Variograms for kriging and clustering of spatial functional data with phase variation

Xiaohan Guo [a],[*], Sebastian Kurtek [a], Karthik Bharath [b]

[a] *Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210, USA*
[b] *School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

ARTICLE INFO

ABSTRACT

Spatial, amplitude and phase variations in spatial functional data are confounded. Conclusions from the popular functional trace-variogram, which quantifies spatial variation, can be misleading when analyzing misaligned functional data with phase variation. To remedy this, we describe a framework that extends amplitude-phase separation methods in functional data to the spatial setting, with a view towards performing clustering and spatial prediction. We propose a decomposition of the trace-variogram into amplitude and phase components, and quantify how spatial correlations between functional observations manifest in their respective amplitude and phase. This enables us to generate separate amplitude and phase clustering methods for spatial functional data, and develop a novel spatial functional interpolant at unobserved locations based on combining separate amplitude and phase predictions. Through simulations and real data analyses, we demonstrate advantages of our approach when compared to standard ones that ignore phase variation, through more accurate predictions and more interpretable clustering results.

1. Introduction

1.1. Motivation

In many disciplines, including environmental science, medicine, biology, geology and econometrics, it is increasingly common to observe functional data with complex spatial dependencies;

---

* Corresponding author.

E-mail addresses: guo.1280@osu.edu (X. Guo), kurtek.1@stat.osu.edu (S. Kurtek), Karthik.Bharath@nottingham.ac.uk (K. Bharath).

such data are commonly referred to as spatial functional data (Delicado et al., 2010). An archetypal example is the well-known Canadian weather data consisting of daily temperature recordings at 35 locations across Canada, described in detail in Ramsay and Silverman (2005). Data representing spatial functional data come in the form of traditional spatio-temporal data (Cressie and Wikle, 2011). However, the functional data analysis framework allows one to directly capture temporal variation through its representation, thus enabling one to view data as discrete space–time realizations of a latent functional random field.

From this perspective, spatial functional data analysis can be regarded as an extension of spatial statistical methods to functional data objects. While standard multivariate spatial statistics can be used once some form of dimension reduction of functional data has been carried out (Nerini et al., 2010), the more popular approaches to model spatial correlations directly on observed functions have been based on the notion of a metric-based *trace-variogram*, which extends the standard variogram used in spatial statistics to the setting of second-order stationary, isotropic functional random fields assuming values in the Hilbert space of square-integrable functions (Giraldo et al., 2011). Accordingly, the standard L^2 metric is used in the definition of the trace-variogram which, when coupled with the spatial distance, captures spatial dependencies between functions (Goulard and Voltz, 1993). Specifically, if $\{f_s, \ s \in \mathcal{D}\} \subseteq L^2$ is a second-order stationary and isotropic random field, on a spatial domain $\mathcal{D}$ with metric d, the L^2 trace-variogram function

$$d(r, s) \mapsto V(d(r, s)) = \frac{1}{2} E(\|f_r - f_s\|^2), \tag{1}$$

quantifies spatial correlation between functions ($\| \cdot \|$ is the usual L^2 norm). The trace-variogram plays a central role in clustering and kriging of spatially correlated functional data (Mateu and Romano, 2017). For example, coefficients of linear combinations of observed functions that define a linear kriging estimate at a new location are determined using an estimate of the trace-variogram (Giraldo et al., 2011).

A key assumption, implicit with the use of the L^2 distance in the trace-variogram in (1), is that the temporal correspondence between functional observations is fixed. Thus, application of currently available L^2 metric-based trace-variogram methods to spatial functional data either assumes that the functions are perfectly aligned or treats phase variation as negligible noise. In reality, however, as with traditional functional data, it is frequently the case that the observed functions are out of phase: there is temporal misalignment of prominent geometric features of the functions, e.g., local extrema. For example, in the well-studied Canadian temperature dataset, this issue can arise when comparing average daily temperatures for two nearby cities, where in addition to spatial dependency of seasonal high and low temperatures, temporal seasonal trends shared between them should also be considered. Further, underlying phase variation in spatial functional data may easily make it non-stationary.

The adverse effects of disregarding phase variation while computing amplitude-related statistical summaries of functional data (e.g., functional mean and functional principal component analysis) using the L^2 distance are well-documented (Marron et al., 2015; Srivastava et al., 2011). The situation is exacerbated in the spatial setting since there are three sources of variation that are potentially confounded: amplitude, phase and spatial, and these have to be appropriately accounted for in the data analysis. A simulated example of kriging of spatial functional data with phase variation is shown in the left panel of Fig. 1. It is clear that the prediction generated by a method that accounts for phase variation (blue) is more accurate than one generated by a method that does not account for phase variation (red). To elaborate, for spatial functional data with phase variation, one is interested in quantifying spatial correlation between two complementary, latent features of the data: amplitude and phase. This necessitates a decomposition of the functional random field $\{f_s, \ s \in \mathcal{D}\}$ into its phase $\{\gamma_s, \ s \in \mathcal{D}\}$ and amplitude $\{f_s \circ \gamma_s, \ s \in \mathcal{D}\}$ components, which should then be used to define appropriate trace-variograms; the phase random field assumes values in the space of warping functions, made precise later, and $\circ$ denotes function composition. In other words, quantifying spatial variability using a trace-variogram V in the presence of phase variation requires a decomposition of V into separate amplitude and phase components that are trace-variograms themselves. Such a decomposition will enable more interpretable clustering relating to amplitude and phase components, and will result in better prediction of functions at unobserved locations. This constitutes the main focus of the paper, which to our knowledge has hitherto not been considered.
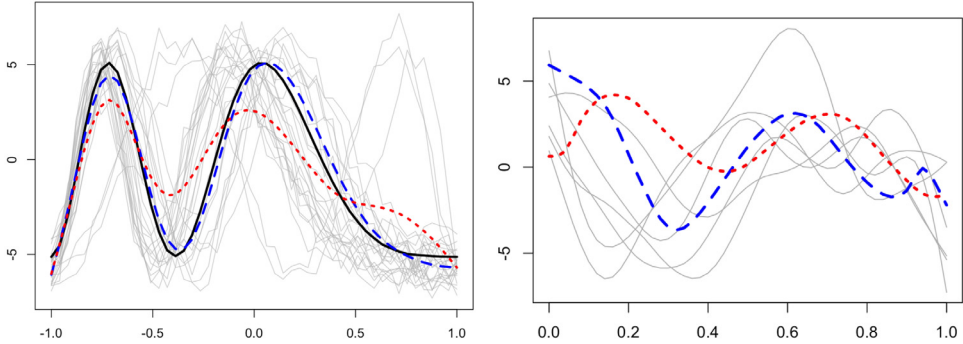
Fig. 1. Left: Spatial prediction of a target function (black) based on a sample of spatial functional data (gray) using kriging methods that account for (blue) and ignore (red) phase variation. Right: The local template estimated using spatial correlation (blue) better represents sample variability in a local area (gray) when compared to the (overall) mean amplitude function (red), and is a better template to use for local alignment of functions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

1.2. Contributions

The key challenge in decomposing the traditional trace-variogram V into separate amplitude and phase trace-variograms, say, V_a and V_p, lies in synthesizing spatial information with the fundamental asymmetry between the absolute and relative notions of amplitude and phase of a function: amplitude variation of a function $f : [a, b] \rightarrow \mathbb{R}$ can be viewed as variability in the set $\{f(t)\}$ of y-axis values as t varies in $[a, b]$, while phase variation tracks variability in locations along the x-axis of amplitude features of f *relative* to another function g. As a consequence, any definition of a trace-variogram for phase, based on the variance of the increment $f_{s_i} - f_{s_j}$ at locations $s_i, s_j \in \mathcal{D}$, will depend on the amplitudes (shapes) of the two functions and hence needs to be defined by conditioning on amplitudes (shapes).

Following viable definitions, estimation of V_a and V_p based on a sample of n observed functions at locations $s_1, \ldots, s_n$ requires a template function to estimate the unobserved warping functions, by aligning the sample functions to the template. In the absence of spatial correlation, a template is typically estimated by the mean amplitude function (Srivastava et al., 2011). However, when sample functions are spatially correlated, a locally defined (with respect to the spatial domain $\mathcal{D}$), data-driven template is desirable to better reflect the confounding between, and eventual disentangling of, amplitude, phase and spatial variations. The right panel in Fig. 1 illustrates the advantage of estimating the template using spatial information (blue) over the mean amplitude function (red); the spatial template better reflects amplitude features of the functions observed in a local area (gray). Such a desideratum is particularly relevant for kriging at a new location s_0 at which no functional datum is observed. Operationally, one could first align spatial functional data using any off-the-shelf registration algorithm to separate the amplitude and phase components, followed by appropriate modeling of spatial dependency. But, such an approach does not use local spatial information in the alignment procedure leading to poor results since spatial dependency amongst functions may arise in the amplitude component, the phase component or both; see Appendix A in the supplement for more on this issue. Accordingly, our contributions are as follows.

1. Aided by a geometric framework for amplitude-phase separation in spatial functional data we define separate amplitude and phase trace-variograms (Section 3.1); the amplitude trace-variogram is invariant to warping and hence captures pure amplitude variation (Lemma 1).
2. We propose an algorithm based on a non-trivial extension of the elastic functional data analysis framework (Srivastava and Klassen, 2016) to compute a spatially-weighted template (Algorithm 1) that enables simultaneous alignment and computation of estimators of the amplitude and phase trace-variograms (Section 3.2).

3. Using the trace-variograms, we propose: (i) linear unbiased estimators for kriging of amplitude and phase, which are combined to form the final kriging estimate, and discuss their properties (Sections 4.1–4.3); and (ii) a method for clustering spatial functional data into amplitude and phase clusters (Section 5).

1.3. Related work and article organization

Spatial functional data analysis has received considerable attention. Adaptation of multivariate spatial data methods to functional clustering, following dimension reduction, was done in Giraldo et al. (2012) and Haggarty et al. (2015). Romano et al. (2010) and Romano et al. (2017) extended the classical dynamic clustering approach in geostatistics to spatial functional data by employing the trace-variogram. On the other hand, Secchi et al. (2013) introduced Bagging Voronoi classifiers for clustering spatial functional data. This method was further improved by Abramowicz et al. (2017) by combining it with k-means registration (Sangalli et al., 2010).

Kriging is based on borrowing information from nearby objects to construct predictions at new spatial locations; the contribution to the predictor from each function depends on the strength of spatial correlation. Giraldo et al. (2011) used the trace-variogram for ordinary kriging of functional observations, which inspired related approaches. Chief amongst these are universal kriging methods wherein observed functions are pre-processed to better manage deviations from the stationarity assumption (Caballero et al., 2013; Menafoglio et al., 2013; Reyes et al., 2015; Menafoglio and Petris, 2016). However, non-stationarity induced by phase variation has not been considered in previous work, and this form of non-stationarity cannot be remedied using the state-of-the-art universal kriging approach (Menafoglio et al., 2013); see simulation results in Section 6.1. Menafoglio et al. (2021) further generalized kriging of functional data to data on a Riemannian manifold.

