4. Let $\alpha \to 0^+$ and $p \to +\infty$ such that $\alpha p\tau \to 1$ with $0 < \tau \leq \lambda_{n-1}$ being fixed. Then, from Equations (4.21) and (4.6), one has $\mathbf{Q} \to \frac{\mathbf{I}_n}{\tau}$. Likewise, from Equation (4.23), it comes $\mathbf{X} \to \mathbf{\Sigma} - \tau\mathbf{I}_n$. The computation of the distances $D_{ij}(u_1, u_2)$ is straightforward: following the same reasoning of the previous derivation, we have

$$\begin{aligned}(4.53) &= \boldsymbol{\delta}_1^\top \mathbf{\Sigma}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^\top \mathbf{\Sigma}\boldsymbol{\delta}_2 - 2\boldsymbol{\delta}_1^\top(\mathbf{\Sigma} - \tau\mathbf{I}_n)\boldsymbol{\delta}_2 - 2\frac{1}{\alpha np(p-1)} \\ &\to (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2)^\top \mathbf{\Sigma}(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2) + 2\tau\boldsymbol{\delta}_1^\top\boldsymbol{\delta}_2.\end{aligned}$$

The first addendum, together with the contribution of the process $\boldsymbol{Z}_E$ (if $\beta \to 1$), gives the distance $d_R(u_1, u_2)$.

5. Let $\alpha \to 0^+$ and $p \to +\infty$, with $\alpha p\tau \to 1$ with $\tau \geq \lambda_1$ fixed. Equation (4.21) gives

$$\mathbf{Q} \to \mathbf{W}\,\mathrm{diag}\,[\lambda_1^{-1}, \ldots, \lambda_{n-1}^{-1}, \frac{1}{\tau}]^\top\mathbf{W}^\top = \mathbf{W}\left(\mathbf{\Lambda}^- + \frac{1}{\tau}\mathbf{J}_n\right)\mathbf{W}^\top = \mathbf{L} + \frac{1}{n\tau}\mathbf{1}_{n\times n}.$$

Therefore, from Equation (4.22):

$$\mathbf{X} = \frac{p\tau}{p-1}\left(\mathbf{\Sigma}\left(\mathbf{L} + \frac{1}{n\tau}\mathbf{1}_{n\times n}.\right) - \mathbf{I}_n\right) \to -\frac{\tau}{n}\mathbf{1}_{n\times n}, \qquad \text{as } p \to +\infty.$$

Hence,

$$(4.53) \to \boldsymbol{\delta}_1^\top \boldsymbol{\Sigma} \boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^\top \boldsymbol{\Sigma} \boldsymbol{\delta}_2 - 2\boldsymbol{\delta}_1^\top \left(-\frac{\tau}{n}\mathbf{1}_{n\times n}\right)\boldsymbol{\delta}_2 - 2\frac{1}{\alpha np(p-1)}$$
$$\to \boldsymbol{\delta}_1^\top \boldsymbol{\Sigma} \boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^\top \boldsymbol{\Sigma} \boldsymbol{\delta}_2 + \frac{2\tau}{n}.$$

The result for the distance $D_{ij}(u_1, u_2)$ is obtained from the last equation by adding and subtracting $2\boldsymbol{\delta}_1^\top \mathbf{X}\boldsymbol{\delta}_2$ and by using the fact that $\beta \to 1$.

□

Proof of Theorem 4.2. The result stems from Theorem 4.1 and the fact that the pseudo-variogram $\mathbf{D}$ satisfies $\mathbf{D}^\top(u_1, u_2) = \mathbf{D}(u_2, u_1)$ and is $\mathbf{1}_p$-conditionally negative semidefinite. □

Proof of Theorem 4.3. The result stems from Theorem 4.2 with $\psi(t) = \exp(-t)$ and $g(t) = t^\theta$ and the fact that the set of positive semidefinite matrix-valued functions is closed under sums, point-wise limits and element-wise products with symmetric positive semidefinite matrices. □

Proof of Theorem 4.5. From Emery et al. (2022, Equations (28)-(30) and Theorem 3), $\boldsymbol{\mathcal{M}}_p$ has a representation of the form (4.35) with $\mathrm{d}F_{ij}(\xi) = R_{ij}(\xi)\mathrm{d}\xi$, where $[R_{ij}(\xi)]_{i,j=1}^p$ is positive semidefinite under the stated conditions (A) or (B). The theorem thus results from the application of Theorem 4.3. □

Proof of Theorem 4.6. Using formula 3.381.4 of Gradshteyn and Ryzhik (2007), one can write

$$\boldsymbol{\mathcal{C}}_p(\mathbf{D}) = \left[\frac{\sigma_{ij}\beta_{ij}^\nu}{\Gamma(\nu)}\int_0^{+\infty} \xi^{\nu-1}\exp(-\beta_{ij}\xi)\exp(-\xi D_{ij}^\theta)\mathrm{d}\xi\right]_{i,j=1}^p.$$

The conditional negative semi-definiteness of $[\beta_{ij}]_{i,j=1}^p$ implies that

$$\left[\frac{\xi^{\nu-1}}{\Gamma(\nu)}\exp(-\beta_{ij}\xi)\right]_{i,j=1}^p$$

is positive semidefinite for any $\xi \geq 0$ (Reams, 1999, Lemma 2.5). The theorem results from the application of the Schur product theorem and Theorem 4.3. □

Proof of Theorem 4.7. One can write (Porcu et al., 2018)

$$\boldsymbol{\mathcal{SG}}_p(\mathbf{D}) = \left[\frac{\sigma_{ij}2^{\nu-1}(\alpha_{ij}\beta_{ij})^{\nu/2}}{\mathcal{K}_\nu\left(\sqrt{\frac{\beta_{ij}}{\alpha_{ij}}}\right)}\int_0^{+\infty}\xi^{\nu-1}\exp\left(-\frac{1}{4\alpha_{ij}\xi}-\beta_{ij}\xi-\xi D_{ij}^\theta\right)\mathrm{d}\xi\right]_{i,j=1}^p.$$

The fact that $[\alpha_{ij}^{-1}]_{i,j=1}^p$ and $[\beta_{ij}]_{i,j=1}^p$ are conditionally negative semidefinite implies that $\left[\exp\left(-\frac{1}{4\alpha_{ij}\xi}-\beta_{ij}\xi\right)\right]_{i,j=1}^p$ is positive semidefinite for any $\xi \geq 0$. The theorem results from the application of the Schur product theorem and Theorem 4.3. □

Proof of Proposition 4.10. We start with the case $p = 1$ and observe that, by Bochner's theorem (Bochner, 1955), a continuous mapping $C : \mathbb{R}^{m'} \to \mathbb{R}$ is positive semidefinite if and only if

$$C(\boldsymbol{x}) = \int_{\mathbb{R}^{m'}} \exp(\mathsf{i}\boldsymbol{x}^\top \boldsymbol{\omega}) G(\mathrm{d}\boldsymbol{\omega}), \qquad \boldsymbol{x} \in \mathbb{R}^{m'},$$

for G being a non-decreasing and bounded measure in $\mathbb{R}^m$. Here, i denotes the unit complex number. Using the above representation, Cambanis et al. (1983) show that a mapping $K : [0, +\infty) \to \mathbb{R}$ belongs to the class Φ_m^1 if and only if Equation (4.41) holds for $p = 1$ and for F being non-decreasing and bounded. In particular, choosing a Dirac measure for F, one finds that $t \mapsto \omega_{m'}(rt)$ belongs to Φ_m^1 for any $r \geq 0$, that is, $\omega_{m'}(r d_M(\cdot, \cdot))$ is positive semidefinite on $\mathbb{R}^{m'}$.

We now address the general case with $p \geq 1$. First, consider a matrix-valued mapping $\mathbf{K}$ belonging to the class Φ_m^p. We invoke the extension of Bochner's theorem to the p-variate case (Cramer, 1940), for which a continuous mapping $\mathbf{C} : \mathbb{R}^{m'} \to \mathbb{R}^{p \times p}$ is positive semidefinite if and only if

$$\mathbf{C}(\boldsymbol{x}) = \int_{\mathbb{R}^{m'}} \exp(\mathsf{i}\boldsymbol{x}^\top \boldsymbol{\omega}) \mathbf{G}(\mathrm{d}\boldsymbol{\omega}), \qquad \boldsymbol{x} \in \mathbb{R}^{m'},$$

for $\mathbf{G}$ a bounded complex Hermitian matrix-valued measure. Hence, the same calculations as much as in Cambanis et al. (1983) show that the representation (4.41) holds for a bounded symmetric real matrix-valued measure $\mathbf{F}$ (that $\mathbf{F}$ is bounded stems from the fact that $\mathbf{K}(0)$ is finite). To finish this part of the proof we need to show that $\mathrm{d}\mathbf{F}(r)$ is positive semidefinite for any $r \geq 0$, which happens if and only if the mapping

$$H_{\boldsymbol{c}}(r) := \boldsymbol{c}^\top \mathbf{F}(\mathrm{d}r)\boldsymbol{c}, \quad r \in [0, +\infty),$$

is non-negative for any arbitrary $\boldsymbol{c} \in \mathbb{R}^p$. To this end, consider a p-variate random field $\boldsymbol{Z}$ in $\mathbb{R}^{m'}$ with isotropic covariance $\mathbf{K}$, and define $Z_{\boldsymbol{c}} := \boldsymbol{c}^\top \boldsymbol{Z}$, which is a univariate random field with isotropic covariance $K_{\boldsymbol{c}} = \boldsymbol{c}^\top \mathbf{K} \boldsymbol{c}$. Hence, $K_{\boldsymbol{c}}$ obeys to the representation provided by Cambanis et al. (1983), with a measure $F_{\boldsymbol{c}}$ such that

$$F_{\boldsymbol{c}} = \boldsymbol{c}^\top \mathbf{F} \boldsymbol{c}.$$

This proves the sufficiency part of the theorem by noting that $F_{\boldsymbol{c}}$ is non-decreasing.

Reciprocally, assume that the representation (4.41) holds with a bounded symmetric matrix-valued measure $\mathbf{F}$ such that $\mathrm{d}\mathbf{F}(r)$ is positive semidefinite for any $r \geq 0$. Being positive semidefinite on $\mathbb{R}^{m'}$ for any $r \geq 0$, $\omega_{m'}(r d_M(\cdot, \cdot))$ is the covariance function of a univariate Gaussian random field Z_r on $\mathbb{R}^{m'}$. Accordingly, the matrix-valued mapping $\omega_{m'}(r d_M(\cdot, \cdot)) \mathbf{1}_{p \times p}$ belongs to Φ_m^p as the covariance function of $Z_r \mathbf{1}_p$, and so do the mappings $\omega_{m'}(r d_M(\cdot, \cdot)) \mathrm{d}\mathbf{F}(r)$ and $\mathbf{K}(d_M(\cdot, \cdot))$ owing to the fact that the set of positive semidefinite functions is closed under sums, element-wise products, and pointwise limits. □

Proof of Proposition 4.11. We provide a proof by direct construction. Let m' be a positive integer. Arguments in Theorem 2.1 of Gneiting (1998) provide the identity

$$\omega_{m'}(t) = \kappa_{m'} I^{(m'-1)} \Omega_{2m'-1}(t), \qquad t \geq 0, \tag{4.54}$$

for $\kappa_{m'}$ a strictly positive constant. Hence, the proof comes from the integral representation (4.41) in concert with a straight use of Fubini's theorem. □

Proof of Proposition 4.12. The results comes from a straight application of Proposition 2 in Emery and Porcu (2023) in concert with Lemma 4.1, Proposition 4.11 and Fubini's theorem. □

Chapter 5

Computationally Efficient Algorithms for Simulating Isotropic Gaussian Random Fields on Graphs with Euclidean Edges[1]

5.1 Introduction

Random fields having networks as an index set are ubiquitous across various applications, such as assessing ecological and environmental impacts in road networks, understanding the propagation of signals through dendrites, or modelling the flow of rivers, communication networks and electrical wires. For a recent comprehensive review on this subject, readers are directed to Baddeley et al. (2021). Several applications involving non-standard networks are discussed by Porcu et al. (2023).

The statistical analysis of network data is challenging. Utilising the coordinates of points on the network akin to coordinates in a Euclidean space where the network is embedded is conceptually flawed, distorting the proximity among points, and potentially producing unrealistic forecasts at locations of the network that have not been observed.

Our paper deals with the simulation of spatial random fields having a (not necessarily linear) network as an index set. In our pursuit of formulating simulation methodologies, our aim is to integrate the extensive body of knowledge and practicality offered by linear networks, while also addressing the challenges posed by more intricate network structures. Consequently, the findings presented throughout this paper retain their validity across various network architectures, as described below,

[1]This chapter is the paper Alegría et al. (2025), accepted for publication in Journal of Computational and Graphical Statistics and available on arXiv:2404.17491.

thereby expanding their score of applicability. This general formulation can prove useful in situations where linear network structures may not accurately represent the complexity of the spatial domain (Ver Hoef et al., 2006; Baddeley et al., 2021; Porcu et al., 2023).

The problem of simulation *on* networks is apparently different from that of simulation *of* networks. For the latter, there is a massive literature at hand, coming from computer science, machine learning and applied mathematics. For the former, the literature is sparse as outlined below.

For a Gaussian random field defined over any network, it is customary to assume that the covariance function is isotropic (Anderes et al., 2020), that is, the covariance between any pair of observations located over two different points of the network is solely a function of the *distance* between the points. While isotropy is intuitive and natural for random fields defined over Euclidean spaces or manifold, defining distances over networks is tricky. The problem can be understood by noting that, essentially, a non-linear network is a (semi) metric graph, that is a graph endowed with a (semi) metric. For real-world applications, the isotropy assumption can serve as a preliminary step in the development of more versatile models for which the covariance is not only a function of the distance between paired observations. Some classical approaches in this regard involve deforming the graph to create intricate patterns or using parameters in the correlation function that vary across the graph. The latter approach is a natural and feasible way to generalise the results presented in this paper, while the former, which requires bijective mappings from a graph to itself, seems more challenging a priori.

This paper challenges the problem of simulation of (approximately) Gaussian random fields having a metric graph as an index set. Specifically, the type of metric graph considered in this paper is termed a *graph with Euclidean edges* (Anderes et al., 2020). These graphs allow for a continuously defined random field where points are defined in both vertices and arcs. This framework proves valuable for analysing networks featuring tunnels or bridges, incorporating weighted edges within the graph structure in a cohesive manner.

Anderes et al. (2020) proved that Gaussian random fields on graphs with Euclidean edges can be endowed with, either, the geodesic distance — although this is rarely possible in practice, as it requires fulfilling specific technical constraints — or the resistance metric, being an extension of the conventional metric used within the framework of electrical networks (Klein and Randic, 1993). The latter metric not only provides a physical interpretation but also allows for the establishment of more flexible correlation structures on the graph, accommodating a broader range of real-world applications, unlike the more restrictive geodesic distance.

