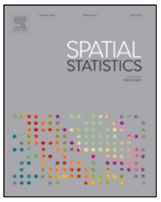
Spatial Statistics 58 (2023) 100784

Contents lists available at ScienceDirect

Spatial Statistics

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



An object-oriented approach to the analysis of spatial complex data over stream-network domains

Chiara Barbi, Alessandra Menafoglio *, Piercesare Secchi

MOX – Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan, 20133, Italy

ARTICLE INFO

Article history:
Received 18 July 2023
Received in revised form 27 September 2023
Accepted 2 October 2023
Available online 13 November 2023

Keywords:
Geostatistics
Functional Data Analysis
Stream distance
Kriging

ABSTRACT

We address the problem of spatial prediction for Hilbert data, when their spatial domain of observation is a river network. The reticular nature of the domain requires to use geostatistical methods based on the concept of Stream Distance, which captures the spatial connectivity of the points in the river induced by the network branching. Within the framework of Object Oriented Spatial Statistics (O2S2), where the data are considered as points of an appropriate (functional) embedding space, we develop a class of functional moving average models based on the Stream Distance. Both the geometry of the data and that of the spatial domain are thus taken into account. A consistent definition of covariance structure is developed, and associated estimators are studied. Through the analysis of the summer water temperature profiles in the Middle Fork River (Idaho, USA), our methodology proved to be effective, both in terms of covariance structure characterization and forecasting performance.

1. Introduction

The need to analyse and extract useful information from extremely complex and varied data has certainly been a central challenge for the statistical community in recent years. The statistical methods formulated for scalar data are not useable in those – increasingly frequent – contexts in

* Corresponding author.
E-mail address: alessandra.menafoglio@polimi.it (A. Menafoglio).

https://doi.org/10.1016/j.spasta.2023.100784

which the data are featured by a high complexity (such as curves, surfaces or images). For this reason, Functional Data Analysis (FDA, Ramsay and Silverman (2005)) and Object Oriented Data Analysis (OODA, Marron and Alonso (2014)) have attracted great interest among researchers and extensive effort has been made in developing functional versions for a wide range of classical statistical methods. Whenever data are georeferenced, however, the complexity of the data is compounded by the need to take into account the dependence between observations induced by their spatial proximity. A relatively large body of literature has recently focused on developing methods of spatial statistics for general types of data objects, including functional data, distributions and data belonging to Riemannian manifolds. These efforts lie within the domain of Object Oriented Spatial Statistics (O2S2, Menafoglio and Secchi (2017)), a recent system of ideas for the analysis of spatial complex data, founded on a strong geometrical approach to the data analysis. The methods developed so far allow one to model the dependence among data, perform dimensionality reduction, as well as perform prediction at unsampled locations within the domain Horváth and Kokoszka (2012), Menafoglio and Secchi (2017), Mateu and Giraldo (2021). However, all these methods are focused on Euclidean spatial domains, or on mildly non-Euclidean spatial regions that, locally, admit a Euclidean representation (see Menafoglio et al. (2018, 2021)). As a matter of fact, vast areas of geosciences study random processes which naturally develop over non-Euclidean settings, where the closeness between data locations is naturally expressed through the shortest path (i.e., the geodesic) induced by the physics of the phenomenon. For instance, when studying aquatic variables in a stream network system, the proximity among sites is better represented by the *water distance* which separates the locations, rather than by the Euclidean shortest path, which does not account for the topology and connectivity of the network. Aiming to enrich that branch of the environmental and ecological research that deals with the characterization of freshwater stream environments, we will analyse the maximum daily water temperatures of the Middle Fork river, a 104-mile-long (167 km) river in central Idaho, USA. Indeed, concerns about climate change and its consequences in terms of habitat alteration, have led to extensive stream temperature monitoring by dozens of natural resources agencies throughout North America and Europe. The development of methods for analysing these data is therefore of great topicality. The data under analysis consist of the maximum daily water temperatures recorded for 47 days at 157 locations of the Middle Fork river, depicted in Fig. 1a together with the observation sites. The temperature profiles for each location are shown in Figs. 1b and 1c.

Although relevant for an increasing number of industrial and environmental applications (see, e.g., Menafoglio and Secchi (2019)), working with non-Euclidean spatial domains poses challenges, because the usual parametric families (e.g., spherical, Matérn) for the covariance among observations may be no longer positive semi-definite under a non-Euclidean metric (Curriero, 2006). Nonetheless, in a few cases, it is possible to derive ad-hoc parametric families, which are well-suited to the topology of the domain under study. This is the case of the models for stream networks proposed and extensively studied by Ver Hoef et al. (2006), Cressie et al. (2006), Peterson et al. (2007), Ver Hoef and Peterson (2010), Peterson and Ver Hoef (2010). These models are built upon a moving average construction of Yaglom (1987), and precisely account for the dependence among observations induced by their water distance (named *stream distance*). This approach yields valid covariance models and proper estimation procedures for spatial data, which can be used whenever their domain of reference can be represented as a binary tree — the water flowing from its root to its leafs. Although these innovative models exhibit an incredible potential, their range of action is still limited to scalar data. As a matter of fact, while sensors typically record relevant variables continuously along time, previous works need to compress this rich set of information into scalar summaries (e.g., the monthly average temperature, the average weekly dissolved oxygen), inevitably leading to a loss of information. The aim of this work is to overcome these limitations, extending the theory of Ver Hoef et al. (2006), Cressie et al. (2006) to general object data, provided that these can be embedded in a (separable) Hilbert space. This setting includes, e.g., the case of functional data (which are typically embedded in the space L^2 of square-integrable functions) as well as that of distributional data (for which the embedding in a Bayes Hilbert space can be used, Van Den Boogaart et al. (2014)). To the authors' knowledge, the only existing work enabling the analysis of functional data over a stream network is that by Haggarty et al. (2014). Motivated by the clustering

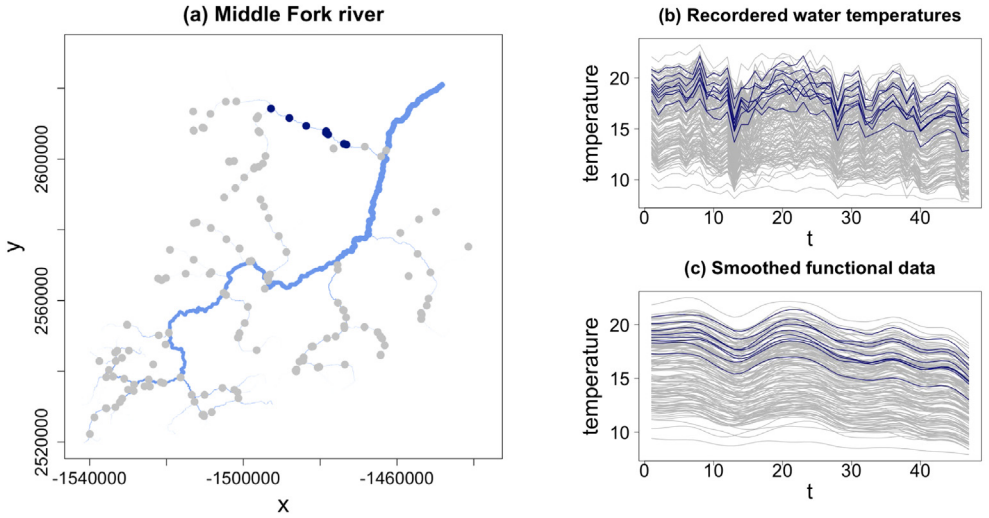


Fig. 1. (a) Middle Fork river. Points indicate the locations of observed data. (b) Water temperatures at the 157 locations of the Middle Fork River from 15 July 2005 to 31 August 2005. (c) Smoothed functional data.

analysis of temporal profiles of nitrate concentrations along the River Tweed (Scotland), these authors propose to model the spatial covariance among observations through the valid models of Ver Hoef et al. (2006), by grounding on integral summaries of the functional data. However, even though the framework of Haggarty et al. (2014) uses typical concepts of FDA, it does not provide a characterization of the infinite-dimensional random field generating the data (being the covariance actually based on scalar summaries), and thus only allows for unsupervised (explorative) analyses. As a key element of innovation with respect to existing literature, we here provide a direct construction of a functional moving average process distributed over a stream network, that creates a solid foundation upon which developing a strategy for variographic analysis and estimation of the spatial covariance structure, which can ultimately be used for the scope of spatial prediction. The remaining part of this work is organized as follows. Section 2 presents a review of the models of Ver Hoef et al. (2006), which is instrumental to the extension of the construction to Hilbert data presented in Section 3. Section 4 proposes estimators for the spatial dependence of the field under stationary and non-stationary assumptions, and presents the associated Kriging predictors. Section 5 discusses two illustrative simulated examples and in Section 6 the proposed methods are employed in the analysis of the water temperature profiles along the Middle Fork river.

2. Stream network models for scalar observations

In this section, a brief review of the models proposed by Ver Hoef et al. (2006) for scalar data distributed over a stream network is given. The reader is referred to Ver Hoef et al. (2006), Peterson et al. (2007), Ver Hoef and Peterson (2010), Peterson and Ver Hoef (2010) for further details. The stream networks considered in this work are topologically modelled as dendritic networks made up of a finite number of stream segments, indexed by $i = 1, 2, \ldots$. Each segment, which is a set of point locations lying on a straight line, and is associated with a unique direction, that is the direction of the water flow. Having assumed the network to be dendritic, there will always be a single most-downstream point, to which from now on we will refer to as *the outlet*. It is therefore possible to define the "upstream distance" for each point in a network as the length of the path (on the network) that connects the point with the outlet. To set the notation, let D be the stream network domain, s a generic spatial location (i.e., a point) in D, and let I be the whole set of stream segment indices. For $i \in I$, the most downstream location in the ith segment is denoted as l_i, whereas

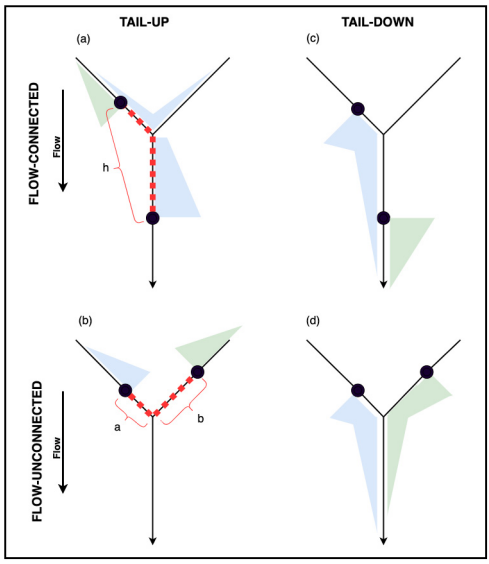


Fig. 2. Representation of flow-connected (a, c) and flow-unconnected (b, d) locations, and moving average functions for tail-up (a, b) and tail-down (c, d) models. Locations on the stream network are indicated as circles; stream-distances between locations are indicated as dashed lines in (a) and (b).
Source: Modified from (Peterson and Ver Hoef, 2010).

the most upstream location is u_i. The index set of stream segments upstream of a point s_i belonging to $i \in I$, will be $U_{s_i} \subseteq I$; segment i is excluded from U_{s_i}. Analogously, $D_{s_i} \subseteq I$ is the index set of all stream segments downstream of s_i, including the segment i containing s_i. Using these definitions, we can say that two locations, s_i and s_j, on a stream network are "flow-connected" (FC) if $D_{s_i} \cap D_{s_j} = D_{s_i}$ or $D_{s_i} \cap D_{s_j} = D_{s_j}$. In other words, water must flow from one site to another in order for the pair to be considered flow connected (see Fig. 2a and 2c). In the following, we denote by B_{s_i,s_j} the set of stream segments between two locations s_i, s_j, including the segment for the upstream location but excluding the segment for the downstream location. The same definition holds if we want to identify the segments between location s_i and segment j, for which we will use the notation $B_{s_i,[j]}$. Finally, it is possible to define the stream distance as the shortest distance between two locations, with the constrain that all displacements are taken along the network.

$$d(s_i, s_j) = \begin{cases} |s_i - s_j| & \text{if } s_i \text{ and } s_j \text{ are flow-connected,} \\ (s_i - u) + (s_j - u) & \text{otherwise.} \end{cases} \tag{1}$$

Here u is the nearest junction downstream which is common to both flow-unconnected locations. Consider now two flow-unconnected locations. Conventionally, we will use a to indicate the shortest distance to u while b indicates the largest one. We use h for the distance between two FC locations (see Fig. 2b).