Indeed, not all functional kriging methods rely on the trace-variogram. Martínez-Hernández and Genton (2020) outlined a comprehensive list of functional kriging methods. Many of the approaches that do not use the trace-variogram focus on prediction via various forms of penalized regression. Aguilera-Morillo et al. (2017) proposed a functional spatial regression model with penalties accounting for spatial and temporal dependency. Bernardi et al. (2017) proposed a regression approach with partial regularization, and used two roughness penalties that separately accounted for spatial and temporal regularity. Compared to trace-variogram-based approaches, the proposed regression models do not explicitly model spatial dependency of the observations, and ensure regularity of the predictions through penalization.

The rest of this paper is organized as follows. Section 2 introduces the notions of amplitude and phase used throughout this paper, defines amplitude and phase distances used in the specification of amplitude and phase trace-variograms, and discusses template-based alignment to separate amplitude and phase variations. Section 3.1 introduces the proposed amplitude and (conditional) phase trace-variograms while Section 3.2 defines their estimators. Section 4 outlines the procedure for amplitude-phase kriging; a key step is the estimation of a spatially-weighted amplitude template (Algorithm 1). Section 5 introduces amplitude-phase hierarchical clustering based on spatially-weighted dissimilarity matrices. Section 6 reports results of extensive simulations while Section 7 considers applications of the proposed methods on two different datasets. Finally, Section 8 offers a brief discussion and outlines directions for future work. The supplement contains empirical performance assessments of Algorithm 1 (Appendix A), proofs of all propositions (Appendix B), a conceptual model-based formulation for kriging and a discussion of convergence for the proposed amplitude kriging estimator (Appendix C), a discussion of invariance of amplitude-phase clustering to the global scale of the amplitude and (conditional) phase trace-variograms (Appendix D), and additional implementation details and kriging/clustering results (Appendix E).

2. Amplitude-phase separation

2.1. Relevant function spaces and distances

We build on the metric-based elastic functional data analysis framework for amplitude-phase separation (Srivastava et al., 2011; Srivastava and Klassen, 2016). Without loss of generality, we

consider the representation space of functional data objects to be $\mathcal{F} = \{f : [0, 1] \to R \mid f$ is absolutely continuous$\}$. The group of warping functions representing phase is $\Gamma = \{\gamma : [0, 1] \to [0, 1] \mid \gamma(0) = 0, \ \gamma(1) = 1, \ \dot{\gamma} > 0\}$ ($\dot{\gamma}$ is the time derivative of γ). For any $f \in \mathcal{F}$, $\gamma \in \Gamma$, the warping of f by γ is given by the group action of composition, $f \circ \gamma$. The group-theoretic formulation of phase enables a definition of the amplitude of a function f as the equivalence class $[f] = \{f \circ \gamma \mid \gamma \in \Gamma\} \subseteq \mathcal{F}$, known as its orbit under the action of Γ; thus, $f \circ \gamma \in [f]$ has the same amplitude as f for each $\gamma \in \Gamma$. The amplitude space then is the quotient space $\mathcal{F}/\Gamma = \{[f] \mid f \in \mathcal{F}\}$.

Separating amplitude and phase requires a metric on the amplitude space $\mathcal{F}/\Gamma$. A convenient way to define one is through a metric d on $\mathcal{F}$ that is invariant to simultaneous warping: for every $\gamma \in \Gamma$, $d(f_1, f_2) = d(f_1 \circ \gamma, f_2 \circ \gamma)$. Under such a metric d, it becomes possible to view the action of the group Γ as performing an isometric operation $\gamma \mapsto f \circ \gamma$, much like an orthogonal transformation $O \mapsto Ox$ for orthogonal matrices O and $x \in \mathbb{R}^d$ that preserves sums of squares of relevant quantities in the multivariate setting.

The standard L^2 metric fails to be invariant and Srivastava et al. (2011) thus proposed to use the extended Fisher–Rao (eFR) metric. Unfortunately, this metric is difficult to use in practice. However, the *square-root slope transform* remarkably reduces the complicated eFR metric on $\mathcal{F}$ to the standard L^2 metric on the transformed space. The transform maps $f \mapsto Q(f) = q = \text{sgn}(\dot{f})|\dot{f}|^{1/2}$ ($\dot{f}$ is the time derivative of f). Given $f(0)$, Q is bijective with inverse $Q^{-1}(q, f(0))(t) = f(t) = f(0) + \int_0^t q(u)|q(u)|du$. Henceforth, for any $f \in \mathcal{F}$, we will refer to $q = Q(f)$ as its square-root slope function (SRSF).

The transformed space $Q(\mathcal{F})$ is a subset of $L^2[0, 1]$, and, by an abuse of notation, is denoted by $\mathcal{Q}$. Under Q, the eFR metric on $\mathcal{F}$ maps to the standard L^2 metric on $\mathcal{Q}$, and thus analysis of SRSFs can be carried out using standard Hilbert space machinery. Warping of $f \in \mathcal{F}$ by γ induces the warping action $(q, \gamma) = (q \circ \gamma)\dot{\gamma}^{1/2}$ on $\mathcal{Q}$ equipped with the L^2 metric, and the action is by isometries since $\|(q, \gamma)\| = \|q\|$ for every $\gamma \in \Gamma$, $q \in \mathcal{Q}$.

The corresponding orbit or *amplitude* of the SRSF q is then given by $[q] = \{(q, \gamma) \mid \gamma \in \Gamma\}$, and the *amplitude space* becomes $\mathcal{Q}/\Gamma = \{[q] \mid q \in \mathcal{Q}\}$. According to this definition, the amplitude of q is an entire equivalence class under the action of Γ; this implies that each member (q, γ) of $[q]$, as γ varies in Γ, represents the amplitude component of the function q. We will use 'amplitude' to refer to both $[q]$ and (q, γ), for any particular γ, and the context will disambiguate the two. Note that the amplitude of a function contains its magnitude (global scale), whereas a sensible notion of 'shape' of a function would be one that is scale-invariant. We thus define the *shape* of a function as the SRSF orbit $[\bar{q}] = \{(\bar{q}, \gamma) \mid \gamma \in \Gamma\}$, where $\bar{q} = q/\|q\|$ corresponds to the scale-normalized function, and the set of shapes of functions in $\mathcal{Q}$ constitute the shape space.

Definition 1 (*Amplitude and Shape Distance*). The amplitude distance between $q_1, q_2 \in \mathcal{Q}$ is defined as $d_a(q_1, q_2) = \inf_{\gamma \in \Gamma} \|q_1 - (q_2, \gamma)\|$, where $\|\cdot\|$ is the L^2 norm, and is a distance on the amplitude space. The shape distance between $q_1, q_2 \in \mathcal{Q}$ is defined as $d_{sh}(q_1, q_2) = d_a(\bar{q}_1, \bar{q}_2) = d_a(q_1/\|q_1\|, q_2/\|q_2\|)$ and is a distance on the shape space.

Amplitude and phase separation through pairwise registration or alignment of f_2 to f_1 (or vice versa) is formulated as the determination of the relative phase obtained by solving

$$\gamma^* = \arg\min_{\gamma \in \Gamma} \|q_1 - (q_2, \gamma)\| = \arg\min_{\gamma \in \Gamma} \|\bar{q}_1 - (\bar{q}_2, \gamma)\|, \tag{2}$$

typically using the dynamic programming algorithm, where q_1 and q_2 are the SRSFs of f_1 and f_2, respectively. The optimal alignment of f_2 with respect to f_1 is then given by $f_2 \circ \gamma^*$.

Alignment of f_2 to f_1 using q_1 and q_2 allows us to compute their relative phase distance. For this, we consider the square-root slope transform ψ of γ: $\gamma \mapsto Q(\gamma) = \psi = \dot{\gamma}^{1/2}$. Since $\int_0^1 \psi^2(t)dt = 1$ and $\psi(t) > 0 \ \forall t$, the square-root transformed warping group $Q(\Gamma) = \Psi$ is the positive orthant of the unit sphere in $L^2[0, 1]$, enabling us to define the (extrinsic) relative phase distance.

Definition 2 (*Phase Distance*). If $\psi^* = \sqrt{\dot{\gamma}^*}$ is the relative phase between $q_1, q_2 \in \mathcal{Q}$, then their (extrinsic) phase distance is $d_p(q_1, q_2) = \|\psi^* - \psi_{id}\|$, where $\psi_{id}(t) = 1$ is the square-root slope transformed identity warping function $\gamma_{id}(t) = t$.

Since Ψ is a subset of the unit sphere in $L^2[0, 1]$, the intrinsic 'arc-length' distance $\cos^{-1}(\langle \psi^*, \psi_{id} \rangle)$ can also be used. We note that Ψ, equipped with the L^2 Riemannian metric, is a Riemannian manifold. Further, the L^2 metric on Ψ corresponds to the Fisher–Rao metric on the warping group Γ (Srivastava et al., 2011).

Due to the nonlinear nature of warping, the L^2 distance between $q_1, q_2 \in \mathcal{Q}$ does not decompose exactly into the respective amplitude and phase distances in Definitions 1 and 2. The elastic framework, however, enables us to extract pure amplitude and phase components, and disentangle them from spatial variation in spatial functional data.

2.2. Template-based alignment of multiple functions

Amplitude-phase decomposition of variability present in a sample $f_1, \ldots, f_n$ can be carried out using the corresponding SRSFs $q_1, \ldots, q_n \in \mathcal{Q}$ (equipped with the L^2 metric) by jointly aligning the sample to a template $\mu_q \in \mathcal{Q}$, which is representative of a population-level amplitude. A natural choice is a representative $\hat{q}$ from the amplitude Karcher mean of $[q_1], \ldots, [q_n]$, which is defined as a local minimizer of the variance functional $[q] \mapsto \sum_{i=1}^{n} d_a^2(q, q_i)$ on the amplitude space $\mathcal{Q}/\Gamma$. In practice, this is carried out by using an algorithm that iterates between aligning $\{q_i\}$ to the current iterate of the representative of the Karcher mean amplitude and updating it (Section 8.3.3 of Srivastava and Klassen, 2016). The output of such an algorithm is the representative $\hat{q}$ and optimal warping functions $\{\hat{\gamma}_i\}$, such that $(q_i, \hat{\gamma}_i)$ are optimally aligned to $\hat{q}$ with respect to the metric d_a. When $\{q_i\}$ are spatially correlated across the spatial domain $\mathcal{D}$, their amplitudes (and hence the relative phases) are dependent on the locations in $\mathcal{D}$, and using a common template in their alignment might be inappropriate. We propose a modified version of the above algorithm (Algorithm 1) that jointly computes a suitable template for alignment of q_i *and* carries out the alignment.