The simulation of random fields over non-linear networks has been addressed to a limited extent only. Møller and Rasmussen (2024) proposed simulation algorithms on linear networks that employ operations of large covariance matrices, such as the Cholesky decomposition, and are challenging from a computational viewpoint when the sample size is large. Bolin et al. (2024) studied a spectral decomposition

for random fields on metric graphs. This approach depends on the eigenfunctions and eigenvalues of a differential operator, covering the case of a covariance function analogous to the Matérn model and depending on geodesic distance. For other families of covariance functions, this approach is demanding, if not unfeasible. Bolin et al. (2023a,b) provided fast simulation methods for a Whittle-Matérn random field, based on finite element and rational approximations and, when the smoothness parameter is an integer, on the Markov property of the random field.

Although the approaches provided above offer insightful techniques that are computationally fast and reliable, we note that they are limited to random fields associated with specific classes of SPDE. This is not an issue *per se*. Yet, alternative paths might allow for exploring simulations of random fields that are not necessarily associated with specific SPDEs and Markovian representations. In particular, in this paper, we will exploit the connection between resistance metric with the variogram of certain classes of *ad hoc* random fields in concert with three alternative stochastic representations. This will lead to three efficient simulation algorithms inspired by classical geostatistical approaches that have been proposed to simulate isotropic random fields (with respect to Euclidean distance) in $\mathbb{R}^d$, for d a positive integer. Our proposal is rich: we are able to simulate Gaussian random fields from a wealth of choices of covariance functions depending on the resistance metric.

We then proceed to illustrate algorithmic effectiveness through examples built upon existing popular networks found in R packages, such as the `spatstat` package (Baddeley and Turner, 2005). Our simulation routines are compatible with the syntax and structure of networks within these packages. The codes are accessible in the GitHub repository `https://github.com/alfredoalegria/FastSimNetworks` and the results within the manuscript are reproducible.

The paper is structured as follows. Section 5.2 presents preliminaries on graphs with Euclidean edges and the resistance metric. Section 5.3 introduces three stochastic representations for random fields on the graph, serving as the foundation for the simulation algorithms described in Section 5.4. Section 5.5 illustrates several numerical examples and offers guidelines for practitioners. Concluding remarks (Section 5.6) close the paper.

5.2 Preliminaries

5.2.1 Graphs with Euclidean Edges

The following lines provide a succinct and non-rigorous illustration of graphs with Euclidean edges. For a formal treatment, the reader is referred to Anderes et al. (2020) and Porcu et al. (2023). The recent works related to random fields that are continuously defined over networks are substantially based on the following facts:

1. A non-linear network can be represented through a special graph structure as depicted below. The graph allows for non-standard topologies that elude the limits of linear network representations.

2. In turn, the graphs proposed in the recent literature, endowed with a proper (semi) metric, become (semi) metric spaces.

3. (Semi) metric spaces can be *embedded* into convenient function spaces where implementing covariance functions become straightforward thanks to the contributions from the early 1940ies (Schoenberg, 1942).

The type of topology considered by Anderes et al. (2020) is termed that of a graph *with Euclidean edges*, denoted with a triple $G = (V, E, \{\varphi_e\}_{e\in E})$ where the elements are blended in the following way:

(a) (V, E) has a graph structure, where V is the set of vertices and E contains the edges. We assume that this graph is simple, undirected and connected, i.e., V is finite, the graph has not repeated edges (multi-edges connecting the same pair of vertices) or edges that join a vertex to itself (loops), the edges have no specified direction, and every pair of vertices is connected by a path.

(b) Each edge $e \in E$ is associated with a unique abstract set, also denoted e, such that V and all the edge sets are mutually disjoint.

(c) Let u and v be vertices connected by $e \in E$. Then, φ_e is a bijective mapping defined on $e \cup \{u, v\}$, which maps e onto an open interval $(\underline{e}, \overline{e}) \subset \mathbb{R}$ and $\{u, v\}$ onto $\{\underline{e}, \overline{e}\}$.

Throughout, when we write $u \in G$, we are referring to an element that can be either a vertex or a point within an edge set, i.e., $u \in V \cup \bigcup_{e\in E} e$.

Some comments are in order. A linear network is formally defined as the union of linear segments, denoted as $\ell_i \subset \mathbb{R}^2$, $i \in I$, and these segments intersect only at their endpoints. Apparently, the definition above shows that a linear network is a special case of a graph with Euclidean edges: the endpoints of the segments constitute the set of vertices, while each edge $e_i \in E$ is defined as the relative interior of the linear segment ℓ_i. The bijection φ_{e_i} corresponds to the inverse of the length-path parametrisation of ℓ_i. However, graphs with Euclidean edges allow for far more flexibility; in particular, they can model networks with bridges or tunnels. Insightful graphical representations of such graphs that are not linear networks can be found in Anderes et al. (2020) and Filosi et al. (2025a) (see Chapter 3).

An additional recall to notation will provide an easier exposition throughout. Two vertices $u, v \in V$ are neighbours if there exists an edge in E connecting them. If $u \in e \in E$, we let $\underline{u}$ and $\overline{u}$ represent the vertices that are connected by e, such that $\underline{u}$ and $\overline{u}$ correspond, respectively, to $\underline{e}$ and $\overline{e}$. When $u \in V$, we define $\underline{u} = \overline{u} = u$.

5.2.2 Gaussian Random Fields on Graphs

A random field on G is an uncountable collection $Y := \{Y(u),\ u \in G\}$ of random variables in the same probability space. It is called Gaussian when $(Y(u_1), \ldots, Y(u_n))^\top$, with $\top$ denoting the transpose of a vector and n a positive integer, is a Gaussian

random vector, for all $u_1, \ldots, u_n \in G$, with a given covariance matrix $\mathbf{\Sigma} = (\Sigma_{ij})_{i,j=1}^{n}$. In this case, the modeling, inference and prediction are solely determined by the mean and the covariance function $k : G \times G \to \mathbb{R}$, which is a positive semidefinite mapping such that $\Sigma_{ij} = k(u_i, u_j)$, for $u_i, u_j \in G$. The function K is called isotropic for the graph G when there exists a pair (C, d), with d a suitable semi metric, such that $K(u, v) = C(d(u, v))$, and for a mapping $C : \sigma_{d.} \to \mathbb{R}$, with $\sigma_{d.}$ being the disk of $d.$, that is the image space for the mapping d.

Apparently, once the suitable (semi) metric d is found, the pair (G, d) becomes a (semi) metric space. As for the functions C such that $C(d(\cdot, \cdot))$ is positive semidefinite, we defer this discussion as we first define the proper metric for the graph used in this paper.

5.2.3 Endowing the Graph with the Resistance Metric

While the geodesic (aka, shortest-path) distance is a physically natural candidate for a metric over a graph with Euclidean edges, such a choice resents from a collection of drawbacks and structural limitations when implementing a suitable covariance function based on that metric (Anderes et al., 2020; Porcu et al., 2023). This fact motivated Anderes et al. (2020) to search for a metric that is methodologically more flexible. A solution is found through a generalisation of the classical resistance metric that is reminiscent to electrical networks (Klein and Randic, 1993).

Formally, the resistance metric on G is defined as

$$d(u, v) = \mathbb{V}\mathrm{ar}\left(Z_G(u) - Z_G(v)\right), \qquad u, v \in G, \tag{5.1}$$

where Z_G is an auxiliary random field on G, adopting the following form

$$Z_G(u) = Z_\mu(u) + \sum_{e \in E} Z_e(u), \qquad u \in G. \tag{5.2}$$

Let us establish a precise definition of the random fields Z_μ and Z_e involved in (5.2):

- Firstly, Z_μ is a Gaussian vector on the vertices $v_1, \ldots, v_n \in V$:

  $$\left(Z_\mu(v_1), \ldots, Z_\mu(v_n)\right)^\top \sim \mathcal{N}_n\left(\mathbf{0}, \mathbf{L}^{-1}\right),$$

  where L, described next, is a variation of the Laplacian matrix. Consider the function $c : V \times V \to [0, \infty)$ such that $c(u, v) = 1/d_G(u, v)$ when u and v are neighbors, with $d_G(u, v)$ being the length of the edge joining these points, and $c(u, v) = 0$ otherwise, and the function $c(u) = \sum_v c(u, v)$. For an arbitrary vertex u_0, one defines

  $$L(u, v) = \begin{cases} 1 + c(u_0) & \text{if } u = v = u_0 \\ c(u) & \text{if } u = v \neq u_0 \\ -c(u, v) & \text{otherwise.} \end{cases}$$

  The resulting matrix is strictly positive definite as a consequence of the extra term 1 in an entry of the main diagonal (Anderes et al., 2020).

- Secondly, Z_μ is extended to the edges by employing a linear interpolation of the form
$$Z_\mu(u) = (1 - \delta(u))Z_\mu(\underline{u}) + \delta(u)Z_\mu(\overline{u}), \tag{5.3}$$
where
$$\delta(u) = \begin{cases} d_G(u, \underline{u})/d_G(\underline{u}, \overline{u}) & \text{if } u \notin V(G) \\ 0 & \text{otherwise.} \end{cases}$$

- Finally, $Z_e(u) = B_e(\varphi_e(u))$ for $u \in e$, and $Z_e(u) = 0$ otherwise, where B_e is a standard independent Brownian bridge independent of Z_μ, such that $B_e(\underline{e}) = B_e(\overline{e}) = 0$.

The resistance metric adheres to fundamental criteria: it is non-negative and symmetric, $d_R(u, v) = 0$ if and only if $u = v$, and it satisfies the triangular inequality. Moreover, it maintains invariance under certain operations, such as splitting an edge into two (with the addition of an extra vertex along that edge's path) or merging edges at a vertex with two incident edges (effectively removing that vertex) (Anderes et al., 2020).

5.2.4 Which Functions C can be Coupled with the Metric d?

Anderes et al. (2020) prove that a sufficient condition for a continuous mapping $C : \sigma_d \to \mathbb{R}$ to ensure $C(d_R(\cdot, \cdot))$ is positive semidefinite on $G \times G$ is that C is the restriction of a completely monotone function ψ on the real line, i.e., the Laplace transform of a non-negative and finite measure μ on $[0, \infty)$:

$$\psi(x) = \int_0^\infty e^{-xt}\, \mu(\mathrm{d}t), \quad x \in [0, \infty). \tag{5.4}$$

Each completely monotone function ψ is non-negative on its domain. Notice that the condition $\psi(0) = 1$ is equivalent to $\mu([0, \infty)) = 1$, that is μ is a probability distribution.

5.3 Useful Stochastic Representations

5.3.1 First Construction (Spectral Representation)

We study the class of random fields on G defined according to

$$Y(u) := \sqrt{-2\ln(V)}\cos(Z_G(u)W + \Lambda), \qquad u \in G, \tag{5.5}$$

where Z_G is the Gaussian random field introduced in (5.2) and involved in the definition of the resistance metric. Here, $V \sim \mathrm{Unif}(0, 1)$, $\Lambda \sim \mathrm{Unif}(0, 2\pi)$ and $W \sim F$, with F standing for a probability measure on $\mathbb{R}$ with no atom at the origin (i.e., W is almost surely different from 0). We assume that Z_G, V, Λ and W are independent.

This construction is inspired by spectral and substitution methods developed in the geostatistics literature to simulate random fields in Euclidean spaces (Lantuéjoul, 2002; Allard et al., 2020). The following proposition demonstrates that this approach enables the generation of isotropic random fields with a broad spectrum of correlation functions.

Proposition 5.1. *The random field introduced in (5.5) possesses a zero mean. Its covariance function, denoted by C_Y, is isotropic with respect to the resistance metric and is given by*

$$C_Y\left(d(u,v)\right) = \int_{\mathbb{R}} \exp\left(-d(u,v)\omega^2/2\right) F(d\omega), \qquad u,v \in G. \tag{5.6}$$

Proof. Note that (5.5) can be written as

$$Y(u) = \sqrt{-2\ln(V)}\big[\cos(Z_G(u)W)\cos(\Lambda) - \sin(Z_G(u)W)\sin(\Lambda)\big]. \tag{5.7}$$

Thus, by considering the fact that $\mathbb{E}(\cos\Lambda) = \mathbb{E}(\sin\Lambda) = 0$ and that Λ is independent of V, W and Z_G, it becomes evident that $Y(u)$ has a zero mean. By also accounting for the fact that $\mathbb{E}(\cos\Lambda\sin\Lambda) = 0$, $\mathbb{E}(\cos^2\Lambda) = \mathbb{E}(\sin^2\Lambda) = 1/2$ and $\mathbb{E}(-\ln(V)) = 1$, one obtains

$$\mathbb{C}\mathrm{ov}\left(Y(u), Y(v)\right) = \mathbb{E}\left[\cos\left(W\{Z_G(u) - Z_G(v)\}\right)\right], \qquad u,v \in G, \tag{5.8}$$

where expectation is taken with respect to W and $(Z_G(u), Z_G(v))^\top$. Since $(Z_G(u) - Z_G(v))/\sqrt{d(u,v)}$ follows a standard Gaussian distribution, (5.8) can be written as

$$\mathbb{C}\mathrm{ov}\left(Y(u), Y(v)\right) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \cos\left(\sqrt{d(u,v)}\omega z\right) \exp\left(-z^2/2\right) \mathrm{d}z\right] F(\mathrm{d}\omega). \tag{5.9}$$

We conclude that the random field Y is isotropic with respect to the resistance metric, and the expression (5.6) is readily obtained by solving the integral within the brackets in (5.9) (Gradshteyn and Ryzhik, 2007, formula 3.896.4). □

Proposition 5.1 shows that the covariance function C_Y is a scale mixture of exponential covariances, therefore it is a completely monotone function. Reciprocally, a straightforward change of variable in (5.6) shows that any completely monotone covariance function can be obtained with the proposed construction. In particular, the Matérn model can be derived, but it would be evaluated at $\sqrt{d}$ rather than d. As a result, the full range of smoothness of the sample paths cannot be recovered, which agrees with Anderes et al. (2020).

Table 5.1 provides a list of parametric families of correlation functions on G, for different choices of the spectral measure F. Some of the entries (1, 6 and 8) have been reported in Anderes et al. (2020). The last seven entries have been established by using formulae 4.2.16, 4.2.25, 4.2.26, 4.4.15, 4.5.3, 4.5.23 and 4.5.28 of Erdélyi (1954) to calculate (5.6).

Concerning the random field distributions, one has the following result.