We are now able to enter the core of the models proposed by Ver Hoef et al. (2006). To build the random process $\{Z(s), s \in D\}$ on the stream network domain D, these authors generalize the moving-average construction of Yaglom (1987), originally designed on $\mathbb{R}^1$, to the topology of D. Yaglom (1987) defines the element $Z(s)$ of a random process on $\mathbb{R}^1$ as

$$Z(s) = \int_{-\infty}^{+\infty} g(x - s|\boldsymbol{\theta}) dW(x) \tag{2}$$

where $W(x)$ is a white noise process and $g(x|\boldsymbol{\theta})$ is called the moving-average (MA) function, which is defined on $\mathbb{R}^1$ and assumed to be squared integrable. To account for the topology of the domain, Ver

Hoef et al. (2006) use the same construction, but compute the integral in (2) piece-wise, summing up the contribution from each segment of the network associated with non-null values of the MA function $g(x|\boldsymbol{\theta})$. The key idea is that the overlap between the MA function of one random variable and that of another give rise to a partial correlation between these two variables. Notice that the moving average function could go in both directions, up and down the stream with respect to flow, and this choice will discriminate whether the final model will be a *tail-up* or *tail-down*, respectively. From the moving average construction (2) it is possible to obtain the autocovariance between two elements of the field $Z(s)$ and $Z(s+h)$ as

$$C_t(h|\boldsymbol{\theta}) = \begin{cases} \int_{-\infty}^{+\infty} (g(x|\boldsymbol{\theta}))^2 dx + \eta & h = 0 \\ \int_{-\infty}^{+\infty} g(x|\boldsymbol{\theta}) g(x-h|\boldsymbol{\theta}) dx & h > 0, \end{cases} \tag{3}$$

where η is the nugget effect. From this construction, when $D \subset \mathbb{R}^1$, several classes of models can be obtained (e.g., spherical, exponential, Mariah; see Yaglom (1987)). Analogously, parametric classes are obtained by Ver Hoef et al. (2006) by computing the integrals in (3) piece-wise along the stream network. Table 1 reports examples of tail-up and tail-down models, which are further specified below. For clarity of notation, recall that, as in classical geostatistics, the (practical) range parameter θ_r represents the minimum separating distance for two observations to be (approximately) uncorrelated, whereas the sill parameter θ_v defines the variance of the process.

2.1. Tail-up models

In the tail-up models (TU), the support of the moving average functions is not null only moving in the upstream direction (Fig. 2a and 2b). Obviously, if two locations are not flow connected, the corresponding tail up moving average will never overlap (Fig. 2b). Hence, null covariance is associated to two flow unconnected random variables, making tail-up models particularly appropriate when the variable of interest is dominated by flow (e.g. organisms or materials that move passively downstream like pollutants, waterborne chemicals and so on). The way that $g(x|\boldsymbol{\theta})$ gets split as we go upstream plays a crucial role to ensure the stationarity of the spatial process. Segment weights ω_k are used to proportionally split the function between upstream segments when the MA function reaches a confluence in the network, eventually obtaining the following expression for the element $Z(s_i)$

$$Z(s_i) = \int_{s_i}^{u_i} g(x - s_i|\boldsymbol{\theta}) dW(x) + \sum_{j \in U_{s_i}} \left(\prod_{k \in B_{s_i,[j]}} \sqrt{\omega_k} \right) \int_{l_j}^{u_j} g(x - s_i|\boldsymbol{\theta}) dW(x). \tag{4}$$

In (4), for each segment i in the network, the weights associated to the two segments j and k in which segment i splits are such that $0 \leq \omega_j, \omega_k \leq 1$ and $\omega_j + \omega_k = 1$. Note that the weights ω_k may be chosen as to reflect specific hydrological characteristics of each segment, such as discharge, watershed area or flow volume (Ver Hoef et al., 2006). The covariance between two random elements $Z(s_i)$, $Z(s_j)$ defined by (4) is then given by

$$C(s_i, s_j|\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } s_i \text{ and } s_j \text{ are not flow connected} \\ C_t(0|\boldsymbol{\theta}) & \text{if } s_i = s_j \\ \pi_{i,j} C_t(h|\boldsymbol{\theta}) & \text{otherwise,} \end{cases} \tag{5}$$

where $\pi_{i,j} = \prod_{k \in B_{s_i,s_j}} \sqrt{\omega_k}$, h is the stream distance between the two flow connected locations on the stream network, and the (unweighted) covariance functions $C_t(0|\boldsymbol{\theta})$ are obtained by using moving average functions in one dimension given in (3).

2.2. Tail-down models

In contrast to the tail-up models, tail-down (TD) models arise when the MA function is non-zero only downstream of a location. This means that the "tail" of the moving average functions points

in the flow direction (Fig. 2c and 2d). Therefore, the tail-down random variable has the following expression:

$$Z(s) = \int_{-\infty}^{s} g(s - x|\boldsymbol{\theta}) dW(x). \tag{6}$$

As shown in Fig. 2c and 2d and by direct computations, autocorrelation in tail-down models is allowed both for flow-connected and flow-unconnected locations. Moreover, since the MA functions do not split at the junctions, introducing a weighting procedure is not needed anymore. Some examples of the tail-down covariance structures are given in Table 1. Due to their characteristic of allowing correlation for both connected and not connected pairs of sites, tail-down models are particularly indicated for modelling variables, such as fish or aquatic insects, that can move both upstream and downstream.

3. Functional random fields over stream networks

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and H a separable Hilbert space, equipped with operations $(+, \cdot)$ and inner product $\langle \cdot, \cdot \rangle$, inducing the norm $\| \cdot \|$. Following (Menafoglio et al., 2013), we consider the case of real-valued functional observations and assume that each element of H is a function $\mathcal{X} : \tau \to \mathbb{R}$, τ being a compact subset of $\mathbb{R}$. Denote by D the spatial domain, and let

$$\{\mathcal{X}_s, s \in D \subseteq \mathbb{R}^d\} \tag{7}$$

be a functional random field valued in H. The theory of random processes on Hilbert spaces is well established when D is a subset of $\mathbb{R}^d$ (see, e.g., Bosq, 2000); in the following, we elaborate on the case of D being a stream network domain, defining the field (7) by direct construction. In this work, we will always assume the square-integrability of the process, i.e., that each element $\mathcal{X}_s$, $s \in D$, of the random field is such $\mathbb{E}[\|\mathcal{X}_s\|^2] < \infty$; we denote this as $\mathcal{X}_s \in L^2(\Omega; H)$. As in the usual setting of geostatistics, we consider that a partial observation of the field is available at given (non-random) spatial locations $s_1, \dots, s_n$ in D, and denote the functional dataset as $\mathcal{X}_{s_1}, \dots, \mathcal{X}_{s_n}$. Following (Bosq, 2000), for any s_1, s_2 in D, we define the cross-covariance operator between the elements $\mathcal{X}_{s_1}$ and $\mathcal{X}_{s_2}$ of (7) as the operator $\mathcal{C}_{s_1,s_2} : H \to H$ acting on the (non-random) element $x \in H$ as $\mathcal{C}_{s_1,s_2} x = \mathbb{E}[\langle \mathcal{X}_{s_1} - m_{s_1}, x \rangle (\mathcal{X}_{s_2} - m_{s_2})]$, with m_{s_1} (m_{s_2}) the mean of the process in s_1 (s_2). The family of cross-covariance operators $\{\mathcal{C}_{s_1,s_2}, s_1, s_2 \in D\}$ fully defines the second-order properties of the field (Bosq, 2000; Horváth and Kokoszka, 2012). A *(global)* measure of dependence for the process (7) is instead provided by the so-called *trace-covariogram* (Caballero et al., 2013; Menafoglio et al., 2013). This is defined as the (real-valued) function $C : D \times D \to \mathbb{R}$:

$$C(s_1, s_2) = \mathbb{E}\big[\langle \mathcal{X}_{s_1} - m_{s_1}, \mathcal{X}_{s_2} - m_{s_2} \rangle\big]. \tag{8}$$

Note that $C(s_1, s_2)$ defines a scalar product on $L^2(\Omega; H)$ and it is positive definite. Moreover, $C(s_1, s_2)$ coincides with the trace of the cross-covariance operator $\mathcal{C}_{s_1,s_2}$ (Menafoglio et al., 2013). Recall also that the field (7) is *second-order stationary* if (*i*) the mean is spatially constant ($\mathbb{E}[\mathcal{X}_s] = m$ for all $s \in D$), and (*ii*) the family of cross-covariance operators is stationary, i.e., if there exist a family of operators $\{\mathcal{C}_h, h \in \mathbb{R}^d\}$ such that $\mathcal{C}_{s_1,s_2} = \mathcal{C}_h$ for all s_1, s_2 satisfying $s_1 - s_2 = h$. The assumption of *global* second-order stationarity requires, instead of condition (*ii*), that (*ii'*) the trace-covariogram is stationary, i.e., that there exist a function $\tilde{C}$ such that $\tilde{C}(h) = C(s_1, s_2)$ for all s_1, s_2 satisfying $s_1 - s_2 = h$.

Hörmann and Kokoszka (2011) show that every functional random process (7) with constant mean can be expressed through the following basis expansion

$$\mathcal{X}_s = m + \sum_{k \geq 1} \xi_k(s) e_k. \tag{9}$$

Here $\{e_k, k \geq 1\}$ is an orthonormal basis of H and the random coefficients $\xi_k(s) = \langle \mathcal{X}_s - m, e_k \rangle$ are the projections of the functional random variable $\mathcal{X}_s$ on the orthonormal basis. These coefficients determine both the stationarity and the covariance structure of the functional process. To ease the notation, we hereafter assume the process to be zero mean.