2.3. Setup and notation

We focus on the setting of dense functional data (Wang et al., 2016), wherein a function at each spatial location is assumed to have been observed on a fine partition of $[0, 1]$. This implies that we are not considering situations wherein some form of function estimation is required that can potentially add another source of variability to amplitude, phase and spatial variations. It is important to first understand the interplay between the variations in this setting before moving to the more challenging one of sparsely observed functional data.

The functional random field $\{f_s, \ s \in \mathcal{D}\}$, on a spatial domain $\mathcal{D} \subseteq R^2$, assumes values in $\mathcal{F}$. Associated with $\{f_s\}$ is its square-root slope transformed version $\{q_s, \ s \in \mathcal{D}\}$ such that $s \mapsto q_s \in \mathcal{Q}$. Then, $\{q_s\}$ is a square-integrable functional random field since $\mathcal{Q} \subseteq L^2([0, 1])$.

Observed functional data $f_{s_i}, \ s_i \in \mathcal{D} \ (i = 1, \ldots, n)$ is first mapped to its corresponding SRSF representation, q_{s_i}, and methodology is entirely developed using q_{s_i}. Henceforth, the subscript i as an index is short for the spatial location s_i (e.g., $q_i, \ \gamma_i$); the subscript s is only used with a functional random field (e.g., q_s). The L^2 norm (inner product) on the function spaces $\mathcal{Q}$ and Ψ is denoted by $\| \cdot \| (\langle \cdot, \cdot \rangle)$, while $\| \cdot \|_2$ denotes the Euclidean norm on $\mathcal{D}$.

3. Amplitude-phase separation of trace-variogram

Denote by $\mu_{q,s}$ the expected value of the random field $\{q_s\} \subseteq \mathcal{Q}$ defined using the Bochner integral. The covariance function $C : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ of $\{q_s\}$ is the positive definite function $C(s, s') = E(\langle q_s - \mu_{q,s}, q_{s'} - \mu_{q,s'} \rangle)$, resulting in the variance function being defined as $\mathrm{var}(q_s) = C(s, s) = E(\|q_s - \mu_{q,s}\|^2)$. The semi-variogram of the process $\{q_s\}$ then is a conditionally negative definite function defined as $\theta_q(s, s') = \frac{1}{2}\mathrm{var}(q_s - q_{s'})$ for $s, s' \in \mathcal{D}$. The random field $\{q_s\}$ is said to be second-order stationary and isotropic if $\mu_{q,s} \equiv \mu_q$, i.e., the mean is constant across the spatial domain $\mathcal{D}$, and $C(s, s')$ is a function of $\|s - s'\|_2$ only for every pair (s, s'). Under this condition, using Fubini's

theorem, the trace-semivariogram V_q corresponds to the integrated pointwise variogram (Giraldo et al., 2011),

$$V_q(h) = \theta_q(h) = \frac{1}{2} \int_0^1 \mathrm{E}\left[q_s(t) - q_{s'}(t)\right]^2 \mathrm{d}t = \frac{1}{2}\mathrm{E}(\|q_s - q_{s'}\|^2), \tag{3}$$

where $h = \|s - s'\|_2$. In other words, V_q is (half the) expected squared L^2 distance between values of the functional random field $\{q_s\}$ at two locations in $\mathcal{D}$. Henceforth, we will simply refer to V_q as the *trace-variogram*. The definition implicitly assumes that q_s and $q_{s'}$ are aligned with zero phase variation, a situation rarely true in practice. Importantly, V_q is invariant to warping of two SRSF functions q_s and $q_{s'}$ by the same $\gamma \in \Gamma$. This is not true when the trace-variogram is defined using the L^2 distance on the random field $\{f_s\}$ as in (1), providing a strong motivation for using the SRSF representation to define separate amplitude and phase trace-variograms.

3.1. Trace-variograms for amplitude and phase

The amplitude and phase components in spatial functional data represent two distinct sources of variation, and importantly, can have different spatial correlations. Furthermore, in contrast to current approaches, phase variation cannot be viewed as noise. For example, in the aforementioned Canadian weather data (Ramsay and Silverman, 2005), phase represents important seasonal trends of temperature fluctuations across the observed sites, and spatial correlation in the phase component is important to explore regional climate change. Thus, our aim is to define complementary amplitude and phase trace-variograms that separately capture spatial correlation in these two components of spatial functional data, and can be used in downstream statistical tasks.

To define amplitude and phase trace-variograms, we treat the functional random field $\{q_s, \ s \in \mathcal{D}\} \subseteq \mathcal{Q}$ as being comprised of two random fields representing amplitude and phase. The amplitude random field is defined as $\{(q_s, \gamma_s), \ s \in \mathcal{D}\}$ and the phase random field as $\{\psi_s, \ s \in \mathcal{D}\}$, where $\gamma_s \in \Gamma$ is a random warping function and $\psi_s = \dot{\gamma}_s^{1/2}$. Amplitude-phase separation of $\{q_s\}$ into $\{(q_s, \gamma_s)\}$ and $\{\psi_s\}$ allows us to capture the spatial dependence in functional data via two different trace-variograms, one for the amplitude and one for the phase. We provide the definitions of the amplitude and phase trace-variograms next.

Definition 3 (*Amplitude Trace-Variogram*). Assuming that the amplitude random field $\{(q_s, \gamma_s), \ s \in \mathcal{D}\}$ is second-order stationary and isotropic, the amplitude trace-variogram is defined as

$$\|s - s'\|_2 = h \mapsto V_a(h) = \frac{1}{2}\mathrm{E}\left(\|(q_s, \gamma_s) - (q_{s'}, \gamma_{s'})\|^2\right). \tag{4}$$

The amplitude trace-variogram is similar to the trace-variogram in (3). The random warping functions γ_s and $\gamma_{s'}$ account for the removal of phase variation from the original random field $\{q_s\}$. Since the amplitude random field is assumed to be stationary, the above definition in essence supposes that any non-stationarity in the functional random field $\{q_s\}$ is induced by phase variation. This is manifestly different from the typical case wherein a spatially dependent mean induces non-stationarity, for which the universal kriging predictor of Menafoglio et al. (2013) may be used. Further, the proposed amplitude trace-variogram is invariant to simultaneous warping of $\{q_s\}$, which is a direct consequence of the isometric action of Γ on $\mathcal{Q}$ under the L^2 metric, as recorded in the following lemma.

Lemma 1. *The amplitude trace-variogram in* (4) *is invariant to simultaneous warping of the functional random field* $\{q_s, \ s \in \mathcal{D}\}$ *by any* $\gamma \in \Gamma$.

While it may be reasonable to assume that the amplitude random field $\{(q_s, \gamma_s)\}$ is second-order stationary and isotropic on $\mathcal{D}$, elements of the phase random field $\{\psi_s\}$ are only relative and generally depend on both (proximity of) spatial locations and similarity in the *shapes of the functions* that constitute the random field $\{q_s\}$. Thus, to account for the relative nature of phase, it is thus sensible to consider the phase random field $\{\psi_s\}$ conditional on the shape random field

associated with $\{q_s\}$ defined as $\mathcal{S} = \{[\bar{q}_s], \ s \in \mathcal{D}\}$ (recall that $\bar{q}_s = q_s/\|q_s\|$). This allows us to handle the non-stationarity in the phase random field $\{\psi_s\}$ due to heterogeneous shapes of the functions in the random field $\{q_s\}$. The relative nature of phase, with respect to amplitude or shape features, has received considerable attention in previous literature, albeit in other statistical contexts. For example, Sangalli et al. (2010) propose a simultaneous approach for clustering and alignment of functional data, where the cluster partitions are determined via amplitude similarity, and the relative phase of each function is estimated with respect to a cluster-specific template. As a result, the procedure accounts for the fact that only functions with similar amplitude have comparable phase components. Strait et al. (2017) and Matuk et al. (2021) further show that using shape constraints to regularize the phase component of functions and/or curves can result in more natural alignment. This leads to the following definition of the conditional phase trace-variogram.

Definition 4 (*Conditional Phase Trace-Variogram*). Assuming that, conditional on the shape random field $\mathcal{S}$, the phase random field $\{\psi_s, \ s \in \mathcal{D}\}$ is second-order stationary and isotropic, the conditional phase trace-variogram is defined as

$$V_p(\|s - s'\|_2, \mathcal{S}) = \frac{1}{2}\mathrm{E}\left(\|\psi_s - \psi_{s'}\|^2 \mid \mathcal{S}\right). \tag{5}$$

The above definition requires a valid definition of distance on $\mathcal{D}$ that uses information of the shape random field $\mathcal{S}$. Inspired by the approach proposed by Schmidt et al. (2011) for traditional spatial data, we consider shape as an additional covariate in order to define a pseudo-metric on a subset $\mathcal{M} := \{(s, [\bar{q}_s]) \mid s \in \mathcal{D}, \ [\bar{q}_s] \in \mathcal{S}\} \subset \mathcal{D} \times \mathcal{S}$. For a fixed $\omega \in R_{\geq 0}$, define a functional $h_\omega : \mathcal{M} \times \mathcal{M} \to R_{\geq 0}$ as

$$h_\omega = h_\omega((s, [\bar{q}_s]), (s', [\bar{q}_{s'}])) = \sqrt{\|s - s'\|_2^2 + \omega \cdot d_{sh}^2(q_s, q_{s'})}, \ \omega \geq 0, \quad (s, [\bar{q}_s]), \ (s', [\bar{q}_{s'}]) \in \mathcal{M}, \tag{6}$$

that provides a combined measure of discrepancy between shapes $[\bar{q}_s]$ and $[\bar{q}_{s'}]$ of two functions at locations s and s' and their spatial distance; ω serves as a tuning parameter that allows us to adjust the importance of the shape covariate. Thus, we consider a modification of the conditional phase trace-variogram V_p defined as

$$V_p(h_\omega) = V_p(\|s - s'\|_2, \mathcal{S}) = \frac{1}{2}\mathrm{E}\left(\|\psi_s - \psi_{s'}\|^2 \mid \mathcal{S}\right).$$

Reminiscent of the pseudo-metric $\mathrm{E}[(x_s - x_{s'})^2]^{1/2}$ on $\mathcal{D}$ for a Gaussian random field $\{x_s, \ s \in \mathcal{D}\}$ (see, e.g., Section 1.3 of Adler and Taylor, 2007), one can view h_ω as a pseudo-distance on the spatial domain $\mathcal{D}$, and its definition is motivated by the fact that the relative phase components of functions with very different shapes are not comparable. In other words, for a fixed $\omega > 0$, when two functions have very different shapes, their phase components are viewed as 'spatially' far away from each other in terms of the pseudo-distance h_ω. Viewing function shape information as an additional covariate (or coordinate), the relative phase components of $\{q_s, \ s \in \mathcal{D}\}$ are further stratified according to the shapes of the associated functions. This idea is analogous to the one used in Sangalli et al. (2010) for simultaneous clustering and alignment of functional data; the main difference lies in the use of a continuous measure of shape discrepancy in our case versus a discrete partition of the function space in theirs. As will be seen in the sequel, the estimator of the proposed conditional phase trace-variogram better captures the interplay between relative phases and spatial dependencies of the sample functions. Henceforth, we refer to the conditional phase trace-variogram simply as the phase trace-variogram.