$F(\mathrm{d}\omega)$	C_Y
$\mathrm{d}\delta_\omega(\{a\})$	$C_Y(d) = \exp(-a^2 d/2)$
$(2a)^{-1}1_{\lvert\omega\rvert<a}d\omega$	$C_Y(d) = \sqrt{\pi}(2a^2 d)^{-1/2}\operatorname{erf}(a\sqrt{d/2})$
$(\pi a)^{-1}(1+\omega^2/a^2)^{-1}\mathrm{d}\omega$	$C_Y(d) = \exp(a^2 d/2)[1-\operatorname{erf}(-a\sqrt{d})]$
$a(\pi\lvert\omega\rvert)^{-1}(\omega^2-a^2)^{-1/2}1_{\lvert\omega\rvert>a}d\omega$	$C_Y(d) = 1-\operatorname{erf}(a\sqrt{d/2})$
$(\lvert\frac{\omega}{a}\rvert^3 1_{\lvert\omega\rvert<a} + \lvert\frac{\omega}{a}\rvert(2-(\frac{\omega}{a})^2)1_{a<\lvert\omega\rvert<2a})\frac{d\omega}{a}$	$C_Y(d) = (a^2 d/2)^{-2}(1-\exp(-a^2 d/2))^2$
$a^\tau\lvert\omega\rvert^{2\tau-1}\exp(-a\omega^2)d\omega/\Gamma(\tau)$	$C_Y(d) = (2a)^\tau(2a+d)^{-\tau}$
$(a\Gamma(1/4))^{-1}\sqrt{2}\exp(-\omega^4/(4a^4))d\omega$	$C_Y(d) = a\sqrt{d}\exp(a^4d^2/8)K_{1/4}(a^4d^2/8)/\Gamma(1/4)$
$a\omega^{-2}\exp(-a^2/(4\omega^2))/(2\sqrt{\pi})d\omega$	$C_Y(d) = \exp(-a\sqrt{d/2})$

Table 5.1: Parametric correlation functions on G in terms of the resistance metric. For each model C_Y, we also provide the associated spectral measure F. Γ is the gamma function, erf is the Gauss error function, K_τ is the modified Bessel function of the second kind of order τ, $\tau > 0$ is a shape parameter, and $a > 0$ is a scale parameter.

Proposition 5.2. *The random field defined in (5.5) has a standard Gaussian univariate distribution. Furthermore, its bivariate distributions are isotropic with respect to the resistance metric. They can be written as mixtures of bivariate Gaussian distributions and have an isofactorial representation with the Hermite polynomials as the factors: for $u, v \in G, u \neq v$, the probability density function of $(Y(u), Y(v))$ can be expanded as*

$$g_{u,v}(y,y') = g(y)g(y')\sum_{\nu=0}^{\infty}\mathbb{E}\left[\rho(u,v)^\nu\right]H_\nu(y)H_\nu(y'), \quad y,y' \in \mathbb{R}, \tag{5.10}$$

where $\rho(u,v) := \cos(WZ\sqrt{d(u,v)})$, Z is a standard Gaussian random variable independent of W, H_ν is the normalized Hermite polynomial of degree ν and g is the standard Gaussian probability density function.

Proof. Owing to the Box-Muller method, one can rewrite (5.7) as

$$Y(u) = Y_1\cos(Z_G(u)W) + Y_2\sin(Z_G(u)W), \quad u \in G,$$

where Y_1 and Y_2 are standard Gaussian random variables that are independent and independent of W and Z_G. Therefore, the distribution of $Y(u)$ (with $u \in G$) conditioned to W and Z_G has the following Fourier transform, where i is the imaginary unit and $\lambda \in \mathbb{R}$:

$$\mathbb{E}\left[e^{\mathrm{i}\lambda Y(u)} \mid W, Z_G\right] = \exp\left[-\frac{\lambda^2}{2}\left(\cos^2(Z_G(u)W) + \sin^2(Z_G(u)W)\right)\right] = e^{-\frac{\lambda^2}{2}}, \tag{5.11}$$

which does not depend on W or Z_G. Accordingly, the non-conditional distribution of $Y(u)$ is standard Gaussian, as its Fourier transform is the same as the last term

in (5.11), i.e., $\lambda \mapsto e^{-\frac{\lambda^2}{2}}$. As for the bivariate distributions, the distribution of $(Y(u), Y(v))$ (with $u, v \in G$) conditionally on W and Z_G, for $\lambda, \eta \in \mathbb{R}$, has the following Fourier transform:

$$\mathbb{E}\left[e^{\mathrm{i}\lambda Y(u)+\mathrm{i}\eta Y(v)} \mid W, Z_G\right] = \exp\left(-\frac{\lambda^2+\eta^2}{2} - \lambda\eta\cos(Z_G(u)W - Z_G(v)W)\right),$$

where $Z := (Z_G(u) - Z_G(v))/\sqrt{d(u,v)}$ follows a standard Gaussian distribution and is independent of W. Hence, for $\lambda, \eta \in \mathbb{R}$, the non-conditional Fourier transform is

$$\begin{aligned}
&\mathbb{E}\left[e^{\mathrm{i}\lambda Y(u)+\mathrm{i}\eta Y(v)}\right] \\
&\quad = \mathbb{E}\left[\exp\left(-\frac{\lambda^2+\eta^2}{2} - \lambda\eta\cos(WZ\sqrt{d(u,v)})\right)\right] \\
&\quad = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\int_{\mathbb{R}} \exp\left(-\frac{\lambda^2+\eta^2}{2} - \lambda\eta\cos(\omega z\sqrt{d(u,v)})\right) e^{-\frac{z^2}{2}}\,\mathrm{d}z\,F(\mathrm{d}\omega),
\end{aligned}$$

with the double integral being convergent insofar as the first exponential in the integrand is upper bounded by 1. It is seen that the Fourier transform depends on u and v only through their mutual distance $d(u,v)$, which establishes that the bivariate distributions of Y are isotropic with respect to the resistance metric, and that it is a scale mixture of bivariate Gaussian Fourier transforms. Given that W is almost surely different from 0, the joint probability density function of $(Y(u), Y(v))$ exists if $u \neq v$ and can be written as:

$$g_{u,v}(y, y') = \mathbb{E}\left[\frac{1}{2\pi\sqrt{1-\rho(u,v)^2}}\exp\left(-\frac{y^2 - 2\rho(u,v)yy' + y'^2}{2(1-\rho(u,v)^2)}\right)\right], \quad y, y' \in \mathbb{R},$$

with $\rho(u,v) := \cos(WZ\sqrt{d(u,v)})$. The proof concludes by noting that the bivariate Gaussian distributions have an isofactorial representation with the Hermite polynomials as the factors (Chilès and Delfiner, 2012, p. 412), which leads to the announced expansion (5.10); the convergence is guaranteed because $\rho(u,v)$ almost surely belongs to the open interval $(-1, 1)$. Formulae 1.320.5, 1.320.7 and 3.896.4 of Gradshteyn and Ryzhik (2007) lead to:

$$\mathbb{E}\left[\rho(u,v)^\nu\right] = \begin{cases} \frac{2}{2^\nu}\mathbb{E}\left[\sum_{\kappa=0}^{\frac{\nu-1}{2}} \frac{\nu!}{\kappa!(\nu-\kappa)!}\exp\left(\frac{(\nu-2\kappa)^2W^2d(u,v)}{2}\right)\right] & \nu \text{ odd} \\ \frac{\nu!}{2^\nu(\nu/2)!^2} + \frac{2}{2^\nu}\mathbb{E}\left[\sum_{\kappa=0}^{\frac{\nu}{2}-1} \frac{\nu!}{\kappa!(\nu-\kappa)!}\exp\left(\frac{(\nu-2\kappa)^2W^2d(u,v)}{2}\right)\right] & \nu \text{ even,} \end{cases}$$

so that a sufficient condition for (5.10) to be expressed analytically is that the moment generating function of W^2 is defined on $\mathbb{R}$ and known (e.g., with the first entry of Table 5.1).

□

5.3.2 Second Construction (Poisson Dilution Model)

Consider the class of random fields on G given by

$$Y(u) := \sum_{x \in X} \epsilon(x) f\left(Z_G(u) - x\right), \qquad u \in G, \tag{5.12}$$

where f is real-valued square-integrable function on $\mathbb{R}$, X is a stationary Poisson point process in $\mathbb{R}$ with intensity $\theta = 1$ independent of Z_G, $\{\epsilon(x) : x \in \mathbb{R}\}$ are mutually independent weights, such that $\mathbb{P}(\epsilon(x) = 1) = \mathbb{P}(\epsilon(x) = -1) = 1/2$ for each $x \in \mathbb{R}$, and are independent of X and Z_G. Let ψ_f be the transitive covariogram of f, defined as

$$\psi_f(h) = \int_{\mathbb{R}} f(x+h) f(x) \mathrm{d}x, \quad h \in \mathbb{R}. \tag{5.13}$$

This construction adapts the dilution method from classical geostatistics (Lantuéjoul, 2002). The following proposition establishes the range of achievable correlation functions.

Proposition 5.3. *The random field defined in (5.12) has a zero mean and is isotropic with respect to the resistance metric. Its covariance function, denoted as C_Y, is given by*

$$C_Y\left(d(u,v)\right) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \psi_f\left(\sqrt{d(u,v)} z\right) \exp(-z^2/2) dz. \tag{5.14}$$

Proof. The random field clearly possesses a zero mean, insofar as $\mathbb{E}(\epsilon(x)) = 0$ for any $x \in X$ and $\epsilon(x)$ is independent of Z_G and X. Let $I_k = [c_k, d_k]$ denote a sequence of closed intervals in $\mathbb{R}$, such that $I_k \subset I_{k+1}$ and $\cup_{k \in \mathbb{N}} I_k = \mathbb{R}$. The length of I_k is denoted by $|I_k|$. We first analyze (5.12), by restricting our attention to $X \cap I_k$. We proceed to calculate the covariance function of the random field

$$Y_k(u) := \sum_{x \in X \cap I_k} \epsilon(x) f\left(Z_G(u) - x\right), \qquad u \in G, k \in \mathbb{N}.$$

By conditioning on the number and location of points of the Poisson process and utilisingthe independence of the weights, we obtain

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}\left(Y_k(u), Y_k(v)\right) \\
&= \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{\exp(-|I_k|)|I_k|^n}{n!} \frac{n}{|I_k|} \int_{I_k} f\left(Z_G(u) - x\right) f\left(Z_G(v) - x\right) \mathrm{d}x \right) \\
&= \mathbb{E}\left(\sum_{n=1}^{\infty} \frac{\exp(-|I_k|)|I_k|^n}{n!} \frac{n}{|I_k|} \int_{Z_G(v)-d_k}^{Z_G(v)-c_k} f\left(x + Z_G(u) - Z_G(v)\right) f(x) \mathrm{d}x \right) \\
&= \mathbb{E}\left(\int_{Z_G(v)-d_k}^{Z_G(v)-c_k} f\left(x + Z_G(u) - Z_G(v)\right) f(x) \mathrm{d}x \right).
\end{aligned}$$

As k approaches infinity, the result expands to $\mathbb{R}$; therefore, we obtain

$$\mathbb{C}\mathrm{ov}\left((Y(u), Y(v)\right) = \begin{cases} \psi_f(0) \text{ if } u = v \\ \mathbb{E}\left[\psi_f\left(\sqrt{d(u,v)}Z\right)\right] \text{ otherwise,} \end{cases}$$

where $Z := (Z_G(u) - Z_G(v))/\sqrt{d(u,v)}$, which follows a standard Gaussian distribution.

□

In Table 5.2, we provide a list of parametric families of correlation functions on G, for different choices of the function f. The entries of the table have been established by using formulae 3.321.2, 3.321.3, 3.321.4, 3.322.1 and 6.511.13 of Gradshteyn and Ryzhik (2007) and formulae 2.19, 7.27 and 7.28 of Chilès and Delfiner (2012).

Table 5.2: Parametric correlation functions on G in terms of the resistance metric, for different choices of the dilution function $f : \mathbb{R} \to \mathbb{R}$. For each model, $a > 0$ is a scale parameter.

$f(t)$	C_Y
$a^{-1/2}1_{\lvert t\rvert \le a/2}$	$C_Y(d) = \mathrm{erf}\left(\frac{a}{a\sqrt{2d}}\right) - \frac{2-2\exp(-(a^2/2)/\sqrt{d})}{\sqrt{2\pi}}\sqrt{d}$
$\exp(-a^2t^2)(2/\pi)^{1/4}\sqrt{a}$	$C_Y(d) = (1 + a^2d)^{-1/2}$
$K_0(a\lvert t\rvert)\sqrt{2a}/\pi$	$C_Y(d) = \exp(a^2d/2)(1 - \mathrm{erf}(a\sqrt{d/2}))$

Remark 5.1. Unlike the conventional Euclidean case, this construction does not offer correlation functions with a compact support. Essentially, we take an additional expectation with respect to a Gaussian probability function, which prevents the occurrence of compactly supported models.

5.3.3 Third Construction (Dilution of a Random Germ)

Consider the class of random fields on G given by

$$Y(u) := \frac{\epsilon f\left(Z_G(u) - X\right)}{\sqrt{\varpi(X)}}, \qquad u \in G, \tag{5.15}$$

where ϵ is a Rademacher random variable, X is a random variable with a positive probability density function ϖ on $\mathbb{R}$, f is real-valued square-integrable function on $\mathbb{R}$, and (Z_G, ϵ, X) are mutually independent. Let ψ_f be the transitive covariogram of f, as per (5.13).

Proposition 5.4. *The random field defined in (5.15) has a zero mean, is isotropic with respect to the resistance metric, and its covariance function C_Y is given by (5.14).*

Proof. The zero mean of $Y(u)$ for all $u \in G$ stems from the zero mean of ϵ and the fact that ϵ is independent of X and Z_G. As for the covariance, one has:

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left(Y(u), Y(v)\right) &= \mathbb{E}\left(\int_{\mathbb{R}} f\left(Z_G(u) - x\right) f\left(Z_G(v) - x\right) \mathrm{d}x\right) \\ &= \mathbb{E}\left(\int_{\mathbb{R}} f\left(x + Z_G(u) - Z_G(v)\right) f(x) \mathrm{d}x\right) \\ &= \begin{cases} \psi_f(0) \text{ if } u = v \\ \mathbb{E}\left[\psi_f\left(\sqrt{d(u,v)}Z\right)\right] \text{ otherwise,} \end{cases}\end{aligned}$$

where $Z := (Z_G(u) - Z_G(v))/\sqrt{d(u,v)}$ follows a standard Gaussian distribution.

□

Remark 5.2. The class of covariance functions attainable with this construction is the same as the class obtained by the Poisson dilution model. With this version, however, we anticipate a simplified algorithm, as it does not require simulating a Poisson point process, which may involve aggregating many points and render its implementation less efficient.

5.4 Simulation Algorithms

5.4.1 Simulation of Z_G

The auxiliary random field Z_G is a key mathematical ingredient in the three constructions presented in the previous section. We employ the following steps to simulate this field:

Step 1. We first simulate Z_G on the vertices of the graph $v_1, \ldots, v_n$. It consists of simulating a Gaussian random vector $\boldsymbol{Z}_G^S$ with zero mean and covariance matrix $\mathbf{L}^{-1}$ (the inverse of the Laplacian matrix). This step is performed by employing the Cholesky factorization of the covariance matrix, $\mathbf{L}^{-1} = \boldsymbol{\Omega}^\top \boldsymbol{\Omega}$. The simulated values are obtained in the following way

$$\left(Z_G^S(v_1), \ldots, Z_G^S(v_n)\right)^\top = \boldsymbol{\Omega}^\top \boldsymbol{Z}_0,$$

with $\boldsymbol{Z}_0$ an n-dimensional vector with uncorrelated standard normal random variables.