Table 1
Covariograms and Semivariograms for tail-up and tail-down models. $\theta_r, \theta_v \in \mathbb{R}^+$ are respectively the range and the sill parameters. The notation FC and FU is used to denote respectively that s_i and s_j are flow-connected or flow-unconnected.

Name	Covariogram	Semivariogram
Tail-up Linear with Sill	$C(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} \theta_v & \text{if } s_i = s_j \\ \pi_{i,j}\theta_v \left(1 - \frac{h}{\theta_r}\right) \mathbb{1}_{(\frac{h}{\theta_r} \le 1)} & \text{if FC} \\ 0 & \text{if FU.} \end{cases}$	$\gamma(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } s_i = s_j \\ \theta_v - \pi_{i,j}\theta_v \left(1 - \frac{h}{\theta_r}\right) & \text{if FC and } h \le \theta_r \\ \theta_v & \text{if } h > \theta_r \text{ or if FU.} \end{cases}$
Tail-up Spherical	$C(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} \theta_v & \text{if } s_i = s_j \\ \pi_{i,j}\theta_v \left(1 - \frac{3}{2}\frac{h}{\theta_r} + \frac{1}{2}\frac{h^3}{(\theta_r)^3}\right) \mathbb{1}_{(\frac{h}{\theta_r} \le 1)} & \text{if FC} \\ 0 & \text{if FU.} \end{cases}$	$\gamma(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } h = 0 \\ \theta_v - \pi_{i,j}\theta_v \left(1 - \frac{3}{2}\frac{h}{\theta_r} + \frac{1}{2}\frac{h^3}{(\theta_r)^3}\right) & \text{if FC and } h \le \theta_r \\ \theta_v & \text{if } h > \theta_r \text{ or if FU.} \end{cases}$
Tail-up Exponential	$C(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} \theta_v & \text{if } s_i = s_j \\ \pi_{i,j}\theta_v exp(-h/\theta_r) & \text{if FC} \\ 0 & \text{if FU.} \end{cases}$	$\gamma(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } h = 0 \\ \theta_v - \pi_{i,j}\theta_v \big(exp(-h/\theta_r)\big) & \text{if FC} \\ \theta_v & \text{if FU.} \end{cases}$
Tail-up Mariah	$C(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} \theta_v & \text{if } s_i = s_j \\ \pi_{i,j}\theta_v \left(\frac{log(h/\theta_r+1)}{h/\theta_r}\right) \mathbb{1}_{(h>0)} & \text{if FC} \\ 0 & \text{if FU.} \end{cases}$	$\gamma(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } h = 0 \\ \theta_v - \pi_{i,j}\theta_v \left(\frac{log(h/\theta_r+1)}{h/\theta_r}\right) & \text{if } s_i \text{ and } s_j \text{ are FC} \\ \theta_v & \text{if } s_i \text{ and } s_j \text{ are FU.} \end{cases}$
Tail-down Linear with Sill	$C_d(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} \theta_v \left(1 - \frac{h}{\theta_r}\right) \mathbb{1}_{(\frac{h}{\theta_r} \le 1)} & \text{if FC} \\ \theta_v \left(1 - \frac{b}{\theta_r}\right) \mathbb{1}_{(\frac{b}{\theta_r} \le 1)} & \text{if FU.} \end{cases}$	$\gamma(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } h = 0 \\ \theta_v \frac{h}{\theta_r} & \text{if FC and } h \le \theta_r \\ \theta_v \frac{b}{\theta_r} & \text{if } b < \theta_r \text{ and if FU} \\ \theta_v & \text{otherwise.} \end{cases}$
Tail-down Spherical	$C_d(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} \theta_v \left(1 - \frac{3}{2}\frac{h}{\theta_r} + \frac{1}{2}\frac{h^3}{(\theta_r)^3}\right) \mathbb{1}_{(\frac{h}{\theta_r} \le 1)} & \text{if FC} \\ \theta_v \left(1 - \frac{3}{2}\frac{a}{\theta_r} + \frac{1}{2}\frac{b}{\theta_r}\right) \left(1 - \frac{b}{\theta_r}\right)^2 \mathbb{1}_{(\frac{b}{\theta_r} \le 1)} & \text{if FU} \end{cases}$	$\gamma(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } h = 0 \\ \theta_v \left(\frac{3}{2}\frac{h}{\theta_r} - \frac{1}{2}\frac{h^3}{(\theta_r)^3}\right) & \text{if FC and } h \le \theta_r \\ \theta_v - \theta_v \left(1 - \frac{3}{2}\frac{a}{\theta_r} + \frac{1}{2}\frac{b}{\theta_r}\right) \left(1 - \frac{b}{\theta_r}\right)^2 & \text{if FU and } b < \theta_r \\ \theta_v & \text{otherwise.} \end{cases}$
Tail-down Exponential	$C_d(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} \theta_v exp(-h/\theta_r) & \text{if FC} \\ \theta_v exp(-(a+b)/\theta_r) & \text{if FU} \end{cases}$	$\gamma(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } h = 0 \\ \theta_v \big(1 - exp(-h/\theta_r)\big) & \text{if FC} \\ \theta_v \big(1 - exp(-(a+b)/\theta_r)\big) & \text{if FU.} \end{cases}$
Tail-down Mariah	$C_d(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} \theta_v \left(\frac{log(h/\theta_r+1)}{h/\theta_r}\right) & \text{if FC, } h > 0 \\ \theta_v & \text{if FC, } h = 0 \\ \theta_v \left(\frac{log(a/\theta_r+1) - log(b/\theta_r+1)}{(a-b)/\theta_r}\right) & \text{if FU, } a \ne b \\ \theta_v \left(\frac{1}{a/\theta_r+1}\right) & \text{if FU, } a = b \end{cases}$	$\gamma(s_i, s_j\|\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } h = 0 \\ \theta_v \left(1 - \frac{log(h/\theta_r+1)}{h/\theta_r}\right) & \text{if FC, } h > 0 \\ \theta_v \left(1 - \frac{log(a/\theta_r+1) - log(b/\theta_r+1)}{(a-b)/\theta_r}\right) & \text{if FU, } a \ne b \\ \theta_v \left(1 - \frac{1}{a/\theta_r+1}\right) & \text{if FU, } a = b. \end{cases}$

3.1. Functional moving-average models on the real line

We now use the direct construction (9) to show the existence of a functional version of the MA random variables defined in (2). We set $D = \mathbb{R}^1$, and consider N independent, zero mean, second-order stationary and isotropic scalar random fields, $\{\xi_k(s), s \in D\}$ for $k = 1, \dots, N$. We further assume that each scalar random field is defined through a MA model

$$\xi_k(s) = \int_{-\infty}^{+\infty} g^{(k)}(x - s|\boldsymbol{\theta}) dW_k(x), \tag{10}$$

In (10), each $g^{(k)}(x-s|\boldsymbol{\theta})$ needs to be square integrable for the stochastic integral to be well defined, i.e., $\int_{-\infty}^{+\infty} |g^{(k)}(x - s|\boldsymbol{\theta})|^2 dx < +\infty$. Let us now focus on a truncated version of (9), obtained as

$$\mathcal{X}_s^{(N)} = \sum_{k=1}^{N} \xi_k(s) e_k, \tag{11}$$

where $\{e_k, k \geq 1\}$ is an orthonormal basis of H. In this case, each $\mathcal{X}_s$ is valued in $H^{(N)}$, where $H^{(N)} = span\{e_1, \dots, e_N\}$ is the finite-dimensional Hilbert space generated by the N orthonormal vectors $e_1, \dots, e_N$. Moreover, $\mathcal{X}_s$ is square-integrable (i.e., $\mathcal{X}_s \in L^2(\Omega; H^{(N)})$) as

$$\begin{aligned} \mathbb{E}\left[\left\| \mathcal{X}_s^{(N)} \right\|^2 \right] &= \mathbb{E}\left[\sum_{k=1}^{N} \left(\int_{-\infty}^{+\infty} g^{(k)}(x - s|\boldsymbol{\theta}) dW_k(x) \right)^2 \right] \\ &= \sum_{k=1}^{N} \int_{-\infty}^{+\infty} \left(g^{(k)}(x - s|\boldsymbol{\theta}) \right)^2 dx < +\infty, \end{aligned} \tag{12}$$

thanks to the Ito isometry and to the fact that each g is deterministic and square integrable. Hence, the variable $\mathcal{X}_s^{(N)}$ has finite second moment, which guarantees the existence of the family of cross-covariance operators for the process $\{\mathcal{X}_s^{(N)}, s \in D\}$. It is worth highlighting that the boundedness of the last sum in (12) is due to the finiteness of the orthonormal basis being considered. Letting $N \to +\infty$, the square-integrability of $\mathcal{X}_s$ is only obtained if the sequence $\{\xi_k(s)\}_{k \geq 1}$ belongs to $l^2(\Omega; \mathbb{R})$ (i.e., if $\sum_{k \geq 1} \mathbb{E}[\xi_k(s)^2] < \infty$). This can be guaranteed including additional assumptions on each moving average function $g(x|\boldsymbol{\theta})$ (see Appendix A). In this case, the MA random field (9) has a well-defined family of cross-covariance operators. Note that the covariance functions $C_k(s_1, s_2) = \mathbb{E}[\xi_k(s_2)\xi_k(s_2)]$ of the scalar random fields appearing in (9) completely characterize the family of cross-covariance operators of the field $\{\mathcal{X}_s, s \in D\}$, thus also its trace-covariogram (see, e.g., Hörmann and Kokoszka, 2011; Menafoglio et al., 2013). In particular, the trace-covariogram of the process (9), obtained from the moving average construction, is $C(s_i, s_j) = \sum_{k=1}^{N} \mathbb{E}\left[\xi_k(s_i)\xi_k(s_j) \right] = \sum_{k=1}^{N} C_t^{(k)}(h|\boldsymbol{\theta})$, where $C_t^{(k)}(h|\boldsymbol{\theta})$ is the autocovariance function of the kth scalar random field $\{\xi_k(s), s \in D\}$ (defined as in (3)). Since the family of valid covariograms is a convex cone and each $C_t^{(k)}(h|\boldsymbol{\theta})$ is a valid covariogram for the kth random field, $C(s_i, s_j)$ is clearly a valid covariance function. On the other hand, any of the valid covariance models available in D can be used to provide a valid covariance model for the field $\{\mathcal{X}_s, s \in D\}$.