Remark 1. Introducing shape information in the phase trace-variogram allows us to account for the potential association between amplitude and phase components in spatial functional data. As defined, the phase trace-variogram considers a functional random field over an infinite-dimensional domain, i.e., the space $\mathcal{M} \subset \mathcal{D} \times \mathcal{S}$. Literature on variography over infinite-dimensional spaces is scarce, and we use the proposal in this paper without formal theoretical justification. That said, we
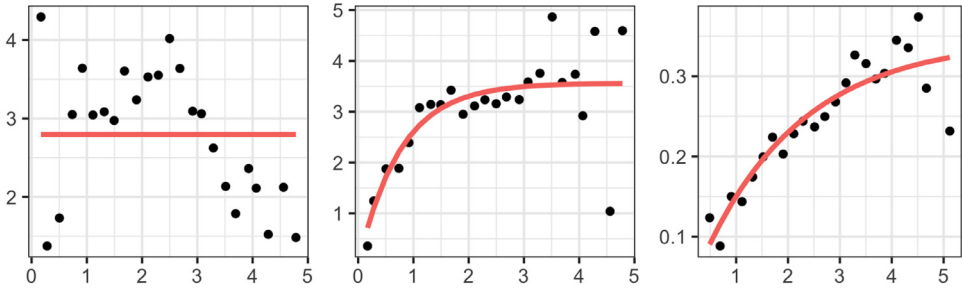
Fig. 2. Decomposition of the L^2 trace-variogram (left) into amplitude (middle) and phase (right) components for simulated functional data with spatially correlated amplitudes and phases. The dots represent the empirical L^2, amplitude, and phase trace-variograms (see definitions in Section 3.2). Estimates of the trace-variograms (red curves) are obtained by fitting a Matérn variogram model to the empirical trace-variograms.

have found through extensive simulations and real data applications that the phase trace-variogram defined in (5) has strong practical value; see Sections 6 and 7, and Appendix E in the supplement. A rigorous examination of the conditional phase trace-variogram is a significant undertaking and beyond the scope of this manuscript; as such, we leave it as future work. Alternatively, one could define the phase trace-variogram by replacing d_{sh} in (6) with a distance on a space of reduced (finite) dimension that captures shape features of the functions. Dimension reduction in this case can be attained either through functional principal component analysis or an appropriate basis decomposition. However, the choice of dimension reduction procedure will have a strong effect on the resulting distance and phase trace-variogram.

The benefits of constructing separate amplitude and phase trace-variograms are illustrated in Fig. 2 using simulated functions, wherein the spatial dependency in the data arises through both the amplitude and phase components. The amplitude components are generated from a second-order stationary and isotropic functional random field, whereas the correlation between phase components arises through both, their spatial locations and the shape features of the associated functions. Failure to disentangle the amplitude and phase variations leads to an empirical trace-variogram (Delicado et al., 2010) that suggests a quadratic pattern for spatial dependency (left panel), which is the truth for neither amplitude nor phase. A fitted Matérn variogram model, shown in red, is constant and fails to capture the spatial correlation that exists in the data. On the other hand, decomposing the trace-variogram into amplitude and phase (the empirical versions of (4) and (5)) appropriately captures the spatial correlatedness (middle and right panels) in these two components.

3.2. Estimating amplitude and phase trace-variograms

We have introduced the definitions of trace-variograms using latent amplitude and phase components. The amplitude and phase of given spatial functional data, however, are not observable and need to be estimated through appropriate alignment procedures that satisfy the requirements of different statistical analysis tasks. Here, given a sample of functions $\{q_i, \; s_i \in \mathcal{D}\}$ $(i = 1, \ldots, n)$, we propose empirical versions of the amplitude and phase trace-variograms that are compatible with the kriging and clustering tasks.

For kriging at a new location, since information from the entire sample $q_1, \ldots, q_n$ is used, we require a template-based multiple alignment approach. For this, it is essential to define a sensible template that captures spatially localized features of the sample. A detailed algorithm for estimating such a template is given in Section 4. Assuming that a template is available, we extract the relative phase components $\{\hat{\gamma}_i\}$ by aligning each function in the sample $q_1, \ldots, q_n$ to the template. The aligned functions $\{(q_i, \hat{\gamma}_i)\}$ and estimated (transformed) warping functions $\{\hat{\psi}_i\}$

are then used to estimate the amplitude and phase trace-variograms, respectively. Specifically, the empirical amplitude trace-variogram is

$$\hat{V}_a(h) = \frac{1}{2|N_a(h)|} \sum_{i,j \in N_a(h)} \|(q_i, \hat{\gamma}_i) - (q_j, \hat{\gamma}_j)\|^2, \qquad (7)$$

where $N_a(h) = \{(s_i, s_j) \mid \|s_i - s_j\| = h\}$. For irregularly spaced data, $N_a(h)$ is modified to $\{(s_i, s_j) : \|s_i - s_j\| \in (h - \epsilon, h + \epsilon)\}$ for a small $\epsilon > 0$. Similarly, a feasible estimator of V_p is

$$\hat{V}_p(h_\omega) = \frac{1}{2|N_p(h_\omega)|} \sum_{i,j \in N_p(h_\omega)} \|\hat{\psi}_i - \hat{\psi}_j\|^2, \qquad (8)$$

where the neighborhood $N_p(h_\omega) = \{((s_i, [\bar{q}_i]), (s_j, [\bar{q}_j])) \mid h_\omega((s_i, [\bar{q}_i]), (s_j, [\bar{q}_j])) \in (h_\omega - \epsilon, h_\omega + \epsilon)\}$ (for a small $\epsilon > 0$) is defined with respect to the pseudo-distance h_ω specified in (6), with a suitable choice of $\omega \geq 0$.

The estimators $\hat{V}_a$ and $\hat{V}_p$ are simplified when only pairwise comparisons of functions $q_1, \ldots, q_n$ are of interest; this is the case for example in clustering methods based on dissimilarity/distance matrices such as hierarchical clustering. Here, a joint alignment of $q_1, \ldots, q_n$ to a template can be avoided, with alignment between q_i and q_j carried out using either of the functions as the template. This circumvents the challenges associated with estimation of a template, and thus reduces the computational and methodological complexity. In this case, the corresponding expressions for $\hat{V}_a$ and $\hat{V}_p$ reduce to

$$\hat{V}_a(h) = \frac{1}{2|N_a(h)|} \sum_{i,j \in N_a(h)} d_a(q_i, q_j)^2, \quad \hat{V}_p(h_\omega) = \frac{1}{2|N_p(h_\omega)|} \sum_{i,j \in N_p(h_\omega)} d_p(q_i, q_j)^2, \qquad (9)$$

where $N_a(h)$ and $N_p(h_\omega)$ are defined as in (7) and (8).

To guarantee that the estimated variograms are conditionally negative definite (Cressie, 2015), we fit a Matérn variogram model to the empirical variograms at a discrete set of distance values; we use ordinary least squares (Cressie, 2015) to estimate the parameters of the Matérn model. In subsequent analyses, i.e., kriging and clustering, the fitted variograms are used instead of the empirical variograms $\hat{V}_a$ and $\hat{V}_p$. In the phase trace-variogram, the tuning parameter ω is selected to minimize the squared error of the parametric Matérn fit to the empirical estimate. Fig. 3 illustrates an example of tuning parameter selection for a simulated dataset. In the left panel, we plot the squared error (y-axis) of the fit versus different values of ω (on the $\log_{10}$ scale on the x-axis). In the middle ($\omega = 0$) and right (optimal $\omega = 10^{2.2}$) panels, spatial correlation patterns of the phase components are captured by scatterplots of the pairwise phase discrepancies $\|\hat{\psi}_i - \hat{\psi}_j\|^2$ (y-axis) versus the pairwise pseudo-distances h_ω (x-axis). In red, we highlight points corresponding to small spatial distances ($\|s_i - s_j\|$), but relatively large phase discrepancies. After introducing shape information as a covariate through the pseudo-distance h_ω, with $\omega = 10^{2.2}$, the red points shift to the right due to shape differences of the corresponding functions. This shows that the large phase discrepancy in the red points in the middle panel is partially due to the shape differences of the corresponding functions. In the right panel, using the optimal value of the tuning parameter, we are able to account for the shape heterogeneity in the given data, and as a result, detect a clearer dependency pattern between the phase components. Note that the scale of ω in the distance h_ω depends on the difference in the scales of the spatial and shape distances. In particular, one can show that the shape distance d_{sh} is bounded above by 2. The spatial distance, on the other hand, depends on the size (and coordinates) of the spatial domain. Thus, there is no absolute scale for the tuning parameter ω.

4. Amplitude-phase kriging

Giraldo et al. (2011) developed a linear unbiased estimator that extends ordinary kriging or spatial interpolation to the functional setting by minimizing the L^2 prediction error. In the presence of phase variation, the L^2-based linear estimator can be biased, since function features such as
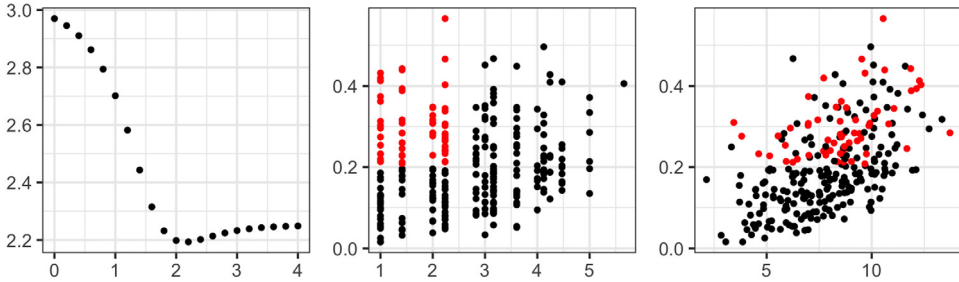
Fig. 3. Left: Squared error (y-axis) of Matérn variogram fit to the empirical phase trace-variogram (8) under different values of $\log_{10}(\omega)$ (x-axis). Middle: The pairwise squared distances $\|\hat{\psi}_i - \hat{\psi}_j\|^2$ (y-axis) versus pairwise pseudo-distances h_ω (x-axis) with $\omega = 0$. Right: Same as middle for $\omega = 10^{2.2}$. Red points represent pairs with small pseudo-distance h_ω, but relatively large phase discrepancy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

local extrema can be misaligned; see an example in the left panel of Fig. 1. Given observations $\{q_i, s_i \in \mathcal{D}\}$ $(i = 1, \ldots, n)$, the goal is to predict an unobserved function q_0 at a new location $s_0 \in \mathcal{D}$ comprising amplitude (q_0, γ_0) and phase γ_0. To address possible misalignment of q_i, we consider a three-stage kriging procedure: (i) predict the amplitude component, (ii) predict the phase component conditional on the predicted amplitude, and (iii) combine the two to obtain the kriging estimate.