Step 2. We extend the simulation on the vertices to locations on the whole graph. This is performed through linear interpolation (see Equation (5.3)).

Step 3. Finally, at each edge e, we add an independent Brownian bridge B_e^S, i.e., a Gaussian random field having a linear semi-variogram with slope $1/2$ and

conditioned to $B_e^S(\underline{e}) = B_e^S(\overline{e}) = 0$. To this end, we discretise$[\underline{e}, \overline{e}]$ into a set of points $\underline{e} = \varepsilon_0 < \varepsilon_1 < \ldots < \varepsilon_{K_e} = \overline{e}$ and use a sequential approach to take advantage of the Markov property of the Brownian bridge (Chilès and Delfiner, 2012, p. 492): for $k \in [1, K_e - 1]$, $B_e^S(\varepsilon_k)$ is a Gaussian random variable with mean equal to the ordinary kriging prediction from the leftmost and rightmost adjacent values $B_e^S(\varepsilon_{k-1})$ and $B_e^S(\overline{e})$, and variance equal to the corresponding ordinary kriging variance. This yields

$$\begin{aligned} Z_e^S(u_k) &= B_e^S(\varepsilon_k) \\ &= \left[\frac{1}{2} + \frac{\overline{e} - 2\varepsilon_k + \varepsilon_{k-1}}{2(\overline{e} - \varepsilon_{k-1})}\right] B_e^S(\varepsilon_{k-1}) + \xi_k \sqrt{\frac{\overline{e} - \varepsilon_{k-1}}{4} - \frac{(\overline{e} - 2\varepsilon_k + \varepsilon_{k-1})^2}{4(\overline{e} - \varepsilon_{k-1})}}, \end{aligned}$$

with ξ_k a standard Gaussian random variable independent of $\xi_1, \ldots, \xi_{k-1}$, and $\{u_k = \varphi_e^{-1}(\varepsilon_k) : k = 1, \ldots, K_e - 1\}$ is a set of points discretising e.

Remark 5.3. Z_G^S is a Gaussian random field with zero mean and variogram given by the resistance metric, as per (5.1). It is an extension to the graph endowed with the resistance metric of the Wiener-Lévy process on Euclidean spaces endowed with the Euclidean distance.

5.4.2 Algorithm 1 (Spectral Method)

The first algorithm relies on the construction (5.5), which allows us to obtain random fields with completely monotone correlation structures (Proposition 5.1). However, these random fields are clearly non-Gaussian, although their univariate distributions are standard Gaussian for any $u \in G$ (Proposition 5.2). In order to obtain an approximately Gaussian random field, we consider an additive combination of several independent copies of (5.5), as detailed in Algorithm 1, based on the central limit theorem (Lantuéjoul, 2002).

5.4.3 Algorithm 2 (Poisson Dilution)

The second algorithm is based on the construction (5.12), which allows us to obtain random fields with a variety of correlation structures (Proposition 5.3). Once selected the dilution function f, a direct application of (5.12) requires defining a sufficiently large interval I to simulate the Poisson point process $X \cap I$, which yields Algorithm 2 below.

Since I is bounded, Algorithm 2 is approximate, i.e., it does not exactly reproduce the target covariance structure, unless f is compactly supported, as in the first entry of Table 5.2. As for the spectral simulation, a central limit approximation is needed to obtain a random field with finite-dimensional distributions close to multivariate normal.

Algorithm 1 Spectral simulation. The simulated random variables $W_1, \ldots, W_M$, $V_1, \ldots, V_M$, and $\Lambda_1, \ldots, \Lambda_M$ are mutually independent and independent of $Z_G^{(1)}, \ldots, Z_G^{(M)}$

Require: A large integer M
Require: A spectral measure F
Require: M independent copies $Z_G^{(1)}, \ldots, Z_G^{(M)}$ of the auxiliary random field Z_G^S simulated on a fine grid $\{u_p : p = 1, \ldots, P\}$ discretising$G$

Simulate $W_1, \ldots, W_M \sim F$
Simulate $V_1, \ldots, V_M \sim \text{Unif}(0, 1)$
Simulate $\Lambda_1, \ldots, \Lambda_M \sim \text{Unif}(0, 2\pi)$
for $p = 1$ to P **do**
    Calculate $Y^S(u_p) = \sum_{m=1}^{M} \sqrt{\frac{-2\ln(V_m)}{M}} \cos\left(W_m Z_G^{(m)}(u_p) + \Lambda_m\right)$
end for

Algorithm 2 Poisson dilution. The simulated random variables $(N, x_1, \ldots, x_N, \epsilon_1, \ldots, \epsilon_N)$ are mutually independent and independent of Z_G^S

Require: A bounded interval $I \subset \mathbb{R}$
Require: A dilution function f
Require: An auxiliary random field Z_G^S simulated on a fine grid $\{u_p : p = 1, \ldots, P\}$ discretising$G$

Simulate $N \sim \text{Poisson}(|I|)$
if $N > 0$ **then**
    Simulate $x_1, \ldots, x_N \sim \text{Unif}(I)$
    Simulate $\epsilon_1, \ldots, \epsilon_N \sim$ Rademacher
    for $p = 1$ to P **do**
        Calculate $Y^S(u_p) = \sum_{n=0}^{N} \epsilon_n f(Z_G^S(u_p) - x_n)$
    end for
else
    Deliver $Y^S(u_p) = 0$ for $p = 1, \ldots, P$
end if

5.4.4 Algorithm 3 (Dilution of a Random Germ)

Algorithm 3 is a direct application of the third proposal in Section 5.3.3 combined with a central limit approximation. In practice, the importance sampling density ϖ should be chosen in such a way that x has a high probability to be in, or close to, the target simulation domain, and a low probability to be far from this domain.

Algorithm 3 Dilution of a random germ. The simulated random variables $(x_1, \ldots, x_M, \epsilon_1, \ldots, \epsilon_M)$ are mutually independent and independent of $Z_G^{(1)}, \ldots, Z_G^{(M)}$

Require: An importance sampling density ϖ supported in $\mathbb{R}$
Require: A dilution function f
Require: A large integer M
Require: M independent copies $Z_G^{(1)}, \ldots, Z_G^{(M)}$ of the auxiliary random field Z_G^S simulated on a fine grid $\{u_p : p = 1, \ldots, P\}$ discretising$G$

1: Simulate $x_1, \ldots, x_M \sim \varpi$
2: Simulate $\epsilon_1, \ldots, \epsilon_M \sim$ Rademacher
3: **for** $p = 1$ to P **do**
4: Calculate $Y^S(u_p) = \sum_{m=1}^{M} \frac{\epsilon_m}{\sqrt{M\varpi(x_m)}} f(Z_G^S(u_p) - x_m)$
5: **end for**

5.4.5 Computational Aspects

Simulation of the auxiliary field Z_G^S

The range of applicability of the algorithm presented in Section 5.4.1 is extensive, capable of efficiently simulating large graphs. Indeed:

1. The Cholesky factorization in Step 1, which has a cubic computational complexity in terms of the number of intersections or vertices, remains computationally feasible for graphs with tens of thousands of intersections. This can translate to hundreds of thousands or even millions of edges across the entire graph.

2. Although multiple independent copies of Z_G^S must be simulated, the Cholesky decomposition of $\mathbf{L}^{-1}$ needs to be performed only once, insofar as the graph's structure, and consequently the matrix $\mathbf{L}^{-1}$, remains constant throughout.

3. The extension to the grid discretising the graph (Steps 2 and 3) is straightforward and computationally cheap, even when considering a dense grid (i.e., a large K_e).

For graphs with a very large number of vertices (say, hundreds of thousands or more), an alternative to the Cholesky factorisation in Step 1 is the Gibbs sampler (Chilès and Delfiner, 2012, Section 7.6.3). This algorithm is iterative and, at each iteration, updates the simulated value of Z_G at a vertex conditionally to the values of Z_G at the remaining vertices. The updating relies on the inverse of the

variance-covariance matrix of $(Z_G(v_1), \ldots, Z_G(v_n))^\top$, which in the present case is just the Laplacian matrix $\mathbf{L}$. Therefore, there is no need for any matrix inversion or Cholesky factorization, which makes this alternative applicable to practically any graph, irrespective of its number of vertices: the numerical complexity is $\mathcal{O}(n^2)$ and the memory footprint is $\mathcal{O}(n)$ if the required components of the Laplacian matrix are calculated 'on the fly', while these figures are $\mathcal{O}(n^3)$ and $\mathcal{O}(n^2)$, respectively, for the Cholesky factorization. The inconvenience is that the simulation becomes approximate, as the Gibbs sampler must be stopped after finitely many loops over the set of vertices. Different blocking and updating strategies can improve the convergence of the sampler to the target distribution (Galli and Gao, 2001).

Simulation of the target random field Y^S on the graph

Once the auxiliary field is simulated, the computational complexity of Algorithms 1 to 3 depends on the parameters proper to each algorithm and on the desired quality for the central limit approximation, see next section and Table 5.3. Besides, as none of the three algorithms requires keeping in memory the simulated random field Y^S at more than one location at a time, the only significant memory requirement is that of the Cholesky factorisation, which is $\mathcal{O}(n^2)$.

Importantly, Algorithms 1 and 3 reproduce the covariance function of the random field exactly. The only approximation arises from the central limit approximation, whose convergence rate can be assessed numerically, see Section 5.5.4. In contrast, Algorithm 2 provides an approximate reproduction of the covariance function and is computationally more demanding than the others, which is confirmed in the next section.

Table 5.3: Comparison of algorithms for simulating random fields on graphs. Here, n is the number of vertices, P represents the total number of points along the graph's edges, M is the number of independent copies used in the central limit approximation, and $|I|$ is the length of the interval over which a Poisson process is simulated in the implementation of Algorithm 2.

	Algorithm 1	Algorithm 2	Algorithm 3
Numerical complexity	$\mathcal{O}(PM)+\mathcal{O}(n^3)$	$\mathcal{O}(PM\|I\|)+\mathcal{O}(n^3)$	$\mathcal{O}(PM)+\mathcal{O}(n^3)$
Memory footprint	$\mathcal{O}(n^2)$	$\mathcal{O}(n^2)$	$\mathcal{O}(n^2)$
Covariance reproduction	Exact	Approximate	Exact
Restriction on covariance	Eq. (5.6)	Eq. (5.14)	Eq. (5.14)

5.5 Applications

5.5.1 Implementation Parameters

We apply our algorithms to the network of the University of Chicago neighbourhood (Baddeley et al., 2021), which consists of 338 vertices and 503 edges. For Algorithm 1, we consider the first entry of Table 5.1 (an exponential correlation), whereas for Algorithms 2 and 3, we consider the second entry of Table 5.2 (a Cauchy correlation). These correlation functions are common choices in geostatistics. The scale parameter a is set to 0.2 and M is set to 1000 for the central limit approximation. In particular, a was chosen so that the practical range is a fraction of the maximum distance in the network, ensuring realistic realisations. In Algorithm 2, we use the interval $I = [-50, 50]$, to achieve good coverage, while Algorithm 3 is implemented with a standard Cauchy importance sampling density, favoring the occurrence of a wide range of values, attributed to its heavy tails. Figure 5.1 shows realisations over 100,600 locations (200 per edge) obtained by each algorithm.

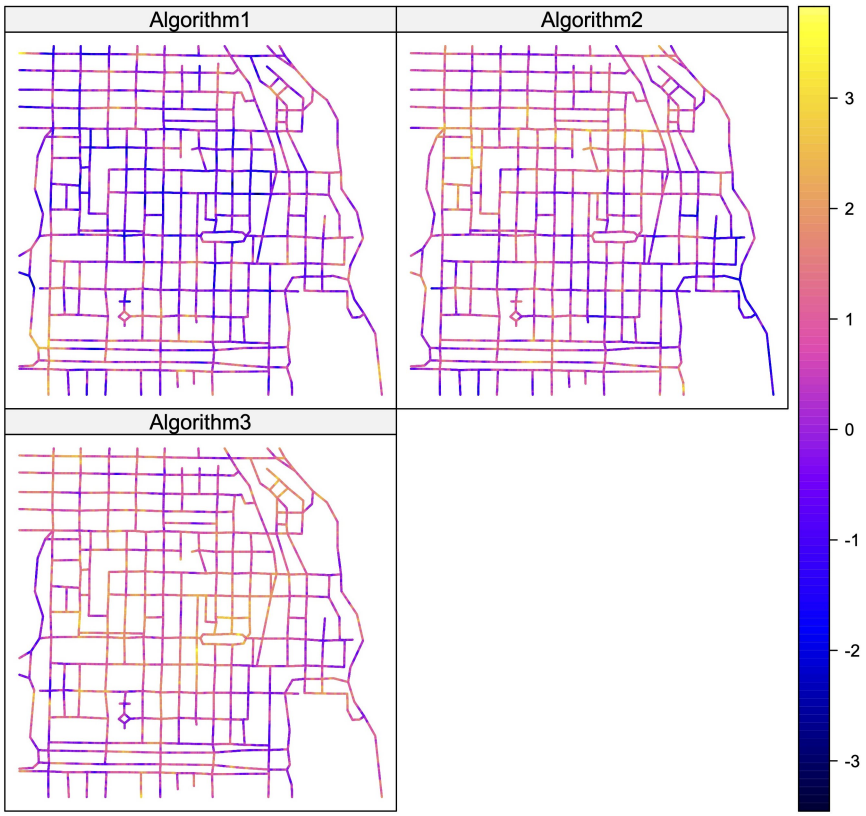


Figure 5.1: Realisations of random fields on a graph with 503 edges discretized into 200 points (totaling 100,600 locations). Algorithm 1 considers an exponential correlation function, while Algorithms 2 and 3 consider a Cauchy correlation function.

5.5.2 Execution Times

To assess the numerical complexity of the algorithms with respect to the number of target points over the network, we discretise each edge into 2^k points, with $k = 5, 6, \ldots, 10$. As a result (Table 5.4), Algorithms 1 and 3 exhibit comparable execution times, while Algorithm 2 is slower because the interval $I = [-50, 50]$, which is necessary for a good coverage, implies the generation of a Poisson process with a large number N of points ($\mathbb{E}(N) = |I| = 100$). The reported times correspond to the total computation time, including both the simulation of the auxiliary field Z_G^S and the time required by each algorithm.

Number of target points	16,096	32,192	64,384	128,768	257,536	515,072
Algorithm 1	1.837	2.762	5.431	10.61	21.49	48.86
Algorithm 2	29.75	66.36	134.8	274.1	604.1	1263
Algorithm 3	1.408	2.396	4.431	8.701	17.91	42.49

Table 5.4: Computing times (in seconds) as a function of the number of target points over the network, with $M = 1{,}000$ in the central limit approximation. The simulation was performed on an Apple M1 Pro chip (10-core CPU) with 16GB of memory.