3.2. Functional tail-up and tail-down models

The approach just introduced can be extended to a stream network domain D. Indeed, the previous arguments still hold true if we assume that each scalar random field is represented as a tail-up model, i.e., as (see also Eq. (4))

$$\xi_k(s) = \int_{s}^{u_i} g^{(k)}(x - s|\boldsymbol{\theta}) dW(x) + \sum_{j \in U_s} \left(\prod_{n \in B_{s,[j]}} \sqrt{\omega_n} \right) \int_{l_j}^{u_j} g^{(k)}(x - s|\boldsymbol{\theta}) dW(x).$$

In this case, when the functional random process is built by direct construction as in (9), the covariance function associated with each random field ξ_k is a combination of the covariance functions of the scalar field defined in (5). The trace-covariogram of the functional process is then easily obtained by linearity as

$$C(s_i, s_j) = \begin{cases} 0 & \text{if } s_i \text{ and } s_j \text{ are not flow connected} \\ \sum_{k\geq 1} C_t^{(k)}(0|\boldsymbol{\theta}) & \text{if } s_i = s_j \\ \pi_{i,j}\left(\sum_{k\geq 1} C_t^{(k)}(h|\boldsymbol{\theta})\right) & \text{otherwise.} \end{cases} \tag{13}$$

Concerning the tail-down models, the procedure is even simpler, since in this case the weights are not needed. Each scalar random field is obtained as (see (6))

$$\xi_k(s) = \int_{-\infty}^{s} g^{(k)}(s - x|\boldsymbol{\theta}) dW(x) \tag{14}$$

where $g^{(k)}(s - x|\boldsymbol{\theta})$ is a unilateral tail-down function with nonzero values only on the negative (i.e., downstream) side of s as in the tail-down model introduced in Section 2.2. Similarly as in the tail-up case, the trace-covariogram function for the functional tail-down process is straightforwardly obtained as

$$C(s_i, s_j) = \sum_{k\geq 1} C_d^{(k)}(s_i, s_j|\boldsymbol{\theta}), \tag{15}$$

where $C_d^{(k)}(s_i, s_j|\boldsymbol{\theta})$ is the covariance function associated to the kth scalar tail-down random field, whose expression may be of the kind presented in Table 1. In the light of expressions (13) and (15), one may wonder whether (and under which conditions) the families $C_t^{(k)}(h|\boldsymbol{\theta})$ and $C_d^{(k)}(s_i, s_j|\boldsymbol{\theta})$ are closed under conic combinations, i.e., if (and when) linear combinations, with positive weights, of valid covariance functions in the same parametric family still belong to the same family. If this was the case, the trace-covariogram of the functional process built in (9) would belong to the same family as those of the 1D processes $\{\xi_k(s), s \in D\}, k = 1, \ldots, N$. Concerning $C_t^{(k)}(h|\boldsymbol{\theta})$, it is well-known from scalar geostatistics that finite conic combinations of valid models are closed if and only if they belong to the same family and share the same range parameter. The same applies to $C_d^{(k)}(s_i, s_j|\boldsymbol{\theta})$, as can be straightforwardly derived from the expressions in Table 1. As such, if the scalar fields $\{\xi_k(s), s \in D\}, k = 1, \ldots, N$, share the same valid model and the same range parameter, the trace-covariogram of the functional process (9) will belong to the same family and share the same range as the scalar fields, but will have a sill equal to the sum of the sills of the scalar fields.

4. Model estimation and spatial prediction

4.1. Estimation of the spatial covariance under stationarity

In this work, we will estimate the parameters of the covariance models proposed in Section 3 by estimating the trace-semivariogram of the field, which is defined, under global second-order stationarity as $\gamma(s_1, s_2) = \frac{1}{2}\mathbb{E}[\|\mathcal{X}_{s_1} - \mathcal{X}_{s_2}\|^2]$ and is related with the trace-covariogram through the well-known relation $\gamma(s_1, s_2) = C(s_1, s_1) - C(s_1, s_2)$ (see, e.g., Menafoglio et al., 2013). As in scalar geostatistics, estimation of the trace-semivariogram can be performed by first determining an empirical estimator and then fitting a valid model. As discussed in Section 3, all the valid models in use in the scalar case can be adopted in the functional case too; for convenience, the semivariogram models derived by Ver Hoef et al. (2006) are reported in Table 1. The semivariograms in Table 1 are defined piecewise, depending on the connectedness of the pair being considered. From now on, the portion of a semivariogram associated to flow-connected (flow-unconnected) locations will be denoted as "the flow-connected (flow-unconnected) portion of the semivariogram".

4.1.1. Empirical semivariograms for stream networks

From Section 2 and the expressions in Table 1, it should be clear that, for both tail-up and tail-down models, the covariance structure among observations does not depend only on the stream distance but also on other characteristics such as flow connectedness, weights attributed to the stream segments, and/or distances to a common junction. This dependence – which motivates the use of the notation $\gamma(s_i, s_j|\boldsymbol{\theta})$ instead of $\gamma(s_i - s_j|\boldsymbol{\theta})$ – highlights the inadequacy in this context of the empirical semivariogram proposed in the functional Euclidean setting (see,e.g. Caballero et al., 2013; Menafoglio et al., 2013). In the scalar setting, Zimmerman and Ver Hoef (2017) propose and discuss modifications of the (scalar) empirical estimator to deal with stream networks and stream distances. They thus derive the *flow-unconnected stream-distance* (FUSD) semivariogram and the *flow-connected stream-distance* (FCSD) semivariogram, able to deal both with the peculiar topology of a stream network and with the stream distance. We here study functional counterparts of these estimators, eventually aiming to fit the parameters of a valid model to the most appropriate empirical estimator.

4.2. Flow-Unconnected Stream-Distance (FUSD) trace-semivariogram

The FUSD empirical trace-semivariogram is computed only from those site-pairs that are flow-unconnected and, for such pairs, it is a function of the stream distance only. The FUSD trace-semivariogram is thus defined as

$$\hat{\gamma}_{FUSD}(h_k) = \frac{1}{2|N(\mathcal{U}_k)|} \sum_{(s_i,s_j)\in N(\mathcal{U}_k)} \left\| \mathcal{X}_{s_i} - \mathcal{X}_{s_j} \right\|^2, \qquad k = 1, \dots, K_{\mathcal{U}}, \tag{16}$$

where $N(\mathcal{U}_k) = \{(s_i, s_j) : d(s_i, s_j) \approx h_k, U_{s_i} \cap U_{s_j} = \varnothing\}$ is the set of flow-unconnected pairs separated by a stream-distance approximately equal to h_k, and $|N(\mathcal{U}_k)|$ is its cardinality. Note that, if $\{\mathcal{X}_s, s \in D\}$ follows a pure tail-up model, the flow-unconnected portion of its semivariogram is constant and corresponds to the sill (see Table 1). In this case, $\hat{\gamma}_{FUSD}$ is an unbiased estimator for the flow-unconnected portion of the trace-semivariogram and an estimate of the sill is obtained as

$$\bar{\gamma}_{FUSD} = \frac{\sum_{k=1}^{K_{\mathcal{U}}} |N(\mathcal{U}_k)| \hat{\gamma}_{FUSD}(h_k)}{\sum_{k=1}^{K_{\mathcal{U}}} |N(\mathcal{U}_k)|}, \tag{17}$$

with $K_{\mathcal{U}}$ the number of bins in which the set of stream distances is partitioned. On the other hand, if $\{\mathcal{X}_s, s \in D\}$ follows a pure tail-down model, the flow-unconnected portion of its semivariogram in general does not depend on the total stream distance (i.e., $h = a + b$, see Fig. 2) but on the two stream distances from the sites within a site-pair to their common junction (i.e., a and b, see Fig. 2). Therefore, in this case, the FUSD empirical semivariogram is not enough to characterize the spatial dependence of flow-unconnected sites. It is worth noticing that an exception occurs if the tail-down component has an exponential semivariogram; indeed, in this case the flow-unconnected part of the semivariogram is a function of the total stream distance only; consequently, $\hat{\gamma}_{FUSD}(\cdot)$ remains unbiased for it.

4.3. Flow-Connected Stream-Distance (FCSD) trace-semivariogram

The FCSD trace-semivariogram differs from the FUSD trace-semivariogram by being computed from site-pairs that are flow-connected rather than flow-unconnected. Thus, it is defined as

$$\hat{\gamma}_{FCSD}(h_k) = \frac{1}{2|N(\mathcal{C}_k)|} \sum_{(s_i,s_j)\in N(\mathcal{C}_k)} \left\| \mathcal{X}_{s_i} - \mathcal{X}_{s_j} \right\|^2, \qquad k = 1, \dots, K_{\mathcal{C}}, \tag{18}$$

where $N(\mathcal{C}_k) = \{(s_i, s_j) : d(s_i, s_j) \approx h_k, U_{s_i} \cap U_{s_j} \neq \varnothing\}$, is the set of flow-connected pairs separated by a stream distance approximately h_k, and $|N(\mathcal{C}_k)|$ is its cardinality.

Note that now, if $\{\mathcal{X}_s, s \in D\}$ follows a pure tail-down model, the flow connected portion of its semivariogram is a function of the stream distance h between locations. In this case, the well-known valid models in use in scalar geostatistics can be used for parametric modelling (see Table 1). Moreover, similarly as for the scalar case, $\hat{\gamma}_{FCSD}(\cdot)$ is an unbiased estimator for the flow-connected portion of the trace-semivariogram of the process. On the other hand, if $\{\mathcal{X}_s, s \in D\}$ is a pure tail-up process, $\hat{\gamma}_{FCSD}(\cdot)$ is not fully appropriate to estimate its covariance structure, because in those cases the flow-connected portion of the trace-semivariogram is a function of the stream distance and of the spatial weights (see Table 1). In Section 4.3.1 we will further expand on this point.

4.3.1. Parameters estimation

In this section we outline the steps to estimate a parametric model for the trace-semivariogram whenever the underlying process that generated the data is assumed to follow either a *pure tail-up* or a *pure tail-down* model. However, note that, in general, mixtures of tail-up and tail-down models may arise and these are, in principle, more flexible to describe the spatial dependence. The extension of the proposed procedure to the case of mixture models will be discussed in Section 7. The identification of the underlying process – based on empirical trace-semivariograms – is a rather hard task. Inspired by the strategy for fluvial variography developed in Zimmerman and Ver Hoef (2017), we propose the following procedure to estimate a valid trace-semivariogram model $\gamma(\cdot, \cdot|\boldsymbol{\theta})$, and the associated trace-covariogram $C(\cdot, \cdot|\boldsymbol{\theta})$.

(i) Estimate the empirical FCSD trace-semivariogram $\hat{\gamma}_{FCSD}(h_k)$, $k = 1, \dots, K_{\mathcal{C}}$. from the observations $x_{s_1}, \dots, x_{s_n}$ using (18).
(ii) Estimate the empirical FUSD trace-semivariogram $\hat{\gamma}_{FUSD}(h_k)$, $k = 1, \dots, K_{\mathcal{U}}$. from the observations $x_{s_1}, \dots, x_{s_n}$ using (16).
    a. If $\hat{\gamma}_{FUSD}(h_k)$, $k = 1, \dots, K_{\mathcal{U}}$, is compatible with a pure nugget model, assume the process to be pure tail-up.
    b. If $\hat{\gamma}_{FUSD}(h_k)$, $k = 1, \dots, K_{\mathcal{U}}$, is not compatible with a pure nugget model, assume the process to be pure tail-down.
(iii) Fit a valid model $\gamma(\cdot, \cdot|\boldsymbol{\theta})$ to the empirical FCSD trace-semivariogram $\hat{\gamma}_{FCSD}(h_k)$ and get $\hat{\boldsymbol{\theta}}$.
(iv) Obtain the trace-covariogram as $C(\cdot, \cdot|\hat{\boldsymbol{\theta}})$ plugging $\hat{\boldsymbol{\theta}}$ in the stream-network trace-covariograms expressions in Table 1.