4.1. Amplitude kriging

Let us first assume that a template has been chosen so that a multiple alignment procedure has been implemented to obtain aligned functions $\{(q_i, \hat{\gamma}_i)\}$ and warping functions $\{\hat{\gamma}_i\}$; for each $i = 1, \ldots, n$, recall that $(q_i, \hat{\gamma}_i)$ is an estimate of the amplitude $[q_i]$ of q_i as a representative of the orbit. Let $\Delta_n := \{(x_1, \ldots, x_n)^T \in R^n | \sum_{i=1}^{n} x_i = 1\}$. We define the linear estimator of the amplitude component (q_0, γ_0) at s_0 as

$$\tilde{q}_0(t) = \sum_{i=1}^{n} \eta_i (q_i, \hat{\gamma}_i)(t), \tag{10}$$

where the coefficient vector $\eta = (\eta_1, \ldots, \eta_n)^T \in \Delta_n$ is implicitly defined as the minimizer of the expected amplitude prediction error functional

$$\eta \mapsto E\left(\|\tilde{q}_0 - (q_0, \gamma_0)\|^2\right). \tag{11}$$

In the ideal setting, if $\{\gamma_i\}$ are known or can be estimated exactly, the situation reduces to the traditional setting without phase variation for determining η that simplifies the optimization in (11) through its relationship with a matrix consisting of evaluations of the trace-variogram (see, e.g., Menafoglio et al. (2013)). The following result demonstrates this.

Proposition 1. *If we assume that γ_i can be estimated exactly such that $(q_i, \hat{\gamma}_i) = (q_i, \gamma_i)$ for every $i = 1, \ldots, n$, then the $\eta \in \Delta_n$ that minimizes (11) also minimizes $\eta \mapsto \eta^T \mathcal{V}_a \eta$, where the $n \times n$ matrix $\mathcal{V}_a$ contains as its elements $V_a(h_{0j}) + V_a(h_{i0}) - V_a(h_{ij})$ with $h_{ij} = \|s_i - s_j\|_2$ $(i, j = 1, \ldots, n)$. As a consequence, the amplitude kriging predictor in (10) depends only on the amplitude trace-variogram $V_a(h)$.*

Unfortunately, it is well-known that the ideal setting considered in Proposition 1, wherein the phase components γ_i can be estimated exactly, only applies in very restrictive modeling scenarios (Kurtek and Srivastava, 2011; Chakraborty and Panaretos, 2021); a more detailed discussion of this issue is given in Appendix C in the supplement. An important implication of this fact is that the

estimation of amplitude and phase components heavily relies on the choice of template, and thus it is essential to estimate an appropriate template in the amplitude-phase separation for kriging.

When aligning independent functional data, as described in Section 2.2, the template is typically estimated by the mean amplitude function (a representative of the mean amplitude orbit), i.e., the minimizer of the variance functional. However, when the sample functions are spatially correlated, the dependence structure must be taken into account in template estimation. When focusing on prediction at a certain location s_0, we seek a local (with respect to the spatial domain $\mathcal{D}$) template that captures the amplitude features of aligned functions that have (strong) spatial correlation with the amplitude component at s_0. In particular, the ideal template for alignment of $q_1, \ldots, q_n$ for amplitude kriging at s_0 is an element of the orbit $[q_0]$. However, q_0 (and its orbit) is not observed as it is the quantity we seek to predict. This implies that the two objectives of (i) estimating a template for alignment of spatial functional data, and (ii) prediction of amplitude at the location s_0 are essentially the same. We outline a procedure, provided as Algorithm 1, that iterates between the following two steps until convergence: (i) alignment of $\{q_i, \ s_i \in \mathcal{D}\}$ given the current estimate of the template, and (ii) prediction of amplitude at s_0 that specifies an update of the template. The algorithm results in a *spatially-weighted amplitude kriging estimator $\tilde{q}_0$ that serves the dual purpose of acting as a local template for alignment and as a predictor of the amplitude component* (q_0, γ_0).

Algorithm 1: Amplitude kriging estimate

---

1 **Input**: s_0 and $(s_1, q_1), \ldots, (s_n, q_n)$;

2 **Output**: Amplitude kriging estimate $\tilde{q}_0$;

3 Initialization: $k = 0$, $\epsilon > 0$ and template $\tilde{q}_0^{(0)}$ with the q_i closest to $s_0 \in \mathcal{D}$;

4 **while** $\|\tilde{q}_0^{(k+1)} - \tilde{q}_0^{(k)}\| > \epsilon$ **do**

5   Align each q_i to $\tilde{q}_0^{(k)}$ using (2) to obtain $(q_i, \hat{\gamma}_i^{(k)})$;

6   Compute $\hat{V}_a(h)$ using $\{(q_i, \hat{\gamma}_i^{(k)})\}$ and $\tilde{q}_0^{(k)} = \sum_i \eta_i(q_i, \hat{\gamma}_i^{(k)})$ using Proposition 1;

7   Set $\tilde{q}_0^{(k+1)} = \tilde{q}_0^{(k)}$, $k = k + 1$;

8 **end**

In Algorithm 1, within each iteration k, the template $\tilde{q}_0^{(k)}$ is fixed, and acts as the given template in Proposition 1. Spatial information is incorporated through the use of $\hat{V}_a(h)$ for determining η. Strictly speaking, the equivalent formulation of the optimization criterion to determine η in Proposition 1 assumes that $\{\hat{\gamma}_i\}$ recover $\{\gamma_i\}$ exactly; we nevertheless use it as it significantly simplifies computations. As a result, Algorithm 1 specifies a (spatially) local alignment procedure that emphasizes functions observed at locations close to s_0 based on a local template.

Remark 2. Explicit convergence analysis for Algorithm 1 is complicated due to the alternating optimizations to compute the phase functions $\{\hat{\gamma}_i\}$, and the weights η to construct the spatially-weighted template, at each iteration. In particular, due to the dependence of η in line 6 on the phase functions estimated in line 5, it is difficult to formalize the entire procedure under a single cost function. We empirically examine convergence properties of Algorithm 1 in Appendix A in the supplement. Additionally, in Appendix C in the supplement, we establish convergence of the algorithm for a one-dimensional model when randomness manifests in q only through a scale parameter.

The advantages of using the local template over a global one (e.g., Karcher mean of amplitude) are confirmed by simulations in Appendix A in the supplement. In particular, it is evidenced there that the local template, estimated via Algorithm 1, better reflects amplitude features of the sample functions observed in a (spatially) local area than a global template that does not take spatial dependence into account. In Appendix A, we also evaluate the influence of the initialization $\tilde{q}_0^{(0)}$ in line 3 on prediction performance.

4.2. Phase kriging

In amplitude kriging, phase variability is removed by aligning all functions with respect to the estimated template, which results in improved prediction of the shape and magnitude of a function

(Section 6). However, $\tilde{q}_0$ is the prediction of an element in $[q_0]$ and not q_0. To obtain the final kriging estimate of q_0, we require an estimator of the phase γ_0. We construct one by carrying out phase kriging using the estimated warping functions $\{\hat{\gamma}_i\}$, with corresponding square-root slope transforms $\{\hat{\psi}_i\}$, computed by aligning $\{q_i\}$ to the amplitude kriging estimate $\tilde{q}_0$.

We want to predict ψ_0 on Ψ, which is a nonlinear Riemannian manifold, using the relative phases $\hat{\psi}_1, \ldots, \hat{\psi}_n$. We deal with the nonlinearity of Ψ by considering the positive extension of Ψ, $\Psi' = \{\tilde{\psi} = a\psi \mid a \in R_+, \ \psi \in \Psi\}$, i.e., we embed Ψ in $L^2[0, 1]$ via Ψ'. Compatible with the linearity of the amplitude kriging estimate $\tilde{q}_0$, we use an extrinsic approach to phase kriging: we compute the corresponding linear phase kriging estimate in Ψ' and then project it back to Ψ. The projection $\Pi : \Psi' \to \Psi$ is defined as $\Pi(\tilde{\psi}) = \arg\min_{\psi \in \Psi} \|\psi - \tilde{\psi}\| = \tilde{\psi}/\|\tilde{\psi}\|$. This projection simply normalizes the magnitude of a point $\tilde{\psi} \in \Psi' \subset L^2[0, 1]$ to result in the closest point on the positive orthant of the unit sphere, $\psi \in \Psi$. Thus, $\Pi(\tilde{\psi}_0) = \tilde{\psi}_0/\|\tilde{\psi}_0\|$ is the phase kriging estimator of ψ_0 based on a linear estimator $\tilde{\psi}_0 \in \Psi'$. Such an estimator represents an extrinsic choice that uses a natural embedding of Ψ into $L^2[0, 1]$ via Ψ'.

Let $\Delta_n^+ := \{(x_1, \ldots, x_n)^T \in R^n | x_i > 0, \ \sum_{i=1}^n x_i = 1\}$. With the estimated phase components $\hat{\psi}_i$, the phase kriging estimate of ψ_0 in Ψ' is defined as

$$\tilde{\psi}_0(t) = \sum_{i=1}^n \zeta_i \hat{\psi}_i(t), \tag{12}$$

where $\zeta = (\zeta_1, \ldots, \zeta_n)^T \in \Delta_n^+$ minimizes the conditional phase prediction error functional

$$\zeta \mapsto \mathrm{E}(\|\tilde{\psi}_0 - \psi_0\|^2 \mid \mathcal{S}). \tag{13}$$

Positivity of ζ_i is required to ensure that the resulting phase prediction $\tilde{\psi}_0$ is positive, i.e., the corresponding $\tilde{\gamma}_0$ is strictly increasing. As with the amplitude kriging estimate in Proposition 1, the following result describes how the vector ζ can be computed, again under the idealized setting where the $\{\gamma_i\}$ can be exactly recovered.