To gain further insight into the behaviour of the computing times, Table 5.5 reports the results obtained for Algorithm 1 when simulations are carried out on networks generated through Delaunay triangulations. We consider networks with the number of vertices ranging from 2,500 to 7,500; the triangulations induced by these vertices yield between 7,475 and 22,468 edges. On each edge we place 45 target sites, so that in the largest case the simulation involves a grid of more than one million points. The reported computing times correspond to $M = 1{,}000$ in the central limit approximation. The results are presented separately for the auxiliary field and for the primary field. It is clear that the main computational burden is associated with simulating the auxiliary random field defined on the vertices, as already discussed, while the addition of an increasing number of target points along the edges produces only a gradual increase in the total computing time. This observation is consistent with the computational order analysis presented in the previous section. A visual illustration of the case with 7,500 vertices, along with additional details on computational aspects, is provided in Section 5.6.

5.5.3 Correlation Structure Reproduction

To investigate the reproduction of the target spatial dependency structure, for each algorithm, we construct $10,000$ realisations. We consider a subset of the graph (see left panel in Figure 5.3) and a set of locations discretisingit (15 points per edge, i.e., 330 points in total), with $M = 3,000$ and the same random field models and parameters as in the previous subsections, and compute their empirical semi-variograms at specific lags. It is seen (first row in Figure 5.2) that, on average, the

Table 5.5: Computing times (in seconds) as a function of the number of vertices in the network, generated via Delaunay triangulation. For each case, the corresponding number of edges is reported, along with the total number of target points obtained by assigning 45 target points per edge. The results are based on Algorithm 1 and a central limit approximation with $M = 1{,}000$. The simulation was performed on an Apple M1 Pro chip (10-core CPU) with 16GB of memory.

			Computing times	
# vertices	# edges	# target points	Auxiliary field	Primary field
2,500	7,475	336,375	33.71	1.33
3,500	10,470	471,150	82.46	1.87
4,500	13,474	606,330	140.4	2.70
5,500	16,470	741,150	218.1	2.93
6,500	19,472	876,240	321.7	3.29
7,500	22,468	1,011,060	456.0	4.08

theoretical semi-variogram γ_2 is reproduced without any bias, a conclusion that is confirmed by statistical testing (Table 5.6) (Emery, 2008).

The same conclusions prevail when looking at the semi-madograms of the simulated random fields (second row in Figure 5.2 and Table 5.6): on average over the $10,000$ realisations, the experimental semi-madograms match the expected model and no bias is perceptible. Recall that, for a Gaussian random field, the semi-madogram γ_1 is given by (Emery, 2005)

$$\gamma_1\sqrt{\pi} = \sqrt{\gamma_2} = \sqrt{C_Y(0) - C_Y},$$

with $\gamma_1(d(u,v)) = \frac{1}{2}\mathbb{E}\left[\mid Y(u) - Y(v) \mid\right]$ and $\gamma_2(d(u,v)) = \frac{1}{2}\mathbb{E}\big[\,(Y(u) - Y(v))^2\,\big]$, $u, v \in G$.

In passing, the fact that the semi-madogram is well reproduced indicate that the bivariate distributions of the simulated random fields are (close to) bi-Gaussian, i.e., the central limit approximation is accurate.

5.5.4 Assessment of Central Limit Approximation

To end the case study, we explore the accuracy of the multivariate normal approximation in terms of the number M of independent copies in the central limit approximation. Both analytical and numerical approaches have been developed in the literature to this end. However, the former (e.g., Lantuéjoul, 2002) are impracticable in our context, due to the difficulty in deriving an analytical expression of the finite-dimensional distributions of the simulated random fields, or a Berry-Esséen bound of the error between the true and simulated distributions. Here, we opt for a

		p-value of T-statistics					
		Lag 20	Lag 40	Lag 60	Lag 80	Lag 100	Lag 120
s-variogram	Algorithm 1	0.427	0.937	0.943	0.696	0.665	0.897
	Algorithm 2	0.641	0.290	0.502	0.768	0.432	0.304
	Algorithm 3	0.456	0.752	0.765	0.703	0.359	0.606
s-madogram	Algorithm 1	0.259	0.709	0.858	0.530	0.455	0.864
	Algorithm 2	0.541	0.196	0.425	0.973	0.476	0.267
	Algorithm 3	0.823	0.837	0.991	0.873	0.386	0.592

Table 5.6: Student test on empirical semi-variograms and semi-madograms at six lag distances, for $10,000$ realisations of the simulated random field. With a significance level of 0.05, the null hypothesis that the average empirical semi-variogram (semi-madogram) matches the theoretical semi-variogram (semi-madogram) is not rejected in all the cases.

numerical validation based on statistical testing and inspired from Arroyo and Emery (2020), as detailed hereinafter.

Given n locations on the graph, we study the distribution of a linear combination of the values simulated at these locations. To simplify the analysis, let us consider a subset of the University of Chicago neighbourhood. We consider two scenarios: $n = 2$ (triangles in Figure 5.3) and $n = 5$ (squares in Figure 5.3) fixed locations. Thus, in each scenario, we consider the zero-mean linear combination $\lambda_1 Y(u_1) + \cdots + \lambda_n Y(u_n)$, where the weights $\lambda_1, \ldots, \lambda_n$ are simulated from a uniform distribution on $[-10, 10]$. These weights are fixed across this experiment, therefore one can calculate the variance of the linear combination once the covariance function of Y is known.

To assess multivariate normality, we proceed as follows. We simulate 100 independent linear combinations, as detailed above, and use the Shapiro-Wilk goodness-of-fit test, with a certain significance level $\alpha \in (0, 1)$, to check the normal assumption. By repeating this procedure 100 times, we can calculate the proportion of rejections and compare this value with the nominal level α. Figure 5.4 shows this comparison for several values of α, and considering each algorithm with $M = 50$, 100 and 500. We also report a 95% confidence interval for the observed proportions of rejected tests, which is constructed by taking into account that, for each α, the number of rejections follows a binomial distribution with size 100 and success probability α.

Both algorithms 1 and 2 offer robust Gaussian approximations even with $M = 50$, as the proportion of rejections almost always lies in-between the 95% confidence bounds derived from the binomial distribution. Conversely, Algorithm 3 requires raising M to achieve a more accurate performance, as the linear combinations with $n = 5$ appear to significantly deviate from normality. This experiment reveals that, for the network under consideration, M of the order of a few hundreds is sufficient to achieve a good multivariate-Gaussian approximation across all algorithms.

This kind of goodness-of-fit testing is deemed useful to determine a suitable

value for M, especially for Algorithms 2 and 3 where the quality of the multivariate-Gaussian approximation may strongly depend on the support of the dilution function and on the network geometry; in contrast, for Algorithm 1, the univariate distribution of the simulated random field is always Gaussian (Proposition 5.2), thus multivariate normality is easier to reach.

5.5.5 Visual Illustration of a Point Pattern

Our findings offer valuable insights into simulating point pattern data on networks, with a particular emphasis on log-Gaussian Cox processes, where the logarithm of the intensity function is governed by a Gaussian random field (see Baddeley et al. (2021) and references therein for a comprehensive overview of potential applications).

To illustrate these insights, consider the intensity random field

$$\Lambda = \rho \exp\left(\sigma Y - \sigma^2/2\right),$$

where Y is a zero-mean Gaussian random field on G, and ρ and σ^2 are positive parameters. By simulating Y, and consequently Λ, with our algorithms, we can obtain a realisation of a Cox process on G driven by Λ. Notice that, under this choice, the intensity random field has a constant expectation ($\mathbb{E}(\Lambda(u)) = \rho$, for all $u \in G$), so that the simulated point process is first-order homogeneous. Also, recall that our code is designed for integration with the architectures of well-established `R` packages, including `spastat` and `coxln`, therefore facilitating the simulation of Cox processes with our algorithms as a prior step.

Figure 5.5 shows an example with $\sigma^2 = 1$ and $\rho = 0.001$. The Gaussian random field Y was simulated using Algorithm 1, with an exponential correlation structure with scale parameter $a = 0.1$. The values of the corresponding realisation of Λ are represented by the varying widths of the grey-coloured segments on the network: larger widths indicate a higher propensity of point occurrence.

In addition to the log-Gaussian Cox process, our algorithms could be seamlessly integrated with other point process models, such as the interrupted Cox process or the permanental Cox process (Møller and Rasmussen, 2024, Section 5.2). Moreover, our approach could complement the advancements outlined in Cronie et al. (2020), where the concept of pseudo-stationarity for point processes on linear networks is introduced.

5.6 Conclusions

The algorithms presented in this study allows the simulation of Gaussian random fields on graphs with Euclidean edges, being isotropic with respect to the resistance metric. In developing these algorithms, we have introduced interesting stochastic representations that offer an alternative path to embeddings in Hilbert spaces for obtaining valid classes of covariance functions on these complex domains.

In particular, Algorithm 1, which allows us to simulate random fields with completely monotone covariance functions, represents a graph-like version of classical spectral algorithms to simulate random fields indexed by Euclidean (Lantuéjoul, 2002) or spherical (Alegría et al., 2020) coordinates. Algorithms 2 and 3, which adapt the dilution principle detailed in Lantuéjoul (2002), allow us to achieve a wide range of covariance functions; however, Algorithm 2 is approximate, unless the dilution function is compactly supported, whereas Algorithm 3, which employs an importance sampling strategy, reproduces the target covariance structure exactly irrespective of the dilution function under consideration. Note that the latter algorithm can be adapted to random fields defined in Euclidean domains as an alternative to the Poisson dilution method when the dilution function is not compactly supported, avoiding biases in the reproduction of the covariance structure due to the restriction of the Poisson process to a bounded interval.

For all the algorithms, the numerical complexity is cubic in the number of vertices, plus an additional cost proportional to the number of locations targeted for simulation on the graph edges. For instance, with standard computer resources, Algorithms 1 and 3 generate realisations across half a million points on the Chicago graph in about 45 seconds. The simulation of an auxiliary random field related to the resistance metric results in additional computational burden. However, the algorithms remain notably fast and feasible compared to one-size-fits-all algorithms such as the classical Cholesky method, which is out of reach. Statistical testing indicates that combining a few hundred independent copies yields satisfactory Gaussian approximations.

To summarise the key takeaways from this work, Algorithms 2 and 3 address the same class of covariance functions; however, Algorithm 3 is preferred due to its greater efficiency, as Algorithm 2 is approximate and has higher computational complexity. Algorithms 1 and 3 exhibit the same computational complexity, reproduce the covariance function exactly, and demonstrate identical behaviour in the central limit approximation. The main difference lies in the families of covariance functions they cover. Therefore, Algorithms 1 and 3 complement each other effectively.

The proposed stochastic representations, and potential variations of them, also hold promise for developing novel covariance models and simulation algorithms in future investigations, particularly when addressing point processes on graphs, vector-valued random fields, and spatio-temporal random fields.

The conventional Matérn covariance model is not suitable for random fields on graphs, as the full range of smoothness is not compatible with this domain. Therefore, the search for a Matérn analogue remains an intriguing challenge. Bolin et al. (2024) address this issue by employing the SPDE approach. We believe that directly modelling the covariance through a parametric structure that incorporates the key features of the Matérn model could also be a promising avenue for further exploration.

The range of applicability of our algorithms is extensive, capable of simulating large urban areas with tens of thousands of intersections/vertices. While the number of streets/edges may be much higher, this does not significantly impact our approach,

as the computational cost mainly depends on the number of vertices, not edges. For example, on a network with 22,472 edges generated by the Delaunay triangulation of 7,500 uniformly distributed vertices on a rectangle, we can simulate a Gaussian random field over 1 million sites (Figure 5.6) in less than 10 minutes on a standard laptop, with Algorithm 1 and the technical specifications provided in Section 5.5. However, certain real-world problems of a considerably larger scale may require refinement of our approach, especially in the simulation step on the vertices and the manipulation of the Laplacian matrix. In Section 5.4.5, we mentioned a concrete strategy to address these challenges using Gibbs sampling. Additionally, alternative approaches, such as the SPDE method, offer valuable options for large-scale simulations.

Acknowledgments

This work was funded and supported by the National Agency for Research and Development of Chile [grants ANID Fondecyt 1250008 (A. Alegría and X. Emery) and ANID PIA AFB230001 (X. Emery)].

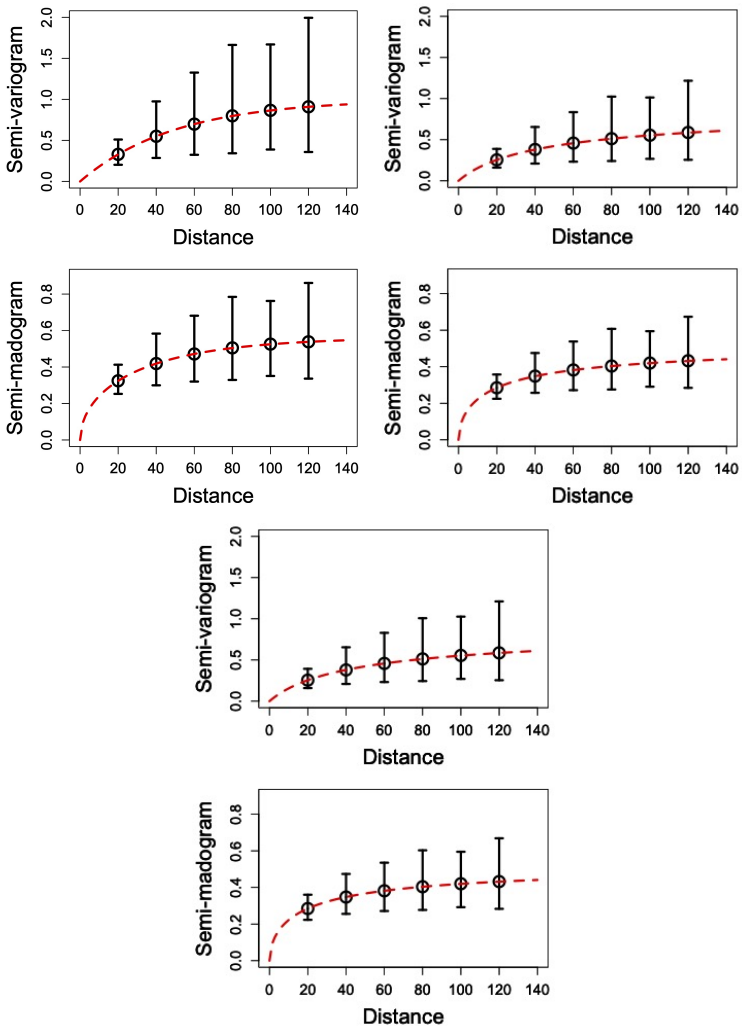


Figure 5.2: 95% confidence intervals (vertical segments) for the empirical semi-variograms (top) and semi-madograms (bottom) at specific lags, based on $10,000$ realisations of a random field, as a function of the resistance metric. The points represent the averages of the empirical semi-variograms and semi-madograms, while the theoretical curves are shown as dashed lines. From left to right, the results correspond to Algorithms 1 to 3, respectively.