Whenever the underlying process can be assumed to follow a pure tail-down model (see point (ii) above) the outlined procedure can be applied without hindrance. This follows from the fact that the flow-connected portion of the tail-down trace-semivariograms in Table 1 has exactly the same expression as the classical models. In case of a pure tail-up model, instead, the approach presents limitations, because the FCSD trace-semivariogram does not account for the spatial weights $\pi_{i,j}$ in (13). Note that step (iii) consists of fitting a standard valid model, and neglects the weights $\pi_{i,j}$ within the variograms of Table 1. These classical geostatistical models are recovered when including weights $\pi_{i,j} = 1$ in the expressions of the theoretical semivariogram in Table 1; in the following, the semivariogram associated with weights $\pi_{i,j} = 1$ will be denoted as *unweighted flow-connected semivariogram*. In fact, the empirical trace-semivariogram (18) is used in step (iii) precisely as an estimator of the unweighted flow-connected semivariogram. Nevertheless, the FCSD semivariogram is a biased estimator for the unweighted flow-connected semivariogram (see Appendix B). As shown in the simulation study presented in the Supplementary Material, this bias may adversely affect the analyst's ability to correctly determine the range of spatial dependence among flow connected sites (i.e., the range estimates tend to be negatively biased). A similar problem is discussed by Zimmerman and Ver Hoef (2017), who eventually proposed an adjusted empirical estimator (FCWA), which accounts for the weights and is unbiased for the unweighted flow-connected semivariogram. A modification of the empirical trace-semivariogram (18) that follows the same line of Zimmerman and Ver Hoef (2017) (named FCWA2), is developed and studied via simulation in the Supplementary Material. These developments are not reported in the outlined procedure because, despite their unbiasedness, the adjusted estimators proved to be characterized

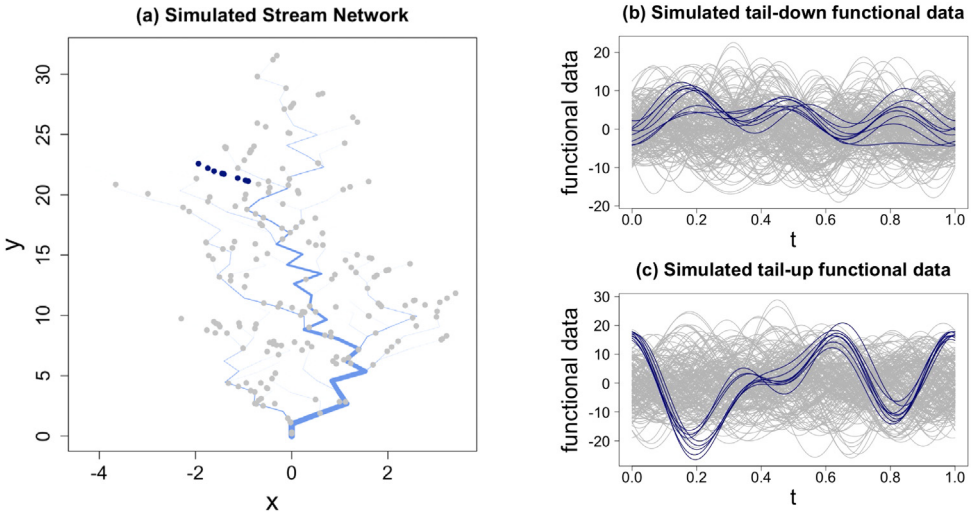


Fig. 3. (a) Simulated stream network. Points indicate the locations of observed data. (b) Simulated functional data for the tail-down model. (c) Simulated functional data for the tail-up model.

by extremely high variance, hindering their use in practice. The same simulations show that the range underestimation does not heavily affect the Kriging performances, thus suggesting that the use of FCSD trace-semivariogram should be anyway preferred to its adjusted (unbiased) version.

4.4. Estimating the spatial dependence in the non-stationary case

The methods developed so far assume second-order stationarity. If stationarity does not hold, we propose to use the non-stationary model of Menafoglio et al. (2013), which decouples the elements of the random field $\{\mathcal{X}_s : s \in D\}$ in a non-stationary mean term m_s (the drift) and a zero-mean stationary residual component δ_s, i.e., $\mathcal{X}_s = m_s + \delta_s$. In the following, we use for the stochastic residual δ_s the models built in Section 3, and denote by $C(s_1, s_2|\boldsymbol{\theta})$, $\gamma(s_1, s_2|\boldsymbol{\theta})$ the trace-covariogram and trace-semivariogram of δ_s, respectively. A model for the drift term is needed to allow for the estimate of $C(\cdot,\cdot|\boldsymbol{\theta})$, $\gamma(\cdot,\cdot|\boldsymbol{\theta})$, as these are assessed from the (estimated) residuals. We consider a linear model, i.e.,

$$\mathcal{X}_s = \sum_{l=0}^{L} a_l f_l(s) + \delta_s \quad s \in D, \tag{19}$$

where $f_0(\boldsymbol{s}) = 1$ for all $s \in D$, $f_l(\cdot)$, $l = 1, \ldots, L$, are known functions of the spatial variable $s \in D$ and $a_l(\cdot) \in H$, $l = 0, \ldots, L$, are functional coefficients independent of the spatial location. Estimation of the linear model (19) can follow the very same lines as in the case of a Euclidean spatial domain, broadly discussed by Menafoglio et al. (2013). These authors propose a generalized least-squares (GLS) estimator for the coefficients a_l, based on the covariance matrix of the residuals $\delta_{s_1}, \ldots, \delta_{s_n}$. The very same procedure can be used in our setting, provided that the covariance matrix Σ is interpreted in terms of the stream-network trace-covariogram $C(\cdot,\cdot|\boldsymbol{\theta})$, i.e., $\Sigma_{i,j} = C(s_i, s_j|\boldsymbol{\theta})$. The iterative algorithm for estimating the model parameters in the non-stationary case is deferred to Appendix C.

4.5. Kriging prediction

Having estimated the model, spatial prediction at a target site s_0 in D can be performed by using the theory of object-oriented kriging presented in Menafoglio et al. (2013) (see also (Menafoglio

and Secchi, 2017) for a recent review). In this setting, the kriging predictor is defined via a linear combination of the data that have been observed: $\mathcal{X}^*_{s_0} = \sum_{i=1}^n \lambda_i^* \mathcal{X}_{s_i}$. The weights $\lambda_1^*, \dots \lambda_n^* \in \mathbb{R}$ are found as to minimize the global variance of the prediction error under the unbiasedness constraint, i.e.,

$$(\lambda_1^*, \dots \lambda_n^*) = \underset{\substack{\lambda_1,\dots\lambda_n \in \mathbb{R}: \\ \mathcal{X}^{\boldsymbol{\lambda}}_{s_0} = \sum_{i=1}^n \lambda_i \mathcal{X}_{s_i}}}{\operatorname{argmin}} \quad Var(\mathcal{X}^{\boldsymbol{\lambda}}_{s_0} - \mathcal{X}_{s_0}) \quad \text{subject to} \qquad \mathbb{E}[\mathcal{X}^{\boldsymbol{\lambda}}_{s_0}] = m_{s_0}. \tag{20}$$

Using the general model (19) – which clearly reduces to the stationary case when $L = 0$ – the global optimum of problem (20) is obtained by solving the following linear system

$$\begin{pmatrix} \Sigma & \mathbb{F} \\ \mathbb{F}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}_0 \\ \boldsymbol{f}_0 \end{pmatrix} \tag{21}$$

where $\mathbb{F}$ is the design matrix of the linear model (19), $\boldsymbol{\lambda}$ is the vector of weights, $\boldsymbol{\mu}$ the vector of Lagrange multipliers, $\boldsymbol{\sigma}_0$ the vector containing the cross-covariances between observations and the target ($\sigma_0^{(i)} = C(s_i, s_0|\boldsymbol{\theta})$), $\boldsymbol{f}_0$ the design vector at the target site ($f_0^{(l)} = f_l(s_0)$).

5. Two simulated examples

The procedure outlined in Section 4.3.1 is here applied to two simulated examples, one for the tail-up case and one for the tail-down case. In both examples, we consider the stream network domain D represented in Fig. 3a, characterized by 250 segments and $n = 200$ observation points; this was generated using the `SNN` package (Ver Hoef et al., 2014) in `R` (R Core Team, 2020). Zero mean functional random processes are simulated by exploiting the construction

$$\mathcal{X}_s = \sum_{k=1}^{N} \xi_k(s) e_k, \tag{22}$$

which is analogous to (9), with $m = 0$. Here, $\{e_k, k \geq 1\}$ denotes the Fourier orthonormal basis of $H = L^2([0, 1])$, and N is set to $N = 7$. Parameters for the generation of the scalar fields $\{\xi_k(s), s \in D\}$, $k = 1, \dots, N$, are specific of the examples, and are detailed below. The fields $\{\xi_k(s), s \in D\}$, $\{\xi_j(s), s \in D\}$ are assumed to be independent for $j \neq k$; each $\{\xi_k(s), s \in D\}$ is finally assumed to be Gaussian.

5.1. Estimating the trace-covariogram in a pure tail-down model

In this example, for each field $\{\xi_k(s), s \in D\}$ appearing (22), a **tail-down exponential model** is used with sill $\theta_v^{(k)}$, range $\theta_r^{(k)}$ and nugget $\eta^{(k)}$ parameters set to $(\theta_v^{(k)}, \theta_r^{(k)}, \eta^{(k)}) = (5, 6.5, 0)$, respectively. Therefore the theoretical model for the functional process (22) is a tail-down exponential model with parameters $(\theta_v, \theta_r, \eta) = (35, 6.5, 0)$ (see the final remarks in Section 3.2). The functional dataset in Fig. 3b was obtained by combining the realizations of the $N = 7$ scalar random fields sampled at the $n = 200$ locations displayed in Fig. 3a. Fig. 4a displays the FCSD and FUSD empirical estimators (see Eq. (18) and (16)), which share – as expected – the same non-trivial structure of spatial dependence. Both trace-semivariograms were obtained considering 15 lags and a maximum distance equal to half the maximum distance in the stream network.

Although, in general, the flow-connected portion of the trace-semivariogram (blue circles) is the only one that should be considered to retrieve the parameter estimates (see the strategy outlined in Section 4.3.1), we may here consider also the FUSD for the purpose, since – for the special case of an exponential tail-down model – it is unbiased for the flow-unconnected portion of the trace-semivariogram (see Section 4.1.1 and Table 1). Fitting the trace-semivariogram parameters to the FCSD yields accurate estimates, $(\hat{\theta}_v, \hat{\theta}_r) = (34.57, 6.73)$, very close to the reference values ($\theta_v = 35$ and $\theta_r = 6.5$). Nevertheless, the estimates obtained via FUSD are affected by a slight overestimation of both the sill and the range $(\hat{\theta}_v, \hat{\theta}_r) = (38.16, 8.26)$. This overestimation may be due to the variability of the FUSD estimator.