Proposition 2. *Assume that there exists a template q such that $\hat{\gamma}_i = \gamma_i$. Then, the vector $\zeta \in \Delta_n^+$ can be obtained by minimizing $\zeta \mapsto \zeta^T \mathcal{V}_p \zeta$, where the $n \times n$ matrix $\mathcal{V}_p$ contains as its elements $V_p(h_{0j,\omega}) + V_p(h_{i0,\omega}) - V_p(h_{ij,\omega})$ with $h_{ij,\omega} = \sqrt{\|s_i - s_j\|_2^2 + \omega \cdot d_{sh}(q_i, q_j)^2}$ $(i, j = 0, 1, \ldots, n)$. The phase predictor in (12) thus depends only on the conditional phase trace-variogram $V_p(h_\omega)$.*

The proofs of Propositions 1 and 2 are presented in Appendix B of the supplement.

Computation of distances $h_{0j,\omega}$, $j = 1, \ldots, n$ and $h_{i0,\omega}$, $i = 1, \ldots, n$ relies on the knowledge of the function shape at s_0; for this, we use the shape of the amplitude kriging estimate $\tilde{q}_0$ at s_0. This aspect of phase kriging reflects the relative nature of phase, as described in Section 3.1, with respect to $\tilde{q}_0$.

Remark 3. An alternative to the proposed extrinsic approach, which we do not consider here, is to construct an intrinsically defined kriging estimator defined directly on Ψ using the geometry of the positive orthant of the Hilbert sphere (see, e.g., Section 7.5.4 of Srivastava and Klassen (2016)). For example, a phase predictor can be defined as a weighted Karcher mean via the intrinsic distance on Ψ, $\tilde{\psi}_0 = \arg\min_{\psi \in \Psi} \sum_{i=1}^n \zeta_i \cos^{-1}(\langle \psi, \psi_i \rangle)^2$, where $\zeta = (\zeta_1, \ldots, \zeta_n)^T \in \Delta_n^+$ is the minimizer of $\mathrm{E}(\cos^{-1}(\langle \tilde{\psi}_0, \psi_0 \rangle)^2 \mid \mathcal{S})$, i.e., the intrinsic counterpart to the extrinsic conditional phase prediction error specified in (13). Unfortunately, Proposition 2 does not hold in this case. In particular, without linearity as in the extrinsic approach, the prediction error cannot be decomposed as a function of the conditional phase trace-variogram. This, in turn, prohibits direct estimation of $\tilde{\psi}_0$.

4.3. Final prediction via combination of amplitude and phase kriging estimates

The predicted amplitude and phase kriging estimates $\tilde{q}_0$ and $\tilde{\psi}_0$ include all information about the magnitude, shape and temporal characteristics of the final prediction, but not the translation,

which is lost due to the square-root slope transformation. To account for this, we use the starting points $f_i(0)$ ($i = 1, \ldots, n$) of the observed functions and apply ordinary kriging (Cressie and Wikle, 2011) to obtain a translation prediction $\hat{T}_0$ of the function f_0.

Recall the inverse of the square-root slope transformation $Q^{-1} : (R \times \mathcal{Q}) \to \mathcal{F}$ from Section 2. The final kriging estimate combines the three estimates of amplitude, phase and translation as follows. First, we combine the amplitude and phase predictions using $q_0^* = (\tilde{q}_0, \tilde{\gamma}_0^{-1})$, where $\tilde{\gamma}_0(t) = \int_0^t \Pi(\tilde{\psi}_0)^2(u)du$ is the phase prediction. The combined kriging estimate of f_0 at site s_0 then is $f_0^* = Q^{-1}(q_0^*, \hat{T}_0)$, where $\hat{T}_0$ is the predicted starting point.

5. Amplitude-phase clustering

Amplitude and phase distances arising from amplitude-phase separation enable separate distance-based amplitude and phase clustering of functional data. Spatially informed adaptations can now be defined through the use of dissimilarity measures by combining the amplitude (phase) distance and amplitude (phase) trace-variogram. Incorporating amplitude-phase separation into clustering can lead to more interpretable clusters. For example, in the famous Canadian weather data (Ramsay and Silverman, 2005) considered in Section 7.2, we note that daily average temperatures at sites with similar extreme temperatures (similar amplitude) need not experience similar seasonal trends. Thus, one would reasonably expect different clustering results corresponding to the two components.

In contrast to clustering independent data, detecting homogeneous partitions of spatially correlated objects must additionally account for spatial dependence by grouping them based on both their similarity as well as proximity on the spatial domain. The proposed amplitude-phase clustering approach can be viewed as an extension of the spatially informed adaptations of clustering and classification for multivariate data (Oliver and Webster, 1989; Bourgault et al., 1992). While several distance-based clustering approaches can be used, we consider hierarchical clustering based on spatially-weighted dissimilarity matrices (Giraldo et al., 2012), by combining the amplitude (phase) distance and amplitude (phase) trace-variogram, and generate spatially-informed amplitude and phase clusters separately. Given observations $\{q_i, \ s_i \in \mathcal{D}\}$ ($i = 1, \ldots, n$), the amplitude and phase dissimilarity matrices are defined as

$$d_{A,ij} = d_a(q_i, q_j) \times V_a(\|s_i - s_j\|), \quad d_{P,ij} = d_p(q_i, q_j) \times V_p(h_{ij,\omega}), \quad i, j = 1, \ldots, n, \tag{14}$$

respectively. Since the dissimilarity matrices measure the discrepancy in amplitude and phase for each pair of functions, it is not necessary to choose a common template for all of the functions for alignment. Instead, we simply choose one of the functions in each pair as a template to compute the amplitude and relative phase distances between them. This ensures that $\hat{V}_a$ and $\hat{V}_p$ expressed in a simplified form via pairwise distances in (9) can be used. In the implementation of hierarchical clustering, we use complete linkage to define the discrepancy between clusters. The number of clusters is chosen by minimizing the average silhouette (Rousseeuw, 1987), which quantifies the difference in similarity of an object to its own cluster versus other clusters.

Because the trace-variogram is an increasing function of the distance h (or h_ω), clustering based on dissimilarities in (14) tends to generate partitions with good spatial contiguity in the presence of strong spatial dependence. Further, it is evident that clustering results do not depend on the overall scale of the trace-variogram, but rather its structure (rate of increase), which reflects the spatial correlatedness among the amplitude (phase) components. In particular, when there is no spatial dependence, the trace-variogram is constant, and the dissimilarity measures in (14) simplify to the amplitude and phase distances (multiplied by a different constant in each case). Furthermore, it is evident that hierarchical clustering based on the amplitude and phase dissimilarity measures defined in (14) are invariant to a global scaling of the amplitude and phase trace-variograms V_a and V_p. Appendix D in the supplement contains a more detailed discussion of this property.

6. Simulations

In the simulation studies, we assess the performance of the proposed amplitude-phase kriging and clustering methods. The fitting of variograms is carried out using functions in the R packages geofd (Giraldo et al., 2020) and geoR (Ribeiro Jr et al., 2020). Joint kriging and alignment in Algorithm 1 is carried out by appropriately modifying the relevant functions in the R package fdasrvf (Tucker, 2021). Hierarchical clustering is performed using the hclust function in R. All core computing tasks in this paper were conducted using a high performance computing cluster.

6.1. Kriging performance

6.1.1. Simulated data

We fix the spatial locations to equally-spaced sites on a 5×5 grid with x, y coordinates taking the values $(-2, -1, 0, 1, 2)$. Spatial functional data f_i ($i = 1, \ldots, 25$) is generated using the model

$$f_i(\gamma_i(t)) = \sum_{j=1}^{K} a_{i,j} \phi_j(t) + e_i(t), \quad t \in [a, b],$$

on the original function space $\mathcal{F}$ (and not the SRSF transformed space $\mathcal{Q}$), where, for each j, the coefficient vector $[a_{1,j}, \ldots, a_{25,j}]$ follows a multivariate normal distribution with a specific mean vector θ and the Matérn covariance $C_{Mat}(\cdot, \cdot; \sigma_a^2, 0.5, \ell_1)$; here, σ_a^2 is the scale parameter, ℓ_1 is the range, and the smoothing parameter is fixed to 0.5. This imposes spatial correlation in the amplitude component of the simulated data. Holding i fixed, the coefficients for the basis ϕ_j, $j = 1, \ldots, K$ are assumed to be independent. We consider two simulation settings based on the choice of basis functions:

1. *Bimodal*: set $K = 1$ and $\phi_1(t) = -\cos(2\pi t)$ on $[a, b] = [-1, 1]$, with the mean vector θ of $[a_{1,1}, \ldots, a_{25,1}]$ identically set to 5;
2. *B-spline*: set $K = 10$ and $\{\phi_j\}_{j=1}^{K}$ to be cubic B-splines on $[a, b] = [0, 1]$ with the mean vector θ for each i equal to $(1, 2, 3, 4, 5, 5, 4, 3, 2, 1)^T$.

We now describe how spatial correlation is induced amongst the warping functions. The phase components γ_i, $i = 1, \ldots, 25$ are chosen to be the cumulative distribution functions of the $Beta(1, e^{b_i})$ density with $\{b_1, \ldots, b_{25}\}$ generated from the correlated uniform distribution on $[-B, B]$, by transforming a random sample from the multivariate normal distribution with covariance $C_{Mat}(\cdot, \cdot; 1, 0.5, \ell_2)$. The parameter B determines the magnitude of phase variation while ℓ_2 controls the range of spatial dependency. Additionally, when the B-spline model is used, we consider two scenarios depending on whether the correlation between phase components depends on the shape of observed spatial functional data:

1. *B-spline Scenario 1*, where spatial phase correlation does not depend on function shapes: we use (6) with $\omega = 0$ to induce correlations between warping functions and set $\ell_1 = \ell_2 = 8^{1/2}$ with e_i generated from a white noise process with variance 0.25;
2. *B-spline Scenario 2*, where spatial phase correlation depends on function shapes: we use (6) with $\omega = 10$ to induce correlations between parameters of the warping functions, and set $\ell_1 = 8^{1/2}$ and ℓ_2 as the median of pairwise distances computed via (6).

6.1.2. Comparison with other methods

We compare predictive performance of the proposed amplitude-phase kriging method (APK) to three competing approaches: (1) ordinary kriging without alignment (OK) (Giraldo et al., 2011), (2) universal kriging without alignment (UK) (Menafoglio et al., 2013), and (3) two-stage kriging (TSK). For (3), we align the observed functions using the joint template-based alignment procedure described in Section 2.2, followed by ordinary kriging (Giraldo et al., 2011) applied to the aligned functions in SRSF space. Additionally, a translation prediction is generated in the same way as in

Table 1

Average prediction errors (SD) using metrics $E1$–$E5$, across 50 different replicates, for amplitude-phase kriging (APK), two-stage kriging (TSK), ordinary kriging (OK) and universal kriging (UK). $E2$ is divided by 100 and $E4$ is multiplied by 10 to adjust the scale. B controls the magnitude of phase variation.