Figure 5.3: (Left) The dashed square shows the zone of the University of Chicago neighbourhood employed in our experiment. (Right) Target locations within this zone.

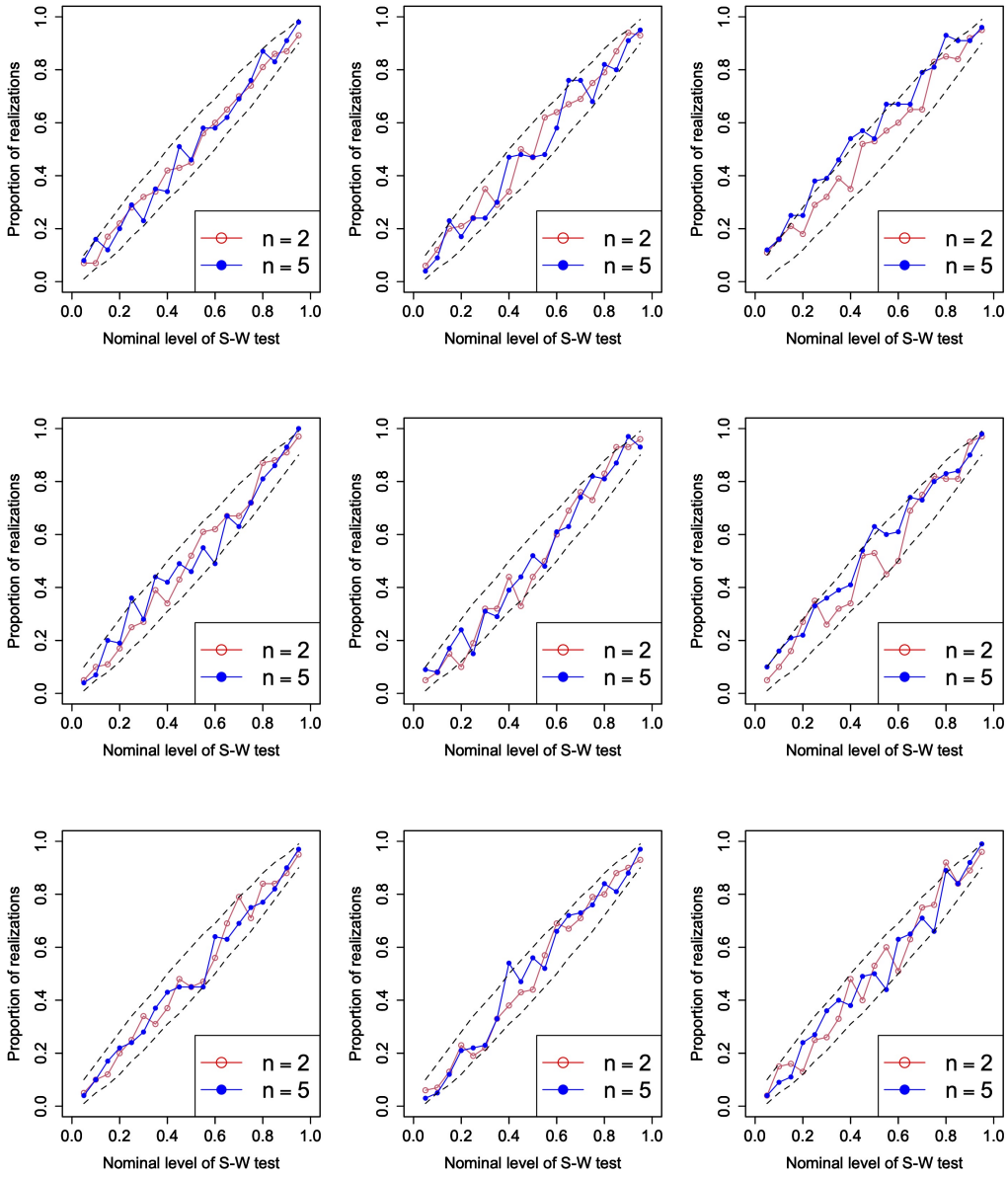


Figure 5.4: Probability-probability plots showing the proportion of rejected Shapiro-Wilk tests versus the nominal significance level of the test. From top to bottom, $M = 50$, 100 and 500. The dashed lines represent the 95% confidence bounds on the observed proportions of rejections. From left to right, the results correspond to Algorithms 1 to 3, respectively.

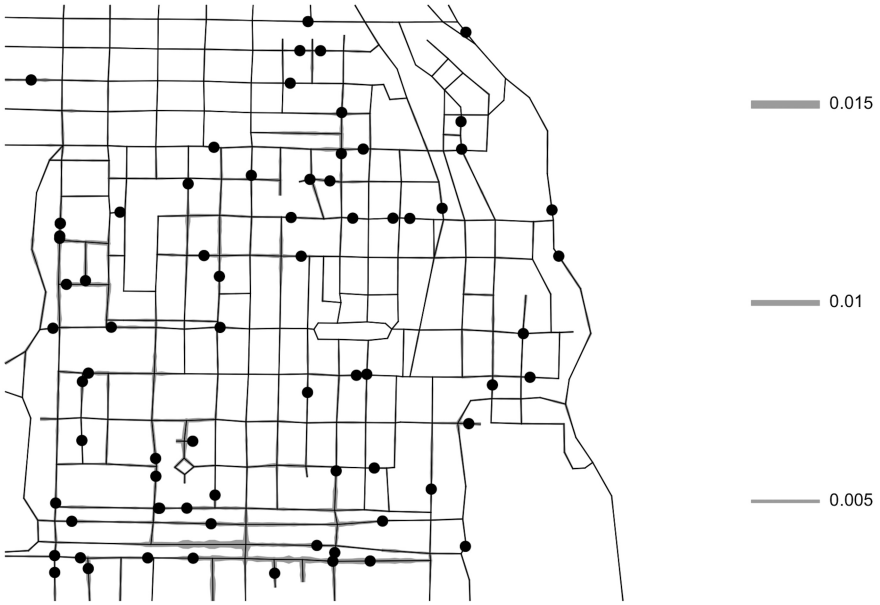


Figure 5.5: Simulation of a Cox process in a neighbourhood of the University of Chicago, based on a log-intensity random field (represented through the grey segments of varying widths) obtained with the proposed spectral algorithm.

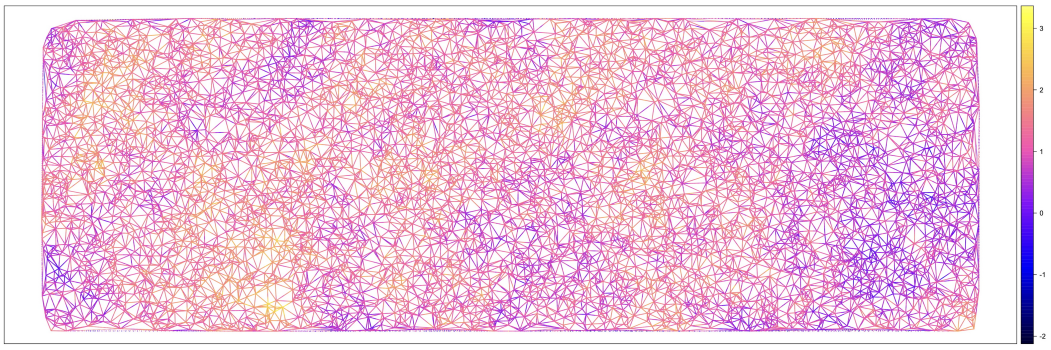


Figure 5.6: Realisation of a Gaussian random field over 1 million sites on a network consisting of 22,472 edges. We applied Algorithm 1 using an exponential correlation function with $a = 0.2$ and performed $M = 1,000$ repetitions in the central limit approximation.

Conclusion

In this thesis we have proposed solutions to some natural matters arising from Anderes et al. (2020), who represented a turning point in the modelling of geophysical networks. In Chapter 2, we have put metric graphs in a wider, structured context, and we have argued that this new topology may play a fundamental role in the development of climate science.

Chapter 3 has explored time-evolving graphs: we have formalised this notion and defined a semi-metric that extends the resistance distance of Anderes et al.. We have explored the properties of such a semi-metric: although it presents some limitations—for instance, it does not satisfy the triangle inequality, it is not invariant by the addition of degree-2 vertices—, it is, to our knowledge, one of the first attempts to formalise and manage time-evolution in graphs. We have then availed of the construction in Anderes et al. (2020) to define a class of (isotropic) covariance functions on metric graphs.

Chapter 4 has delved into multivariate processes, keeping the marginal construction of Anderes et al. (2020) and having a homogeneous structure. We have defined an *ad-hoc* multi-dimensional distance and we have shown that it enjoys several properties analogous to the ones satisfied by univariate distances. Again, we have made use of the same idea (Anderes et al., 2020) to define matrix-valued covariance functions.

Chapter 5, instead of extending the framework introduced by Anderes et al. (2020), has focused on efficient methods to simulate Gaussian processes on graphs with Euclidean edges. It explores three different algorithms and compares their computational complexities and statistical accuracy.

Even though the articles presented above have explored some branches stemming from the framework of metric graphs, it is worth underlying that the research field in question is quite new (yet evolving): there still remains much to be investigated. First, the constructions presented in Chapters 3 and 4 are still in they early stages. Indeed, both constructions are not invariant by addition and removal of degree-2 nodes. *A priori*, one could overlook this deficiency, or even benefit from the greater flexibility it implies, yet such a property would make distances and processes defined on graphs *intrinsic*. Furthermore, all the covariance functions defined on metric graphs via these methods have non-differentiable, albeit continuous, sample paths.

This is a limitation, mainly as it affects kriging estimation, but also from a theoretical perspective. At the moment, the only approach that has been able to handle such a problem is the one by Bolin et al.. However, such an approach has not been able to deal with multivariate or space-time scenarios yet. In addition, Chapter 5 has only treated the base setting of univariate processes on space-only graphs: it would be of interest to extend the presented algorithms to multivariate and space-time settings.

Moreover, the presented articles have not been tested on real-world datasets yet. The main reason is that we have faced many difficulties in finding datasets that could benefit from our approaches. For instance, time-evolving setting requires a dataset that satisfies the following conditions: 1) the graph has to show a significant time-evolving topology, 2) the topology has to be completely known and 3) data have to be continuously-indexed over the graph and follow, at least approximately, a Gaussian distribution. Regarding the multivariate approach, it may be easier to find fitting graph structures, but it still needs continuously-indexed data and approximate normality, in addition to multivariate observations.

Another key aspect is the estimation of model parameters. For instance, in time-evolving setting, it is essential to provide estimates of the parameters α, β, ρ and ϕ, which influence the resulting distance (and, as a consequence, the covariance functions). Similar problems arise for the multivariate setting. Last but not least, it is essential to provide valid methods for the estimation of the completely monotonic functions that, composed with the distance, give the covariance functions.

In conclusion, we have moved the first steps in the still-unexplored field of covariance functions on metric graphs. However, plenty is still to be explored and discovered: we leave the above-mentioned issues for future research.

Bibliography

Abel, D. The role of networked governance for local climate policy output. evidence from europe. *Environmental Politics*, 35(1):1–22, 2025.

Acuto, M. Covid-19: Lessons for an urban(izing) world. *One Earth*, 2(4):317–319, 2020.

Alegría, A., Emery, X., and Lantuéjoul, C. The turning arcs: A computationally efficient algorithm to simulate isotropic vector-valued Gaussian random fields on the d-sphere. *Statistics and Computing*, 30:1403–1418, 2020.

Alegría, A., Emery, X., Filosi, T., and Porcu, E. Computationally Efficient Algorithms for Simulating Isotropic Gaussian Random Fields on Graphs with Euclidean Edges. *Journal of Computational and Graphical Statistics*, 2025. Accepted for publication, arXiv preprint at arXiv.2404.17491.

Allard, D., Emery, X., Lacaux, C., and Lantuéjoul, C. Simulating space-time random fields with nonseparable Gneiting-type covariance functions. *Statistics and Computing*, 30(5): 1479–1495, 2020.

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

Anderes, E., Møller, J., and Rasmussen, J. G. Isotropic covariance functions on graphs and their edges. *Annals of Statistics*, 48(4):2478–2503, 2020.

Anis, A., Gadde, A., and Ortega, A. Efficient sampling set selection for bandlimited graph signals using graph spectral proxies. *IEEE Transactions on Signal Processing*, 64(14): 3775–3789, 2016. doi: 10.1109/TSP.2016.2546233.

Arroyo, D. and Emery, X. Algorithm 1013: An R implementation of a continuous spectral algorithm for simulating vector Gaussian random fields in Euclidean spaces. *ACM Transactions on Mathematical Software*, 47(1):8, 2020.

Baddeley, A. and Turner, R. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12:1–42, 2005.

Baddeley, A., Nair, G., Rakshit, S., and McSwiggan, G. Stationary point processes are uncommon on linear networks. *Stat*, 6(1):68–78, 2017.

Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., and Davies, T. M. Analysing point patterns on networks—A review. *Spatial Statistics*, 42:100435, 2021.

Bai, T. and Tahmasebi, P. Graph neural network for groundwater level forecasting. *Journal of Hydrology*, 616:128792, 11 2022. doi: 10.1016/j.jhydrol.2022.128792.

Barabási, A.-L. Scale-free networks: a decade and beyond. *Science*, 325(5939):412–413, 2009.

Barabási, A.-L. and Albert, R. Emergence of scaling in random networks. *Science*, 286(5439): 509–512, 1999.

Berezin, Y., Gozolchiani, A., Guez, O., and Havlin, S. Stability of climate networks with time. *Scientific reports*, 2(1):666, 2012.

Berg, C. Stieltjes-Pick-Bernstein-Schoenberg and their connection to complete monotonicity. *Positive Definite Functions: From Schoenberg to Space-Time Challenges*, pages 15–45, 2008.

Berlinet, A. and Thomas-Agnan, C. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer, 2004.

Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A., and Pedreschi, D. Foundations of multidimensional network analysis. *2011 International Conference on Advances in Social Networks Analysis and Mining*, pages 485–489, 2011.

Bernstein, S. Sur les fonctions absolument monotones. *Acta Mathematica*, 52:1–66, 1929.

Bevilacqua, M., Faouzi, T., Furrer, R., and Porcu, E. Estimation and prediction using generalized wendland covariance functions under fixed domain asymptotics. *The Annals of Statistics*, 47(2):828–856, 2019.

Board, F. S. Task force on climate-related financial disclosures. *Final Report: Recommendations of the Task Force on Climate-Related Financial Disclosures*, 2017.