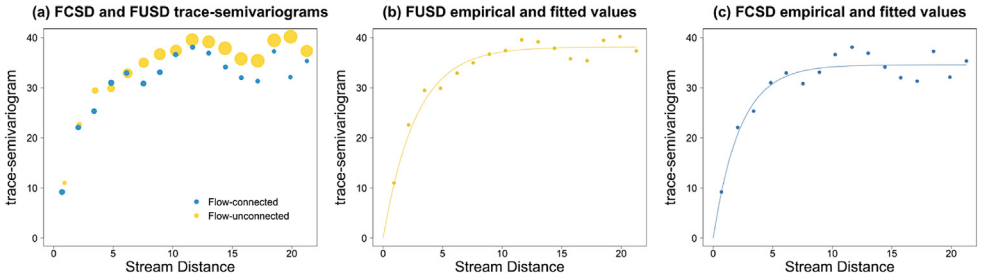


Fig. 4. (a) Empirical FCSD trace-semivariogram (blue) and FUSD trace-semivariogram (yellow). The dot's sizes are proportional to the number of pairs for each binned distance class. (b) Empirical and fitted FUSD trace-semivariograms. (c) Empirical and fitted FCSD trace-semivariograms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.2. Estimating the trace-covariogram in a pure tail-up model

We here consider a functional tail-up model, built as in (22), and assuming for each field $\{\xi_k(s), s \in D\}$ a tail-up spherical model with sill $\theta_v^{(k)}$, range $\theta_r^{(k)}$ and nugget $\eta^{(k)}$ parameters set to $(\theta_v^{(k)}, \theta_r^{(k)}, \eta^{(k)}) = (10, 8.5, 0)$, respectively. Therefore, the resulting functional process (22) is again a tail-up spherical model, but characterized by the parameters $(\theta_v, \theta_r, \eta) = (70, 8.5, 0)$. The functional dataset in Fig. 3c was obtained by combining the $N = 7$ independent realizations of the scalar random fields at the $n = 200$ sampling locations in D, with the first $N = 7$ elements of the Fourier basis $\{e_k, k = 1, \dots, 7\}$. The empirical FUSD trace-semivariogram (16) and the empirical FCSD trace-semivariogram (18) are displayed in Fig. 5a. The FUSD semivariogram appears to be flat, as expected in a pure tail-up model, and it can be used to unbiasedly estimate the variogram sill via expression (17). On the contrary, the flow connected pairs are featured by a non-trivial spatial dependence. Indeed, the FCSD semivariogram exhibit a clear downward concavity near the origin, settling towards a sill not far from the value of the FUSD semivariogram. To retrieve estimated parameters, we fit a spherical model to the FCSD, as shown in Fig. 5b. This leads to the following parameters estimates: $(\hat{\theta}_v, \hat{\theta}_r) = (67.87, 5.53)$. Note that, while the estimated sill is close to the reference value $\theta_v = 70$, the range is underestimated, the reference value being $\theta_r = 8.5$. This tendency is confirmed by the results of the simulation study in the Supplementary Material, and it is due to the fact that FCSD trace-semivariogram neglects the weights π_{ij} appearing in (13), but the simulation process clearly accounts for them. Finally, the sill estimated from the average of the FUSD trace-semivariogram (see Eq. (17)) is $\hat{\theta}_v = 68.35$, again rather close to the reference value. we broadly explored, by simulation, the performance of FCSD trace-semivariogram to estimate the spatial dependence of the field; we also assessed the effect of its bias on the results of Kriging prediction, which appears to be negligible. The results are deferred to the Supplementary Material for brevity of exposition.

6. Analysis of Middle Fork river water temperatures

6.1. Middle Fork river and dataset

In this section, the proposed methods are employed in the analysis of the water temperature profiles of the Middle Fork river. The data consist of the maximum daily water temperatures recorded between 15 July 2005 and 31 August 2005 at several locations of the Middle Fork river, a 104-mile-long (167 km) river in central Idaho, USA (Fig. 1). The data, which can be found and downloaded at https://www.fs.fed.us/rm/boise/AWAE/projects/SSN_STARS/software_data.html, has been pre-processed and created as part of an NCEAS Workshop in April 2011 (National Center for Ecological Analysis and Synthesis). The daily maximum water temperature (in $^\circ C$) have been recorded at each of the 157 locations for 47 days in the aforementioned summer period (15 July

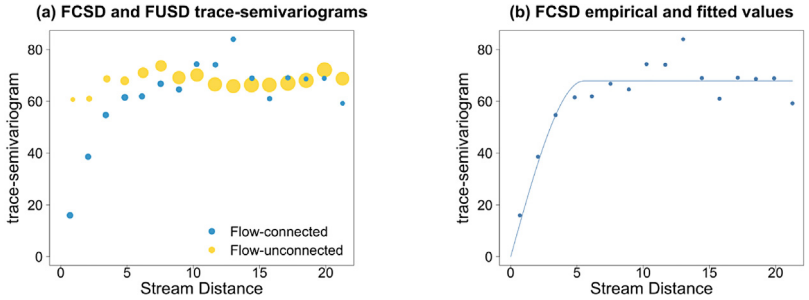


Fig. 5. (a) Empirical FCSD trace semivariogram and FUSD trace semivariogram. The dots sizes are proportional to the number of pairs for each class of distances. (b) Empirical FCSD and fitted trace-semivariogram.

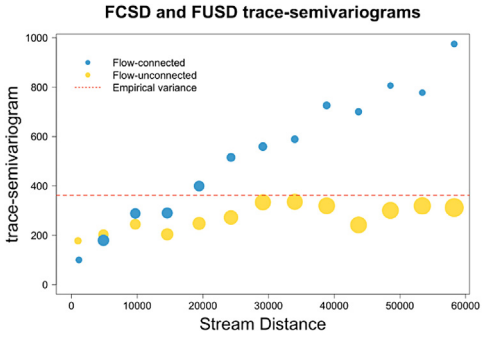


Fig. 6. Trace-semivariograms for the Middle Fork temperatures. Empirical estimates of the flow-connected and flow-unconnected trace-semivariograms; the empirical variance is reported as a dashed red line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2005–31 August 2005). We embed the data in the Hilbert Space $H = L^2$ of the square integrable functions endowed with the usual scalar product. The raw data (Fig. 1b) were smoothed via spline smoothing with a roughness penality (Fig. 1c). The number of basis functions ($nb = 49$) and the smoothing parameter ($\lambda = 5$) were chosen through a leave-one-out cross validation approach to avoid overfitting. The watershed area accumulated downstream (km^2), was used to compute the weights for the tail up models, as a proxy variable for flow volume (see Ver Hoef and Peterson (2010)). Two covariates, namely the elevation of the upper stream segment node on which a temperature sensor was located (m) and the upstream distance between the stream outlet and the site (m), were used to model the drift term.

6.2. Geostatistical analysis

The stationarity of the random field is evaluated from the empirical trace-semivariograms, computed by considering 13 distance classes with bins of equal size up to a maximum distance of 63.11 km (Fig. 6). Visual inspection of the trace-semivariograms suggest that a non-stationarity assumption is appropriate for the random field, since the FCSD trace-semivariogram (blue dots in Fig. 6) seems to increase without bound, beside crossing the flow unconnected trace semivariogram (yellow dots in Fig. 6), indicating a trend contamination aligned with flow. This behaviour (crossing components and unbounded growth) is indeed evidence of an unmodelled drift in upstream distance (Zimmerman and Ver Hoef, 2017). Following the approach devised in Section 4.4, a drift term is thus included in the model. We here consider as covariate for the drift term the variables $\{x, y\}$ ={elevation, distance upstream}, which are appropriate to describe a drift term aligned with

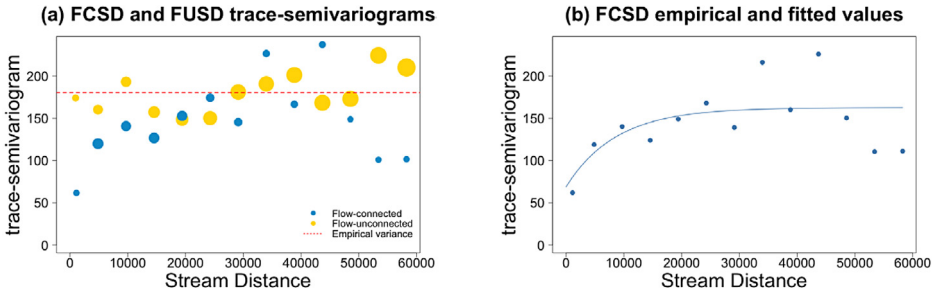


Fig. 7. (a) Empirical trace semivariogram of the residuals obtained with an OLS estimate of the drift. (b) Empirical and fitted trace-semivariograms of the residuals obtained with WLS.

flow (see Zimmerman and Ver Hoef (2017) for the scalar case). For the selection of the functional form for the drift, we follow the approach of Menafoglio et al. (2013), who suggest to consider polynomial forms for the drift term, and select the optimal one through cross-validation. Here, each candidate model is evaluated in terms of kriging performances, quantified through the sum of squared errors

$$SSE_i = \|\mathcal{X}_{s_i} - \mathcal{X}^*_{s_i}\|^2, \quad i = 1, \dots, n, \tag{23}$$

where $\mathcal{X}^*_{s_i}$ stands for the kriging prediction of $\mathcal{X}_{s_i}$ when this is left out of the sample. At this stage, a simplified version of the Universal Kriging predictor with a covariance structure of pure nugget is employed, thus providing the prediction which would have been obtained via FDA linear models (indeed in this case the UK predictor reduces to the drift estimate). We thus considered as candidate models the 31 polynomials of order lower than 2 (excluding the case of the sole intercept as drift). For each candidate drift, the empirical trace-semivariograms of the residuals are computed to verify that the optimal drift model selected according to the introduced criterion gives rise to a stationary residual. The best model is the following:

$$m(s,t) = a_0(t) + a_1(t)x + a_2(t)y + a_3(t)x^2 + a_4(t)y^2 + a_5(t)xy, \quad t \in \tau = [1, 47]. \tag{24}$$

Fig. 7a reports the empirical FCSD and FUSD trace-semivariogram of the residuals of model (24), when these are estimated via OLS. The shape of the flow connected trace-semivariogram is not that of a pure nugget, suggesting that the residuals are spatially correlated. On the other hand, the flow unconnected empirical semivariogram is compatible with a pure nugget model, suggesting the use of a tail-up model for the field (see Section 4.3.1). In the following, we shall thus consider an exponential tail-up model with sill θ_v, range θ_r and nugget η (see Table 1).