Bimodal						
B	Method	$E1$	$E2$	$E3$	$E4$	$E5$
0.5	APK	1.12 (0.33)	**0.46 (0.12)**	**0.55 (0.19)**	**0.15 (0.03)**	10.4 (5.00)
	TSK	**1.11 (0.34)**	**0.46 (0.12)**	0.56 (0.19)	0.18 (0.04)	14.48 (6.56)
	OK	2.34 (1.00)	0.86 (0.32)	0.98 (0.30)	0.17 (0.04)	**8.99 (4.36)**
	UK	2.50 (0.99)	0.91 (0.32)	1.07 (0.33)	0.18 (0.05)	9.20 (4.36)
1	APK	1.48 (0.63)	1.31 (1.19)	**0.81 (0.24)**	0.55 (0.17)	31.42 (14.64)
	TSK	**1.45 (0.59)**	**1.30 (1.15)**	0.82 (0.23)	**0.54 (0.15)**	32.89 (14.11)
	OK	7.94 (3.88)	4.18 (2.19)	4.24 (1.93)	0.88 (0.37)	19.51 (8.05)
	UK	6.86 (2.64)	3.71 (1.86)	3.72 (1.14)	0.83 (0.24)	**19.43 (7.62)**
B-spline Scenario 1 (independent)						
B	Method	$E1$	$E2$	$E3$	$E4$	$E5$
0.5	APK	1.49 (0.42)	**2.32 (0.99)**	**2.26 (0.44)**	1.00 (0.25)	2.61 (0.57)
	TSK	1.55 (0.45)	2.35 (0.99)	2.36 (0.47)	1.12 (0.32)	3.06 (0.79)
	OK	**1.20 (0.21)**	2.34 (0.97)	2.48 (0.52)	**0.99 (0.22)**	**1.84 (0.29)**
	UK	1.32 (0.22)	2.44 (0.97)	2.67 (0.57)	1.05 (0.25)	1.99 (0.30)
1	APK	1.63 (0.42)	**2.96 (1.82)**	**2.52 (0.53)**	**1.23 (0.28)**	3.65 (0.95)
	TSK	1.63 (0.42)	3.00 (1.77)	2.61 (0.55)	1.39 (0.34)	4.09 (1.09)
	OK	**1.50 (0.27)**	3.28 (1.74)	3.12 (0.68)	1.40 (0.22)	**2.71 (0.65)**
	UK	1.59 (0.29)	3.32 (1.58)	3.17 (0.67)	1.43 (0.26)	2.83 (0.63)
B-spline Scenario 2 (dependent)						
B	Method	$E1$	$E2$	$E3$	$E4$	$E5$
0.5	APK	1.53 (0.46)	**2.26 (0.95)**	**2.30 (0.42)**	**1.04 (0.25)**	2.85 (0.64)
	TSK	1.56 (0.46)	2.31 (0.95)	2.37 (0.44)	1.17 (0.34)	3.26 (0.75)
	OK	**1.27 (0.24)**	2.39 (0.99)	2.62 (0.60)	1.06 (0.27)	**2.12 (0.49)**
	UK	1.36 (0.24)	2.48 (1.02)	2.74 (0.61)	1.10 (0.27)	2.23 (0.48)
1	APK	**1.66 (0.47)**	**2.96 (1.64)**	**2.55 (0.49)**	**1.34 (0.27)**	4.27 (1.42)
	TSK	1.70 (0.52)	2.99 (1.55)	2.66 (0.52)	1.44 (0.34)	4.44 (1.42)
	OK	1.77 (0.59)	3.70 (1.82)	3.64 (1.10)	1.66 (0.36)	**3.34 (1.05)**
	UK	1.74 (0.39)	3.52 (1.68)	3.46 (0.78)	1.61 (0.32)	**3.34 (1.02)**

amplitude-phase kriging. Then, the SRSF and translation predictions are combined via Q^{-1} to yield a prediction in the original function space.

Performance metrics. To assess performance, we apply leave-one-out cross-validation. Let $f^{[-i]*}$ denote the prediction of f_i using all observations except the ith, and $q^{[-i]*}$ and q_i denote their SRSFs; $f_*^{[-i]}$ is $f^{[-i]*}$ after optimal alignment to f_i and $\dot{f}$ is the time derivative of f. To measure the accuracy of predictions, we compute the following five error metrics:

- Amplitude least squares: $E1 = n^{-1} \sum_{i=1}^{n} \|f_*^{[-i]} - f_i\|^2$;
- Amplitude Sobolev least squares: $E2 = n^{-1} \sum_{i=1}^{n} \|\dot{f}_*^{[-i]} - \dot{f}_i\|^2$;
- Amplitude mean squared error: $E3 = n^{-1} \sum_{i=1}^{n} d_a(q^{[-i]*}, q_i)^2$;
- Phase mean squared error: $E4 = n^{-1} \sum_{i=1}^{n} d_p(q^{[-i]*}, q_i)^2$;
- L^2 prediction error: $E5 = n^{-1} \sum_{i=1}^{n} \|f^{[-i]*} - f_i\|^2$.

The first three are amplitude errors while the fourth one is the phase error. The last metric is simply based on the standard mean squared error, i.e., the L^2 distance. We use a variety of amplitude/phase error metrics for fair comparison. Note that the mean squared error $E5$ accounts for a combination of amplitude and phase errors and tends to be more sensitive to phase.

Results. The advantages of amplitude-phase kriging over other methods are summarized in Table 1; the table reports average prediction errors (with standard deviations in parentheses) over 50
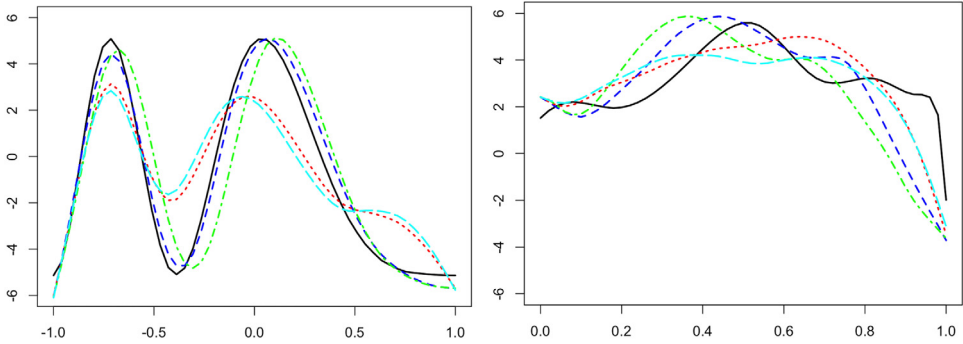
Fig. 4. Example predictions obtained via amplitude-phase kriging (blue), ordinary kriging (red), two-stage kriging (green) and universal kriging (cyan). Left: Bimodal simulation with $B = 1$. Right: B-spline Scenario 1 simulation with $B = 1$. The true function is in black. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

simulation runs. Best performance is highlighted in bold. Compared to ordinary and universal kriging, the improvement in amplitude errors is large when significant phase variation is present in the data. Although the two-stage method has similar performance to the proposed method in terms of amplitude prediction, amplitude-phase kriging shows a clear advantage in predicting the phase of target functions ($E4$).

Data simulated using the B-spline basis (under both scenarios) exhibits higher shape variation than the bimodal case, and represents the more challenging setting for both amplitude-phase kriging and the two-stage method. Nonetheless, amplitude-phase kriging outperforms ordinary and universal kriging in most cases, even when phase variation is small. In particular, when the phase components are dependent on function shapes, amplitude-phase kriging has a clear advantage in terms of the amplitude errors $E2$ and $E3$, and the phase error $E4$. Further, amplitude-phase kriging yields smaller amplitude and phase errors than two-stage kriging in this case; this is due to the spatially-informed alignment via Algorithm 1.

While the proposed approach does not outperform ordinary or universal kriging in terms of the L^2 prediction error $E5$, it has been noted in Srivastava and Klassen (2016) that the L^2 distance, which is used to define this error metric, is not a good measure of amplitude and/or phase differences. Furthermore, since ordinary and universal kriging are optimal under the L^2 metric, the results based on these measures are naturally biased towards these methods. The amplitude-phase kriging errors are mainly due to phase prediction, which is especially challenging on the boundary of the spatial domain since fewer neighbors are available.

Remark 4. In the simulated setting involving bimodal data, the performance of two-stage kriging is very similar to that of the proposed amplitude-phase kriging method, especially in terms of the amplitude errors $E1$, $E2$, and $E3$. The data in this case follows a one-dimensional model, and as such, the shapes of all functions are the same across the entire spatial domain. Thus, the global Karcher mean is very similar to the proposed spatially-weighted template.

Fig. 4 shows predictions generated by the four different methods for a single target function based on the bimodal simulation with $B = 1$ (left), and the B-spline Scenario 1 simulation with $B = 1$ (right). In general, ordinary and universal kriging fail to capture important features of functions in the predictions when phase variation is present in the data. In the left panel, the ordinary and universal kriging predictions severely underestimate the two peaks and the valley. In the right panel, the two methods yield predictions that are 'flat' over a large portion of the domain and fail to capture any of the shape patterns in the true function. The two-stage method appears to perform relatively well in terms of amplitude prediction, but does not provide a viable phase prediction. The proposed amplitude-phase kriging, on the other hand, successfully captures
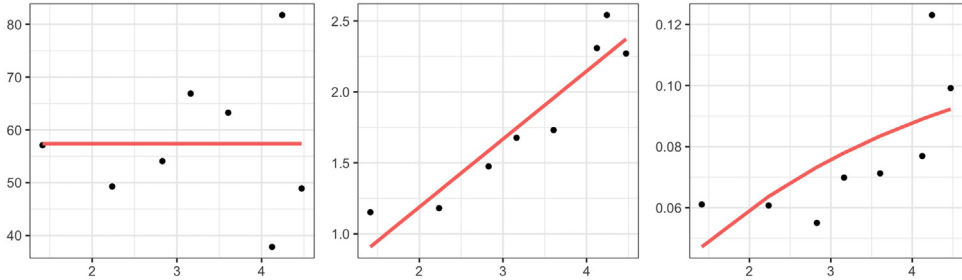
Fig. 5. Estimation of trace-variograms for prediction at site 13 under the Bimodal simulation scenario with $B = 1$; site 13 was left out and the rest of the observations were used to estimate the trace-variograms. The dots represent the empirical L^2 (ordinary kriging), amplitude and phase trace-variograms. Estimates of the trace-variograms (red curves) were obtained by fitting a Matérn variogram model to the empirical trace-variograms. For the phase trace-variogram, the estimated optimal value of the tuning parameter ω is 0.

prominent function features, as well as their magnitude, and provides satisfactory phase predictions. These improvements often result in significant decreases in the various amplitude and phase error metrics. Appendix A in the supplement reports results of another simulation study that directly explores the benefits of spatially-informed alignment in amplitude-phase kriging.