Boccaletti, S., Bianconi, G., Herrero, R., Genio, C., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Z.Wang, and Zanin, M. The structure and dynamics of multilayer networks. *Physics Reports*, 544:1–122, 11 2014. doi: 10.1016/j.physrep.2014.07.001.

Bochner, S. *Harmonic Analysis and the Theory of Probability*. California Monographs in Mathematical Sciences. University of California Press, 1955.

Bocquet, M. and Sakov, P. Joint state and parameter estimation with an iterative ensemble kalman smoother. *Nonlinear Processes in Geophysics*, 20(5):803–818, 2013. doi: 10.5194/npg-20-803-2013.

Boers, N. Observation-based early-warning signals for a collapse of the atlantic meridional overturning circulation. *Nature Climate Change*, 11:680–688, 08 2021. doi: 10.1038/s41558-021-01097-4.

Boers, N., Goswami, B., Rheinwalt, A., Bookhagen, B., Hoskins, B., and Kurths, J. Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature*, 566, 02 2019. doi: 10.1038/s41586-018-0872-x.

Bolin, D. and Lindgren, F. Spatial Models Generated by Nested Stochastic Partial Differential Equations, with an Application to Global Ozone Mapping. *The Annals of Applied Statistics*, 5(1):523–550, 2011.

Bolin, D. and Wallin, J. Spatial confounding under infill asymptotics, 2024. arXiv preprint at arXiv.2403.18961.

Bolin, D., Kovács, M., and Kumar, V. Regularity and numerical approximation of fractional elliptic differential equations on compact metric graphs. *Mathematics of Computation*, 93, 02 2023a. doi: 10.1090/mcom/3929.

Bolin, D., Simas, A., and Wallin, J. Statistical inference for Gaussian Whittle-Matérn fields on metric graphs. *arXiv preprint arXiv:2304.10372*, 2023b.

Bolin, D., Simas, A. B., and Wallin, J. Gaussian Whittle–Matérn fields on metric graphs. *Bernoulli*, 30(2):1611–1639, May 2024. ISSN 1350-7265. doi: 10.3150/23-BEJ1647.

Bolin, D., Riera-Segura, L., and Simas, A. B. A new class of non-stationary Gaussian fields with general smoothness on metric graphs, 2025. arXiv preprint at arXiv.2501.11738.

Borovitskiy, V., Karimi, M. R., Somnath, V. R., and Krause, A. Isotropic gaussian processes on finite spaces of graphs. In Ruiz, F., Dy, J., and van de Meent, J.-W., editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 4556–4574. PMLR, 2023.

Bronstein, M. M. et al. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Cambanis, S., Keener, R., and Simons, G. On-symmetric multivariate distributions. *Journal of Multivariate Analysis*, 13(2):213–233, 1983.

Castruccio, S. and Guinness, J. An Evolutionary Spectrum Approach to Incorporate Large-scale Geographical Descriptors on Global Processes. *J. R. Statist. Soc. C*, 66(2):329–344, 2017.

Chami, I., Abu-El-Haija, S., Perozzi, B., Ré, C., and Murphy, K. Machine learning on graphs: A model and comprehensive taxonomy. *The Journal of Machine Learning Research*, 23(1): 3840–3903, 2022.

Cherifi, H., Palla, G., Szymanski, B., and Lu, X. On community structure in complex networks: challenges and opportunities. *Applied Network Science*, 4, 12 2019. doi: 10.1007/s41109-019-0238-9.

Chilès, J.-P. and Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*, volume 713. John Wiley & Sons, 2012.

Chishtie, F. A., Brunet, D., White, R. H., Michelson, D., Jiang, J., Lucas, V., Ruboonga, E., Imaash, S., Westland, M., Chui, T., et al. Advancing heatwave forecasting via distribution informed-graph neural networks (DI-GNNs): Integrating extreme value theory with GNNs. *arXiv preprint arXiv:2411.13496*, 2024.

Clarke, D. J. and Hill, R. V. *Cost-benefit analysis of the African risk capacity facility*, volume 1292. Intl Food Policy Res Inst, 2013.

Climate Technology Centre and Network. About CTCN. `https://www.ctc-n.org/about-ctcn`, 2024. Accessed June 2025.

Cramer, H. On the theory of stationary random processes. *Annals of Mathematics*, 41(1): 215–230, 1940. ISSN 0003486X.

Cressie, N. and Wikle, C. K. *Statistics for Spatio-Temporal Data*. Wiley Series in Probability and Statistics, 2011.

Cressie, N., Frey, J., Harch, B., and Smith, M. Spatial prediction on a river network. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(2):127, 2006.

Cronie, O., Moradi, M., and Mateu, J. Inhomogeneous higher-order summary statistics for point processes on linear networks. *Statistics and computing*, 30(5):1221–1239, 2020.

Daley, D. J., Porcu, E., and Bevilacqua, M. Classes of compactly supported covariance functions for multivariate random fields. *Stochastic Environmental Research and Risk Assessment*, 29:1249–1263, 2015.

Damaraju, E., Caprihan, A., Lowe, J. R., Allen, E. A., Calhoun, V. D., and Phillips, J. P. Functional connectivity in the developing brain: a longitudinal study from 4 to 9 months of age. *Neuroimage*, 84:169–180, 2014.

Dattani, S., Rodés-Guirao, L., Ritchie, H., Ortiz-Ospina, E., and Roser, M. Life expectancy. *Our World in Data*, 2023. URL `https://ourworldindata.org/life-expectancy`.

De Domenico, M. et al. Mathematical formulation of multilayer networks. *Physical Review X*, 3(4):041022, 2013.

de Prado, M. M. L. *Causal factor investing: can factor investing become scientific?* Cambridge University Press, 2023.

Defferrard, M., Bresson, X., and Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, volume 29 of *NeurIPS 2016*, pages 3844–3852. Curran Associates, Inc., 2016.

Deng, N., Zhou, W., and Haenggi, M. The ginibre point process as a model for wireless networks with repulsion. *IEEE Transactions on Wireless Communications*, 14(1):107–121, 2014.

Devriendt, K. Effective resistance is more than distance: Laplacians, simplices and the Schur complement. *Linear Algebra and its Applications*, 639:24–49, 2022.

Deza, J. I., Barreiro, M., and Masoller, C. Inferring interdependencies in climate networks constructed at inter-annual, intra-season and longer time scales. *The European Physical Journal Special Topics*, 222(2):511–523, 2013.

Divakaran, A. and Mohan, A. Temporal link prediction: A survey. *New Generation Computing*, 38(1):213–258, 2020. ISSN 0288-3635, 1882-7055. doi: 10.1007/s00354-019-00065-z.

Dong, X., Thanou, D., Toni, L., Bronstein, M., and Frossard, P. Graph signal processing for machine learning: A review and new perspectives. *IEEE Signal processing magazine*, 37 (6):117–127, 2020.

Donges, J. F., Zou, Y., Marwan, N., and Kurths, J. The backbone of the climate network. *Europhysics Letters*, 87(4):48007, 2009a.

Donges, J. F., Zou, Y., Marwan, N., and Kurths, J. Complex networks in climate dynamics. *The European Physical Journal Special Topics*, 174(1):157–179, 2009b.

Donges, J. F., Heitzig, J., Beronov, B., Wiedermann, M., Runge, J., Feng, Q. Y., Tupikina, L., Stolbova, V., Donner, R. V., Marwan, N., et al. Unified functional network and nonlinear time series analysis for complex systems science: The pyunicorn package. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(11), 2015.

Dörr, C. and Schlather, M. Characterization theorems for pseudo cross-variograms. *Journal of Applied Probability*, 60(4):1219–1231, 2023.

Emery, X. Variograms of order ω: A tool to validate a bivariate distribution model. *Mathematical Geology*, 37(2):163–181, 2005.

Emery, X. Statistical tests for validating geostatistical simulation algorithms. *Computers & Geosciences*, 34(11):1610–1620, 2008.

Emery, X. and Porcu, E. The schoenberg kernel and more flexible multivariate covariance models in euclidean spaces. *Computational and Applied Mathematics*, 42(4):148, 2023.

Emery, X., Porcu, E., and White, P. New validity conditions for the multivariate Matérn coregionalization model, with an application to exploration geochemistry. *Mathematical Geosciences*, 54(6):1043–1068, 2022.

Erdélyi, A. *Tables of Integral Transforms, Volume I.* McGraw-Hill, New York, 1954.

Evensen, G. *Data Assimilation: The Ensemble Kalman Filter*. Springer, Berlin, Heidelberg, 2 edition, 2009. ISBN 978-3-642-03710-8. doi: 10.1007/978-3-642-03711-5.

Feldhoff, J. H. et al. Geometric detection of coupling directions by means of inter-system recurrence networks. *Physics Letters A*, 376(4):3504–3513, 2012.

Filosi, T., Agostinelli, C., and Porcu, E. Temporally-evolving generalised networks and their kernels. *arXiv preprint arXiv:2309.15855*, 2025a.

Filosi, T., Porcu, E., Emery, X., Agostinelli, C., and Alegrìa, A. Vector-valued gaussian processes and their kernels on a class of metric graphs. *arXiv preprint arXiv.2501.10208*, 2025b.

Folkman, J. Role of angiogenesis in tumor growth and metastasis. *Seminars in Oncology*, 29 (6 Suppl 16):15–18, 2002.

Friedman, J., Hastie, T., and Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008. doi: 10.1093/biostatistics/kxm045.

Galli, A. and Gao, H. Rate of convergence of the Gibbs sampler in the Gaussian case. *Mathematical Geology*, 33(6):653–677, 2001.

Gardner, B., Sullivan, P. J., and Lembo, Jr, A. J. Predicting stream temperatures: geostatistical model comparison using alternative distance metrics. *Canadian Journal of Fisheries and Aquatic Sciences*, 60(3):344–351, 2003.

Georgopoulos, L. and Hasler, M. Distributed machine learning in networks by consensus. *Neurocomputing*, 124:2–12, 2014.

Ghesquiere, F., Mahul, O., Forni, M., and Gartley, R. Caribbean catastrophe risk insurance facility: A solution to the short-term liquidity needs of small island states in the aftermath of natural disasters. *World Bank, Washington DC*, 2006.

Ghosh, A., Boyd, S., and Saberi, A. Minimizing effective resistance of a graph. *SIAM Review*, 50(1):37–66, 2008.

Gneiting, T. On α-symmetric multivariate characteristic functions. *Journal of Multivariate Analysis*, 64(2):131–147, 1998.

Gneiting, T. Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, 97(458):590–600, 2002.

Gradshteyn, I. S. and Ryzhik, I. M. *Tables of Integrals, Series, and Products*. Academic Press, Amsterdam, seventh edition, 2007.

Haas, M., Goswami, B., and von Luxburg, U. Pitfalls of climate network construction—a statistical perspective. *Journal of Climate*, 36(10):3321–3342, 2023.

Hamilton, W. L., Ying, R., and Leskovec, J. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584*, 2017.

Hammond, D. K., Vandergheynst, P., and Gribonval, R. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011. doi: 10.1016/j.acha.2010.04.005.

Hanneke, S., Fu, W., and Xing, E. P. Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585 – 605, 2010. doi: 10.1214/09-EJS548.

Hellmuth, M. E., Osgood, D. E., Hess, U., Moorhead, A., and Bhojwani, H. *Index insurance and climate risk: Prospects for development and disaster management*. Columbia University Press, 2009.

Hlinka, J., Hartman, D., Vejmelka, M., Runge, J., Marwan, N., Kurths, J., and Paluš, M. Reliability of inference of directed climate networks using conditional mutual information. *Entropy*, 15(6):2023–2045, 2013.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Penna, S. D., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., et al. Dynamic functional connectivity: promise, issues, and interpretations. *NeuroImage*, 80:360–378, 2013.

James, A. L. and Roulet, N. T. Investigating hydrologic connectivity and its association with threshold change in runoff response in a temperate forested watershed. *Hydrological Processes: An International Journal*, 21(25):3391–3408, 2007.

Jordan, A., Huitema, D., Schoenefeld, J., Van Asselt, H., and Forster, J. Governing climate change polycentrically. *Governing climate change: Polycentricity in action*, pages 3–26, 2018.

Jorgensen, P. E. T. and Pearse, E. P. J. A Hilbert space approach to effective resistance metric. *Complex Analysis and Operator Theory*, 4(4):975–1013, 2010.

Kazadi, A., Doss-Gollin, J., Sebastian, A., and Silva, A. Floodgnn-gru: a spatio-temporal graph neural network for flood prediction. *Environmental Data Science*, 3:e21, 2024.

Kelman, I. Climate change and the sendai framework for disaster risk reduction. *International Journal of Disaster Risk Science*, 6(2):117–127, 2015.

Kim, S. H., Lee, H. S., Kang, B. J., Song, B. J., Kim, H.-B., Lee, A., and Jin, M.-S. Dynamic contrast-enhanced mri perfusion parameters as imaging biomarkers of angiogenesis in invasive ductal carcinoma. *PLOS ONE*, 11(12):e0168632, 2016.

Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2017.

Kivelä, M. et al. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014.

Klein, D. and Randic, M. Resistance distance. *Journal of Mathematical Chemistry*, 12:81–95, 1993.

Lakshminarayanan, B., Pritzel, A., and Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems (NeurIPS 2017)*, volume 30, 2017.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023. doi: 10.1126/science.adi2336.

Lantuéjoul, C. *Geostatistical Simulation: Models and Algorithms*. Springer, 2002.

Lauritzen, S. L. *Graphical Models*. Number 17 in Oxford Statistical Science Series. Clarendon Press, Oxford, 1996. ISBN 0-19-852219-3.

Lilleborge, K., Martino, S., Fuglstad, G.-A., Lindgren, F., and Ingebrigtsen, R. Joint modelling of line and point data on metric graphs, 2025. arXiv preprint at arXiv.2505.01175.

Lim, M., Abdullah, A., Jhanjhi, N., and Khan, K. Link prediction in time-evolving criminal network with deep reinforcement learning technique. *IEEE Access*, PP:1–1, 2019. doi: 10.1109/ACCESS.2019.2958873.

Lindgren, F., Rue, H., and Lindström, J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4):423–498, 2011.

Lindquist, M. A., Xu, Y., Nebel, M. B., and Caffo, B. S. Evaluating dynamic bivariate correlations in resting-state fmri: a comparison study and a new approach. *NeuroImage*, 101:531–546, 2014.

Ludescher, J. et al. Improved El Niño forecasting by cooperativity detection. *PNAS*, 110 (29):11742–11745, 2013.

Mankad, S. and Michailidis, G. Structural and functional discovery in dynamic networks with non-negative matrix factorization. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 88(4):042812, 2013. ISSN 1550-2376. doi: 10.1103/PhysRevE.88.042812.

Masselink, R. J., Heckmann, T., Temme, A. J., Anders, N. S., Gooren, H. P., and Keesstra, S. D. A network theory approach for a better understanding of overland flow connectivity. *Hydrological Processes*, 31(1):207–220, 2017.