Having chosen the drift and the covariance models, the model parameters are estimated by means of the generalized least square criterion outlined in Section 4.4 and in Appendix C. Fig. 7b displays the FCSD empirical trace-semivariogram together with the fitted variogram model, characterized by estimated parameters: $(\hat{\theta}_v, \hat{\theta}_r, \hat{\eta}) = (68.83, 25885.44, 93.59)$. Note that, as broadly discussed in Section 4.3.1, interpretation of $\hat{\theta}_r$ requires particular care, as it could be affected by a negative bias. Kriging is eventually performed at a grid of new locations along the stream network, by plugging-in the estimated parameters $\hat{\boldsymbol{\theta}} = (\hat{\theta}_v, \hat{\theta}_r, \hat{\eta})'$ in the linear system (21). To evaluate the performance of the Kriging predictor, a leave-one-out cross validation (LOOCV) procedure is applied, considering as measure of discrepancy between the true value and the predicted one the SSE defined in (23) and the relative *SSE* performance index, given by $SSE_i^{(rel.)} = \frac{SSE_i}{\|\mathcal{X}_{s_i}\|^2}$. The statistics shown in Table 2 prove the satisfactory forecasting performance of the method. Fig. 8a displays with colours the SSE_i for each location on the river; a part for a few locations associated with a high estimation error ($SSE > 700$, red dots), which possibly mark influential/outlying data, the kriging predictor works properly. Fig. 9 shows the original data (Fig. 9a) together with the UK estimates (Fig. 9b) and the corresponding kriging residuals (Fig. 9c). Note that the significant reduction of the total SSE,

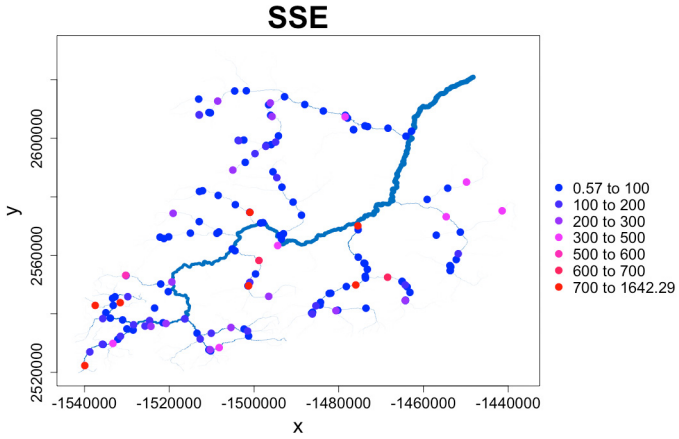


Fig. 8. SSE leave-one-out error for each location on the Middle Fork River.

Table 2
Summary indices of the distribution of SSE and $SSE^{(rel)}$.

	SSE	$SSE^{(rel)}$
Min	0.574	$3.44 \cdot 10^{-5}$
Median	43.346	$4.19 \cdot 10^{-3}$
Mean	132.897	$1.6 \cdot 10^{-2}$
Sum	20 864.89	2.52

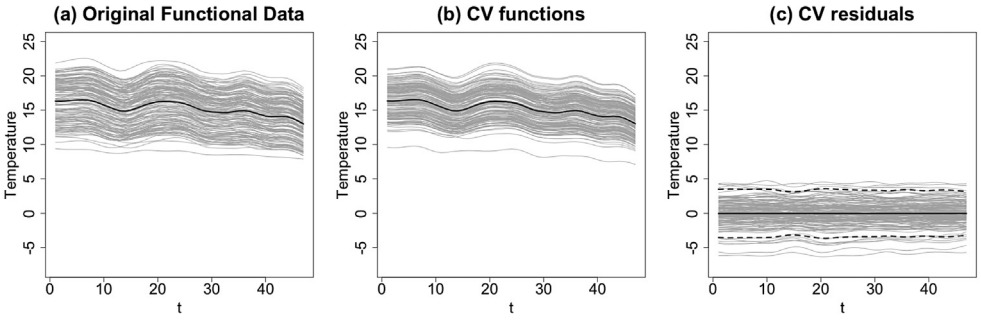


Fig. 9. Cross-validation analysis. (a) Original data and relative sample mean (bold). (b) Data predicted via Universal Kriging and their mean (bold). (c) Difference between original and predicted data, their mean $\hat{m}_r$ (in black) and the (approximate) point-wise confidence band $m_r(z) + 2\hat{\sigma}_r(z)$(dashed), where $\hat{\sigma}_r(z)$ is the (point-wise) standard deviation estimated from the cross-validation residual.

$SSE = \sum_{i=1}^{n} SSE_i$, attained with a tail-up covariance structure ($SSE = 20864.89$) as opposed to a pure nugget ($SSE = 29485.93$), confirms the ability of the former to capture in a greater extent the stochastic variability of the residual process. Finally, Fig. 10 provides a representation of the observed (Fig. 10a) and predicted (via LOOCV, Fig. 10b) average temperatures (the average being taken over the summer period) together with the corresponding marginal distributions (Fig. 10c). Cross-validation results exhibit a narrower range of values than the data and this is a sign of the Kriging smoothing effect. Comparison between Fig. 10a and b confirms the validity of the proposed method, which is able to reproduce the main spatial patterns in the data.

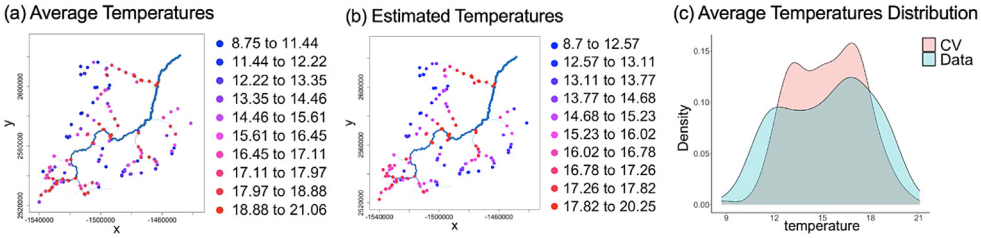


Fig. 10. (a): Observed average temperatures for the locations on the Middle Fork River. (b): Estimated average temperatures via leave-one-out Universal Kriging for the Middle Fork data. (c): Distributions of the observed average temperatures together with the cross-validation average temperatures.

7. Conclusion

In this work new geostatistical methods for complex data distributed over a stream network have been proposed. First of all, the theoretical construction presented in Section 3 allows one to develop a strategy for variographic analysis and estimation of the spatial covariance structure, which proved to be effective in terms of prediction performance for both tail-up and tail-down models. When tested on real data, the methodology achieved a good prediction performance. However, further research should be carried out for tail-up models, aiming at a better empirical estimator of the semivariogram that takes into account the weights characterizing the river topology. This could potentially allow one to achieve better estimates for the range of correlation of the field. In addition, extension to mixed models should certainly be the object of future works. To this end, an adaptation of the procedure proposed in this work is envisioned as a nested structure for the fitting of the FCSD trace-semivariogram. Further research may also be devoted to relax the hypothesis of homogeneous covariance structure when dealing with large stream networks. In fact, allowing varying dependence structures on different sub-networks could lead to interesting developments in the direction of modelling strong spatial non-stationarities. Concerning the topological description of the stream network, one should note that the current work, as well as the literature focused on the scalar case, only enables one to analyse data over one-dimensional stream segments. However, allowing the representation of stream segments to be also equipped with information about their depth and thickness, might notably enrich the geostatistical analysis, especially in cases where the stream network includes large sub-streams and lakes. The application of the developed models to contexts other than those of a river network is possible, and would be extremely topical in contexts like electricity grids, traffic and transportation systems, and road networks. Indeed, whenever it is advisable to define the stream distance with respect to the topology of a network featured by the presence of a flow rather than based on a Euclidean distance, the proposed approach should be considered. Here, extensions of the considered class of models will deserve further research to allow for the analysis of data distributed over non-binary trees and general networks for which valid covariance models are yet to be studied.

Finally, an alternative viewpoint to the prediction problem faced in this work could consist of understanding the data as a time series of tree-structured data objects. In this setting, predicting the space–time (temperature) field would consist of reconstructing the target variable within each tree (e.g., through scalar methods), and then building a time-dependent (predictive) model for the tree-structured objects, properly accounting for their temporal dependence. To the authors' knowledge, this latter goal would need a significant methodological effort to be attained. While for the particular case study which stimulated our research the approach pursued in this work seems natural, this alternative viewpoint certainly deserves to be further investigated, and might open new venues in the context of object oriented time-series modelling.

Acknowledgements

Alessandra Menafoglio and Piercesare Secchi acknowledge the support by MUR, grant Dipartimento di Eccellenza 2023–2027.

Appendix A. Infinite dimensional functional process

We here discuss the conditions that allows one to consider an infinite-dimensional construction for the functional process, i.e., to represent the element $\mathcal{X}_s$ as the limit as $N \to \infty$ of (11) and obtain a well-defined global covariance function. Therefore, we formally define the functional random field as

$$\mathcal{X}_s = \sum_{k=1}^{\infty} \xi_k(s) e_k, \tag{A.1}$$

where $\{e_k\}$ is an orthonormal basis of H and $\{\xi_k(s), s \in D\}$, $k \geq 1$, denote independent, zero-mean, second-order stationary and isotropic scalar random fields, defined through the moving average construction described in Section 2.

A minimal assumption for the *existence* of the cross covariance operators C_{s_i,s_j} is that the random variables $\mathcal{X}_s$ have finite second moments, i.e., $\mathbb{E}[\|\mathcal{X}_s^2\|] < \infty$, for all s in D. For the functional random process (A.1) this is equivalent to the following condition

$$\sum_{k=1}^{\infty} \mathbb{E}[\xi_k(s)^2] = \sum_{k=1}^{\infty} \int_{-\infty}^{+\infty} \left(g^{(k)}(x - s|\boldsymbol{\theta})\right)^2 dx < \infty, \quad \text{for all } s \in D. \tag{A.2}$$

In this context, the dependence of $g(\cdot|\boldsymbol{\theta})$ on the parameters $\boldsymbol{\theta}$ plays an important role since one should impose conditions on $\boldsymbol{\theta}$ to ensure the finiteness of the series. It is worth highlighting that the autocovariance function of each scalar random field $\{\xi_k(s), s \in D\}$ admits a compact form

$$\mathbb{E}[\xi_k(s)\xi_k(s+h)] = C^{(k)}(h|\boldsymbol{\theta}) = \theta_v^{(k)} \rho_k(h/\theta_r^{(k)}), \tag{A.3}$$

where $\rho_k(\cdot)$ is a positive correlation function that depends on the type of moving average function of the kth random field. Plugging-in (A.3) in (A.2), we get

$$\sum_{k=1}^{\infty} \mathbb{E}[\xi_k(s)^2] = \sum_{k=1}^{\infty} \int_{-\infty}^{+\infty} \left(g^{(k)}(x - s|\boldsymbol{\theta})\right)^2 dx = \sum_{k=1}^{\infty} C^{(k)}(0|\boldsymbol{\theta}) = \sum_{k=1}^{\infty} \theta_v^{(k)}$$

Therefore $\mathcal{X}_s$ belongs to $L^2(\Omega; H)$ provided that we assume the summability of the series of $\theta_v^{(k)}$, i.e.,

$$\sum_{k=1}^{\infty} \theta_v^{(k)} < \infty. \tag{A.4}$$

Under condition (A.4), one can prove by direct computations that $\theta_v^{(k)}$ are the eigenvalues of the covariance operator $\mathcal{C}_{s,s}$. Moreover, under the square integrability assumption, the cross-covariance operator $\mathcal{C}_{s_i,s_j}$ exists and it is a symmetric trace-class Hilbert–Schmidt operator (Bosq, 2000). Finally, its trace is well defined by $\sum_{k=1}^{\infty} \langle C_{s_i,s_j} e_k, e_k \rangle$, as the series converges absolutely for every orthonormal basis in H and the sum does not depend on the choice of the basis (Zhu, 2007). The identity

$$C(s_i, s_j) = \sum_{k=1}^{\infty} \langle C_{s_i,s_j} e_k, e_k \rangle$$

can be proved as in Menafoglio et al. (2013), by exploiting the Parseval identity and the Lebesgue's dominated convergence theorem for series. Note that to apply the latter theorem, the requirement $\mathbb{E}[\|\mathcal{X}_s\|^2] < \infty$ is crucial.