Maturana, H. R. and Varela, F. J. *Autopoiesis and Cognition: The Realization of the Living*, volume 42 of *Boston Studies in the Philosophy of Science*. D. Reidel Publishing Company, Dordrecht, 1980. ISBN 978-90-277-1015-2. doi: 10.1007/978-94-009-8947-4.

Mearns, L., Giorgi, F., McDaniel, L., and Shields, C. Climate scenarios for the southeastern US based on GCM and regional model simulations. *Issues in the Impacts of Climate Variability and Change on Agriculture: Applications to the southeastern United States*, pages 7–35, 2003.

Medaglia, J. D., Lynall, M.-E., and Bassett, D. S. Cognitive network neuroscience. *Journal of cognitive neuroscience*, 27(8):1471–1491, 2015.

Mendes, R. V., Mendes, H. C., and Araujo, T. Signal processing on graphs: Transforms and tomograms. *arXiv preprint arXiv:1406.2185*, 2014.

Menegatto, V., Oliveira, C., and Porcu, E. Gneiting class, semi-metric spaces and isometric embeddings. *Constructive Mathematical Analysis*, 3(2):85–95, 2020.

Meyer, C. Generalized Inversion of Modified Matrices. *SIAM Journal on Applied Mathematics*, 24, 1973.

Miller, K. and Samko, S. Completely monotonic functions. *Integral Transforms and Special Functions*, 12(4):389–402, 2001.

Møller, J. and Rasmussen, J. G. Cox processes driven by transformed Gaussian processes on linear networks—A review and new contributions. *Scandinavian Journal of Statistics*, 51 (3):1288–1322, 2024.

Montembeault, M., Joubert, S., Doyon, J., Carrier, J., Gagnon, J.-F., Monchi, O., Lungu, O., Belleville, S., and Brambati, S. M. The impact of aging on gray matter structural covariance networks. *Neuroimage*, 63(2):754–759, 2012.

Moradi, M. and Mateu, J. First-and second-order characteristics of spatio-temporal point processes on linear networks. *Journal of Computational and Graphical Statistics*, 29(3): 432–443, 2020.

Moshe, Z., Metzger, A., Elidan, G., Kratzert, F., Nevo, S., and El-Yaniv, R. Hydronets: Leveraging river structure for hydrologic modeling. *arXiv preprint arXiv:2007.00595*, 2020.

Muhammad, A. Managing river basins with thinking machines. In *2016 IEEE conference on Norbert Wiener in the 21st century (21CW)*, pages 1–6. IEEE, 2016.

Newman, M., Barabási, A.-L., and Watts, D. J. *The structure and dynamics of networks*. Princeton university press, 2011.

NOAA National Centers for Environmental information, Climate at a Glance: Global Mapping, 2025. `https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/global/mapping`, published June 2025, retrieved on June 26.

Ortega, A., Frossard, P., Kovacevic, J., Moura, J., and Vandergheynst, P. Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE*, 106(5): 808–828, 2018.

Ostrom, E. Polycentric systems for coping with collective action and global environmental change. In *Global justice*, pages 423–430. Routledge, 2017.

Perraudin, N. and Vandergheynst, P. Stationary signal processing on graphs. *IEEE Transactions on Signal Processing*, 65(13):3462–3477, 2017.

Perry, P. O. and Wolfe, P. J. Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: Series B*, 75(5):821–849, 2013.

Peterson, E. E., Theobald, D. M., and ver Hoef, J. M. Geostatistical modelling on stream networks: Developing valid covariance matrices based on hydrologic distance and stream flow. *Freshwater Biology*, 52(2):267–279, 2007.

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., et al. Modelling dendritic ecological networks in space: An integrated network perspective. *Ecology Letters*, 16(5): 707–719, 2013.

Pinder, T., Turnbull, K., Nemeth, C., and Leslie, D. Gaussian processes on hypergraphs. *arXiv preprint arXiv:2106.01982*, 2021.

Platt, W. J., Evans, G. W., and Rathbun, S. L. The population dynamics of a long-lived conifer (pinus palustris). *The American Naturalist*, 131(4):491–525, 1988.

Porcu, E., Bevilacqua, M., and Hering, A. S. The shkarofsky-gneiting class of covariance models for bivariate gaussian random fields. *Stat*, 7(1):e207, 2018.

Porcu, E., Furrer, R., and Nychka, D. 30 years of space–time covariance functions. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(2):e1512, 2021.

Porcu, E., Emery, X., and Peron, A. P. Nested covariance functions on graphs with euclidean edges cross time. *Electronic Journal of Statistics*, 16(2):4222–4246, 2022.

Porcu, E., White, P. A., and Genton, M. G. Stationary nonseparable space-time covariance functions on networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(5):1417–1440, 2023.

Porcu, E., Bevilacqua, M., Schaback, R., and Oates, C. J. The matérn model: A journey through statistics, numerical analysis and machine learning. *Statistical Science*, 39(3): 469–492, 2024.

Porcu, E., Filosi, T., and Simon, H. A Triad of Networks and a Triad of Fusions for the Other Climate Crisis, 2025. arXiv preprint at arXiv:2510.09728.

Preti, M. G., Bolton, T. A. W., and Van De Ville, D. The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage*, 160:41–54, 2017.

Radebach, A., Donner, R. V., Runge, J., Donges, J. F., and Kurths, J. Disentangling different types of El Niño episodes by evolving climate network analysis. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 88(5):052807, 2013.

Rakshit, S., Nair, G., and Baddeley, A. Second-order analysis of point patterns on a network using any distance metric. *Spatial Statistics*, 22:129–154, 2017.

Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., and Thuerey, N. Weatherbench: A benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11):e2020MS002203, 2020. doi: 10.1029/2020MS002203.

Reams, R. Hadamard inverses, square roots and products of almost semidefinite matrices. *Linear Algebra and its Applications*, 288:35–43, 1999.

Rico, J., Barateiro, J., and Oliveira, A. Graph neural networks for traffic forecasting: The research progress. *ISPRS International Journal of Geo-Information*, 12(3):100, 2023.

Rossi, R. A., Gallagher, B., Neville, J., and Henderson, K. Modeling dynamic behavior in large evolving graphs. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 667–676. ACM, 2013. ISBN 978-1-4503-1869-3. doi: 10.1145/2433396.2433479.

Roudbari, N. S., Punekar, S. R., Patterson, Z., Eicker, U., and Poullis, C. From data to action in flood forecasting leveraging graph neural networks and digital twin visualization. *Scientific reports*, 14(1):18571, 2024.

Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M., Hartman, D., Marwan, N., Paluš, M., and Kurths, J. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature communications*, 6(1):8502, 2015.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., et al. Inferring causation from time series in earth system sciences. *Nature communications*, 10(1):2553, 2019a.

Runge, J. et al. Detecting causal associations in large nonlinear time series datasets. *Nature Communications*, 10:2553, 2019b. doi: 10.1038/s41467-019-10105-3.

Santos-Fernandez, E., Hoef, J. M. V., McGree, J. M., Isaak, D. J., Mengersen, K., and Peterson, E. E. Ssnbayes: An r package for bayesian spatio-temporal modelling on stream networks. *arXiv preprint arXiv:2202.07166*, 2022.

Sarkar, A., Vandal, T., Ganguly, S., Nemani, R., and Ganguly, A. R. Downscaling extreme precipitation with deep generative models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04):139–146, 2020. doi: 10.1609/aaai.v34i04.5755.

Schäfer, B., Matthiae, M., Timme, M., and Witthaut, D. Dynamically induced cascading failures in power grids. *Nature Communications*, 9(1):1975, 2018.

Schoenberg, I. J. Metric Spaces and Completely Monotone Functions. *Annals of Mathematics*, 25(39):811–841, 1938.

Schoenberg, I. J. Positive Definite Functions on Spheres. *Duke Math. Journal*, 9:96–108, 1942.

Segarra, S., Marques, A. G., Mateos, G., and Ribeiro, A. Network topology inference from spectral templates. *IEEE Transactions on Signal and Information Processing over Networks*, 3(3):467–483, 2017.

Shaw, T. A. and Stevens, B. What climate science needs. *Nature Climate Change*, 13: 105–112, 2023. doi: 10.1038/s41558-022-01573-9.

Shaw, T. A. and Stevens, B. The other climate crisis. *Nature*, 628(8006):35–44, 2025. doi: 10.1038/s41586-025-08680-1.

Shi, J., Stebliankin, V., Wang, Z., Wang, S., and Narasimhan, G. Graph transformer network for flood forecasting with heterogeneous covariates. *arXiv preprint arXiv:2310.07631*, 2023.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., Kerminen, A., and Jordan, M. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10), 2006.

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., and Vandergheynst, P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013. doi: 10.1109/MSP.2012.2235192.

Sit, M., Demiray, B., and Demir, I. Short-term hourly streamflow prediction with graph convolutional gru networks. *arXiv preprint arXiv:2107.07039*, 2021.

Stanković, L., Mandic, D., Daković, M., Brajović, M., Scalzo, B., Li, S., Constantinides, A. G., et al. Data analytics on graphs part iii: Machine learning on graphs, from graph topology to applications. *Foundations and Trends® in Machine Learning*, 13(4):332–530, 2020.

Star, S. L. and Griesemer, J. R. Institutional ecology, translations and boundary objects: Amateurs and professionals in berkeley's museum of vertebrate zoology, 1907-39. *Social studies of science*, 19(3):387–420, 1989.

Stein, M. L. *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.

Stolbova, V., Martin, P., Bookhagen, B., Marwan, N., and Kurths, J. Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka. *Nonlinear Processes in Geophysics*, 21(4):901–917, 2014.

Sun, A. Y., Jiang, P., Yang, Z.-L., Xie, Y., and Chen, X. A graph neural network approach to basin-scale river network learning: The role of physics-based connectivity and data fusion. *Hydrology and Earth System Sciences Discussions*, 2022:1–35, 2022.

Suri, D. and Mangal, M. Powergnn: A topology-aware graph neural network for electricity grids. *arXiv preprint arXiv:2503.22721*, 2025.

Taghizadeh, M., Zandsalimi, Z., Nabian, M. A., Shafiee-Jood, M., and Alemazkoor, N. Interpretable physics-informed graph neural networks for flood forecasting. *Computer-Aided Civil and Infrastructure Engineering*, 2025.

Tang, J. and Zimmerman, D. Space-time covariance models on networks. *Electronic Journal of Statistics*, 18(1):490–514, 2024. ISSN 1935-7524.

Thanou, D., Dong, X., Kressner, D., and Frossard, P. Learning heat diffusion graphs. *IEEE Transactions on Signal and Information Processing over Networks*, 3(3):484–499, 2017.

Tsonis, A. A. and Roebber, P. J. Architecture and dynamics of a climate network. *Physica A*, 333:497–504, 2004.

Tsonis, A. A. and Swanson, K. L. Topology and predictability of El Niño and La Niña networks. *Physical Review Letters*, 100(22):228502, 2008.

Tsonis, A. A., Swanson, K. L., and Roebber, P. J. What do networks have to do with climate? *Bulletin of the American Meteorological Society*, 87(5):585–596, 2006.

Tsonis, A. A., Wang, G., Swanson, K. L., Rodrigues, F. A., and Costa, L. d. F. Community structure and dynamics in climate networks. *Climate dynamics*, 37(5):933–940, 2011.

Uddin, L. Q. and Karlsgodt, K. H. Future directions for examination of brain networks in neurodevelopmental disorders. *Journal of Clinical Child & Adolescent Psychology*, 47(3): 483–497, 2018.

United Nations Framework Convention on Climate Change. Adaptation communications. `https://unfccc.int/topics/adaptation-and-resilience/workstreams/adaptation-communications`, 2024. Accessed June 2025.

Van Der Mheen, M., Dijkstra, H. A., Gozolchiani, A., Den Toom, M., Feng, Q., Kurths, J., and Hernandez-Garcia, E. Interaction network based early warning indicators for the atlantic MOC collapse. *Geophysical Research Letters*, 40(11):2714–2719, 2013.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R. Deepsd: Generating high resolution climate change projections through single image super-resolution. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 1663–1672, New York, NY, USA, 2017. Association for Computing Machinery. doi: 10.1145/3097983.3098004.

Vandal, T., Kodra, E., Ganguly, A. R., Michaelis, A., Nemani, R., and Ganguly, S. Intercomparison of machine learning methods for statistical downscaling: The case of daily and extreme precipitation. *Theoretical and Applied Climatology*, 137(1–2):557–570, 2019. doi: 10.1007/s00704-018-2613-3.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. Graph attention networks. In *International Conference on Learning Representations (ICLR)*, 2018.

Ver Hoef, J., Dumelle, M., Higham, M., et al. Indexing and partitioning the spatial linear model for large data sets. *PLoS ONE*, 2023.

Ver Hoef, J. M., Peterson, E., and Theobald, D. Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics*, 13(4):449–464, 2006.

Wan, H., Xiang, L., Cai, Y., Xie, Y., and Xu, R. Temporal and spatial feature extraction using graph neural networks for multi-point water quality prediction in river network areas. *Water Research*, 281:123561, 2025.

Wang, Y.-X., Sharpnack, J., Smola, A. J., and Tibshirani, R. J. Trend filtering on graphs. *Journal of Machine Learning Research*, 17(105):1–41, 2016.

Wiedermann, M., Radebach, A., Donges, J. F., Kurths, J., and Donner, R. V. A climate network-based index to discriminate different types of El Niño and La Niña. *Geophysical Research Letters*, 43(13):7176–7185, 2016.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Philip, S. Y. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2021. doi: 10.1109/TNNLS.2020.2978386.

Xiao, S., Yan, J., Yang, X., Zha, H., and Chu, S. Modeling the intensity function of point process via recurrent neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31. Association for the Advancement of Artificial Intelligence, 2017.

Yu, Y., Li, P., Huang, D., and Sharma, A. Street-level temperature estimation using graph neural networks: Performance, feature embedding and interpretability. *Urban Climate*, 56: 102003, 2024.

Yuan, M. and Lin, Y. Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35, 2007. doi: 10.1093/biomet/asm018.

Zanfei, A., Brentan, B. M., Menapace, A., Righetti, M., and Herrera, M. Graph convolutional recurrent neural networks for water demand forecasting. *Water Resources Research*, 58(7): e2022WR032299, 2022.

Zastavnyi, V. P. Analog of Schoenberg's theorem for a-conditionally negative definite matrix-valued kernels. *Mathematical Notes*, 114(1):66–76, 2023. ISSN 0001-4346, 1573-8876.

Zastavnyi, V. P. On positive definiteness of some functions. *Journal of Multivariate Analysis*, 73(1):55–81, 2000. ISSN 0047-259X.

Zhang, F. The schur complement and its applications. In *The Schur complement and its applications*, volume 4 of *Numerical Methods and Algorithms*. Springer-Verlag, 2005. ISBN 978-0-387-24271-2.

Zheng, Y., Capra, L., Wolfson, O., and Yang, H. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3): 38, 2014.