Appendix B. Bias of the FCSD empirical estimator

We here discuss on the bias of the FCSD empirical estimator proposed in Section 4.1.1 when $\{\mathcal{X}_s, s \in D\}$ is represented by a tail-up model. Recall that the expression of the trace-semivariogram expression for tail-up models is

$$\gamma(s_i, s_j) = \begin{cases} 0 & \text{if } s_i = s_j \text{ (i.e if } h = 0) \\ \theta_v & \text{if } s_i \text{ and } s_j \text{ are not flow connected} \\ \theta_v - \pi_{i,j} C_t(h|\boldsymbol{\theta}) & \text{otherwise.} \end{cases} \tag{B.1}$$

Here, $C_t(h|\boldsymbol{\theta})$ is the trace-covariogram for the moving average functional process on the real line, which is related to the *unweighted flow-connected trace-semivariogram* $\gamma(h)$ by the relation

$$C_t(h|\boldsymbol{\theta}) = C_t(0) - \gamma(h) = \theta_v - \gamma(h). \tag{B.2}$$

Plugging-in (B.2) in (B.1) we get

$$\gamma(s_i, s_j) = \begin{cases} 0 & \text{if } s_i = s_j \text{ (i.e if } h = 0) \\ \theta_v & \text{if } s_i \text{ and } s_j \text{ are not flow connected} \\ \theta_v - \pi_{i,j}\big(\theta_v - \gamma(h)\big) & \text{otherwise.} \end{cases} \tag{B.3}$$

From this expression we can easily see that, in the absence of weights (i.e., setting $\pi_{i,j} = 1$) $\gamma(s_i, s_j)$ effectively would correspond to $\gamma(h)$, which is targeted by FCSD. Concerning the flow-connected portion of expression (B.3), i.e., focusing on locations flow-connected s_i, s_j, one has

$$\frac{1}{2}\operatorname{Var}_H(\mathcal{X}_{s_i} - \mathcal{X}_{s_j})_H = \theta_v - \pi_{i,j}\theta_v + \pi_{i,j}\gamma(h), \tag{B.4}$$

that, rearranging the terms, reads

$$\gamma(h) = \theta_v + \frac{1}{2\pi_{i,j}}\operatorname{Var}_H(\mathcal{X}_{s_i} - \mathcal{X}_{s_j}) - \frac{1}{\pi_{i,j}}\theta_v.$$

Given that the FCSD empirical estimator (18) is an unbiased estimator for $\frac{1}{2}\operatorname{Var}_H(\mathcal{X}_{s_i} - \mathcal{X}_{s_j})$, it straightforwardly follow that it is biased for the unweighted flow-connected semivariogram $\gamma(h)$, unless $\pi_{i,j} = 1$ for all i, j. Unbiased estimators named FCWA and FCWA2 are derived and studies in the Supplementary Material, adjusting for the bias of the FCSD according to Eq. (B.4).

Appendix C. Drift estimation

We here briefly recall the procedure which can be used to estimate the linear model in (19). The model for the vector of observations $\boldsymbol{\mathcal{X}} = (\mathcal{X}_{s_1}, \dots, \mathcal{X}_{s_n})^T$ can be expressed as

$$\boldsymbol{\mathcal{X}} = \mathbb{F}a + \boldsymbol{\delta}, \tag{C.1}$$

where $\boldsymbol{a} = (a_0, \dots, a_L)^T$ is the vector of (functional) coefficients, $\boldsymbol{\delta} = (\delta_{s_1}, \dots, \delta_{s_n})^T$ is the random vector of spatially-correlated residuals and $\mathbb{F} \in \mathbb{R}^{n\times(L+1)}$ is the design matrix, i.e., $\mathbb{F}_{i,l} = (f_l(s_i))$. Menafoglio et al. (2013) propose to estimate the functional coefficients $\boldsymbol{a}$ given $\boldsymbol{\mathcal{X}}$ based on a generalized least square criterion (GLS) with weighting matrix Σ^{-1}, i.e., the inverse of the $n \times n$ covariance matrix Σ of $\boldsymbol{\mathcal{X}}$. Since $\hat{\boldsymbol{a}}^{GLS}$ depends itself on Σ, which is usually unknown, the following iterative algorithm, can be used for its actual computation.

Algorithm 1 (*Menafoglio et al., 2013*)**.** Given a realization $\boldsymbol{x} = (x_{s_1}, \dots, x_{s_n})$ of $\boldsymbol{\mathcal{X}}$, represented as in (19):

1. Estimate the drift vector $\boldsymbol{m}$ through the OLS method and set $\hat{\boldsymbol{m}} = \hat{\boldsymbol{m}}^{OLS}$, with $\hat{\boldsymbol{m}}^{OLS} = \mathbb{F}(\mathbb{F}^T\mathbb{F})^{-1}\mathbb{F}^T\boldsymbol{x}$.
2. Compute the residual estimate $\hat{\boldsymbol{\delta}} = (\hat{\delta}_{\boldsymbol{s}_1}, \dots, \hat{\delta}_{\boldsymbol{s}_n})$ by difference: $\hat{\delta} = \boldsymbol{x} - \hat{m}$.

3. Estimate the trace-semivariogram $\gamma(\cdot,\cdot)$ of the residual process $\{\boldsymbol{\delta}_s, s \in D\}$ from $\hat{\delta}$ first with the FCSD empirical estimator (18) and then fitting to this a valid model $\gamma(\cdot;\boldsymbol{\theta})$, obtaining $\hat{\boldsymbol{\theta}}$. Plug-in $\hat{\boldsymbol{\theta}}$ in the stream-network trace-covariogram expression of Σ (see Table 1, $\Sigma_{i,j} = C(s_i, s_j|\boldsymbol{\theta})$) yielding $\hat{\Sigma}$ (with $\hat{\Sigma}_{i,j} = C(s_i, s_j|\hat{\boldsymbol{\theta}})$).
4. Estimate the drift vector $\boldsymbol{m}$ with $\hat{\boldsymbol{m}}^{GLS}$, obtained from x using: $\hat{\boldsymbol{m}}^{GLS} = \mathbb{F}(\mathbb{F}^T \Sigma^{-1}\mathbb{F})^{-1}\mathbb{F}^T \Sigma^{-1}\boldsymbol{x}$.
5. Repeat 2.-4. until convergence.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spasta.2023.100784.

References

Bosq, D., 2000. Linear Processes in Function Spaces. Springer-Verlag, New York, NY.
Caballero, W., Giraldo, R., Mateu, J., 2013. A universal kriging approach for spatial functional data. Stochast. Environ. Res. Risk Assess. 27, 1553–1563.
Cressie, N., Frey, J., Harch, B., Smith, M., 2006. Spatial prediction on a river network. J. Agric. Biol. Environ. Stat. 11, 127–150.
Curriero, F., 2006. On the use of non-Euclidean distance measures in geostatistics. Math. Geol. 38, 907–926.
Haggarty, R., Miller, C., Scott, E., 2014. Spatially weighted functional clustering of river network data. Series C, J. R. Stat. Soc. Series C, 64.491–506,
Hörmann, S., Kokoszka, P., 2011. Consistency of the mean and the principal components of spatially distributed functional data. Bernoulli 19, 1535–1558.
Horváth, L., Kokoszka, P., 2012. Inference for Functional Data with Applications. In: Springer Series in Statistics, Springer, New York, NY.
Marron, J.S., Alonso, A.M., 2014. Overview of object oriented data analysis. Biom. J. 56 (5), 732–753.
Mateu, J., Giraldo, R., 2021. Geostatistical Functional Data Analysis. Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd.
Menafoglio, A., Gaetani, G., Secchi, P., 2018. Random domain decompositions for object-oriented Kriging over complex domains. Stochast. Environ. Res. Risk Assess. 32, 3421–3437.
Menafoglio, A., Pigoli, D., Secchi, P., 2021. Kriging Riemannian data via random domain decompositions. J. Comput. Graph. Statist. 30 (3), 709–727.
Menafoglio, A., Secchi, P., 2017. Statistical analysis of complex and spatially dependent data: A review of Object Oriented Spatial Statistics. European J. Oper. Res. 258, 401–410.
Menafoglio, A., Secchi, P., 2019. O2S2: A new venue for computational geostatistics. Appl. Comput. Geosci. 2, 100007.
Menafoglio, A., Secchi, P., Dalla Rosa, M., 2013. A universal Kriging predictor for spatially dependent functional data of a Hilbert space. Electron. J. Stat. 7, 2209–2240.
Peterson, E., Theobald, D., Ver Hoef, J., 2007. Geostatistical modelling on stream networks: Developing valid covariance matrices based on hydrologic distance and stream flow. Freshw. Biol. 52, 267–279.
Peterson, E., Ver Hoef, J., 2010. A mixed-model moving-average approach to geostatistical modeling in stream networks. Ecology 91, 644–651.
R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis. In: Springer Series in Statistics, Springer, New York, NY.
Van Den Boogaart, K.G., Egozcue, J., Pawlowsky-Glahn, V., 2014. Bayes Hilbert Spaces. Australian New Zealand J. Stat. 56, 171–194.
Ver Hoef, J., Peterson, E., 2010. A moving average approach for spatial statistical models of stream networks. J. Amer. Statist. Assoc. 105, 6–18.
Ver Hoef, J., Peterson, E.E., Clifford, D., Shah, R., 2014. SSN: An R package for spatial statistical modeling on stream networks. J. Stat. Softw. 56 (3), 1–45.
Ver Hoef, J., Peterson, E., Theobald, D., 2006. Spatial statistical models that use flow and stream distance. Environ. Ecol. Stat. 13, 449–464.
Yaglom, A.M., 1987. Correlation Theory of Stationary and Related Random Functions. In: Springer Series in Statistics, vol. 1, Springer-Verlag, New York, NY.
Zhu, K., 2007. Operator Theory in Function Spaces, second ed. American Mathematical Society.
Zimmerman, D.L., Ver Hoef, J.M., 2017. The torgegram for fluvial variography. J. Comput. Graph. Statist. 13, 253–264.