We further illustrate why amplitude-phase kriging yields better predictions than ordinary kriging in the presence of phase variation. Here, we use a single simulation run for the bimodal scenario with $B = 1$. Fig. 5 displays the empirical L^2 (left), amplitude (middle) and phase (right) trace-variograms; the fitted Matérn models are shown in red. In Fig. 6, we show the magnitude of optimal kriging coefficients for the observed data when trying to predict at site 13. Again, we consider ordinary, amplitude and phase kriging in the left, middle and right panels, respectively. Due to the 'flat' estimate of the L^2 trace-variogram, ordinary kriging assigns very similar coefficients to all of the observed functions, i.e., it fails to capture the spatial dependence in the data. On the other hand, amplitude and phase kriging result in reasonable coefficient estimates: observations in the spatial neighborhood of site 13 have largest kriging coefficients due to the strong spatial dependence in the data. The resulting estimators are shown in dashed blue at site 13. It is evident that the ordinary kriging prediction underestimates the magnitude of the two extrema; the amplitude kriging prediction is much better at capturing these features. This result is similar to the one presented in the left panel of Fig. 4. Since all of the functions in the bimodal simulation scenario have the same shape, the estimated optimal value of the tuning parameter ω is 0. Thus, function shapes do not contribute to the phase trace-variogram and phase prediction. We provide a similar set of results for the B-spline scenario with irregularly-spaced sites in Appendix E in the supplement; the findings are very similar.

6.2. Clustering

6.2.1. Simulated data

Let n denote the number of spatial sites where data was observed and I the number of clusters. Then, $n = \sum_{i=1}^{I} n_i$, where n_i is the number of functions in cluster i. Motivated by the fact that amplitude and phase in real data scenarios may exhibit different clustering patterns, we simulate the true partitions with respect to amplitude and phase separately. Our aim is to validate that the proposed amplitude-phase clustering method is able to reveal the true underlying partitions of both amplitude and phase simultaneously, irrespective of whether the spatial partitions of each component agree.

We consider two different designs: (i) where amplitude and phase cluster partitions are the same (agree), and (ii) where they are not (disagree). For (i), sites are on a 4×4 grid with integer coordinates 1, 2, 3, 4, and are partitioned into four equally sized clusters via the lines $x = 2.5$ and $y = 2.5$. For (ii), 30 sites are chosen uniformly on $[0, 4]^2$; the amplitudes are partitioned by the lines
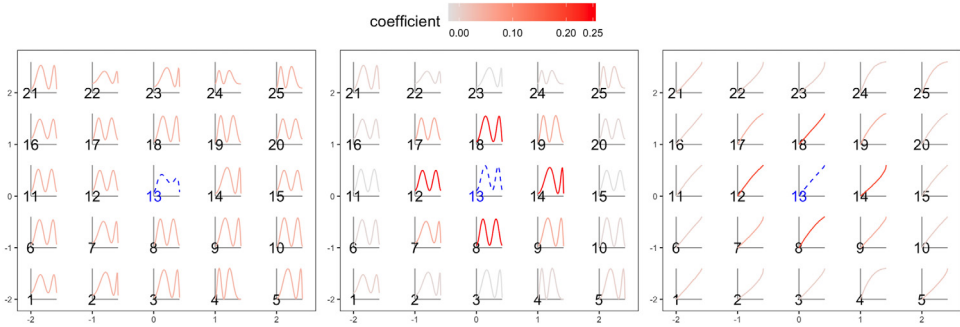
Fig. 6. Ordinary (left), amplitude (middle) and phase (right) kriging maps for prediction at site 13 under the bimodal simulation scenario with $B = 1$. The magnitude of kriging coefficients to construct the estimators (dashed blue) are shown in red.
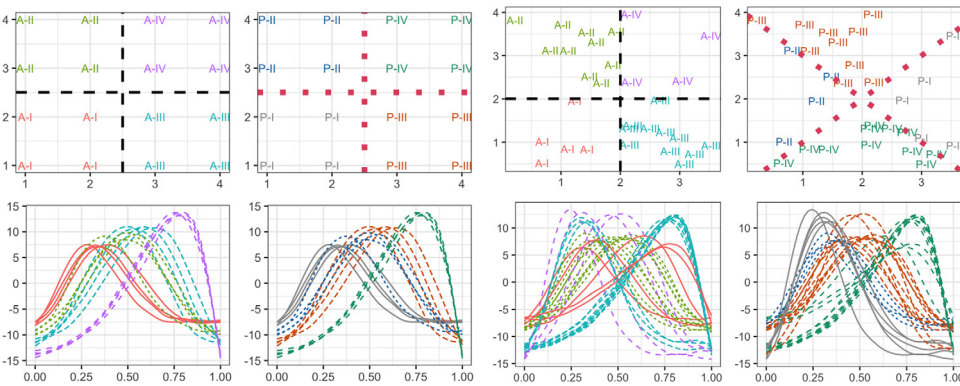

Fig. 7. Top row: Ground truth amplitude (column 1) and phase (column 2) partitions for the agree design with spatial sites on a 4×4 grid with integer coordinates. Ground truth amplitude (column 3) and phase (column 4) partitions for the disagree design with uniformly sampled spatial sites on the domain $[0, 4]^2$. Black dashed lines delineate the boundaries of the amplitude clusters, while red dotted lines delineate the boundaries of the phase clusters. Bottom row: Example dataset generated for the agree and disagree designs, with colors corresponding to the true amplitude or phase clusters displayed in the top row.

$x = 2$ and $y = 2$, while the phases are partitioned by the lines $y = x$ and $y = 4 - x$. The top row in Fig. 7 displays the two designs: the left two panels correspond to the ground truth amplitude and phase partitions for the agree design, respectively, while the right two panels display the same for the disagree design. In the bottom row, we display one example of simulated data for these two designs. The colors in each panel correspond to the ground truth clustering according to amplitude or phase.

Let f_{ij} be the jth functional observation in cluster i. We generate spatial functional data with domain $[0, 1]$ as $f_{ij}(t) = \{(a_{ij}\mu + e_{ij}) \circ \gamma_{ij}\}(t)$ ($i = 1, \ldots, I$; $j = 1, \ldots, n_i$). We set $\mu(t) = -\cos(2\pi t)$, $a_{ij} = i\delta_a + \epsilon_{a,ij}$, and γ_i as the cumulative distribution function of $Beta(1, e^{b_{ij}})$, where $b_{ij} = i\delta_b + \epsilon_{b,ij}$; δ_a and δ_b are fixed parameters that control the amplitude and phase differences between clusters, respectively. The vector $\{\epsilon_{a,ij}\}$ is generated from a multivariate normal distribution with a mean vector $(5, \ldots, 5)^T$ and Matérn covariance $C_{Mat}(\cdot, \cdot; \sigma_a^2, 0.5, \ell)$. The vector $\{\epsilon_{b,ij}\}$ follows the correlated uniform distribution on $[-B, B]^n$ with the same correlation range ℓ; e_{ij} is a zero mean Gaussian process with a diagonal covariance. We fix $\sigma_a^2 = 1$, $B = 1$, $\sigma_e = 0.5$ and $\ell = 8^{1/2}$. In the bottom row of Fig. 7, we display one example of simulated data for the agree and disagree designs. The colors in each panel correspond to the ground truth clustering according to amplitude or phase.

Table 2

Average rand indices (SD) for estimated partitions based on amplitude-phase clustering (APC) and L^2-based clustering (L^2C), with respect to the true amplitude and phase clusters.

δ_a	δ_b	Method	Agree		Disagree	
			Amplitude	Phase	Amplitude	Phase
1	0.1	APC	**0.828 (0.106)**	0.751 (0.101)	**0.808 (0.107)**	**0.711 (0.088)**
		L^2C	0.772 (0.079)	**0.772 (0.079)**	0.731 (0.085)	**0.711 (0.067)**
	0.5	APC	0.870 (0.091)	**0.958 (0.052)**	**0.752 (0.086)**	**0.887 (0.083)**
		L^2C	**0.910 (0.072)**	0.910 (0.072)	0.701 (0.046)	0.877 (0.07)
2	0.1	APC	**0.945 (0.067)**	0.767 (0.104)	**0.916 (0.076)**	0.708 (0.087)
		L^2C	0.808 (0.082)	**0.808 (0.082)**	0.779 (0.085)	**0.742 (0.068)**
	0.5	APC	**0.949 (0.071)**	**0.955 (0.059)**	**0.835 (0.085)**	**0.879 (0.087)**
		L^2C	0.908 (0.080)	0.908 (0.080)	0.707 (0.055)	0.873 (0.071)
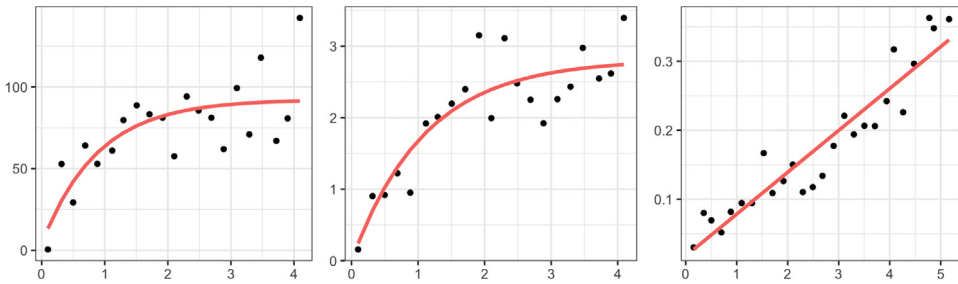
Fig. 8. Estimation of trace-variograms for clustering under the disagree design with $\delta_a = 2$ and $\delta_b = 0.5$. The dots represent the empirical L^2, amplitude and phase trace-variograms. Estimates of the trace-variograms (red curves) were obtained by fitting a Matérn variogram model to the empirical trace-variograms.

6.2.2. Comparison with another method

We repeat each simulation 100 times, and compare amplitude-phase clustering (APC) to the L^2 distance-based method (L^2C) (Giraldo et al., 2012) using the rand index (Rand, 1971).

Results. The means and standard deviations of the rand indices for each design, and different choices of δ_a and δ_b, are shown in Table 2; best performance is highlighted in bold. The proposed amplitude-phase clustering approach outperforms the L^2 distance-based method in most scenarios, even when the amplitude and phase partitions agree. When the true partitions are different, the amplitude-phase clustering is far superior, especially for the larger values of δ_a and δ_b. The L^2-based approach is always forced to compromise between the true amplitude and phase clusters, while the proposed approach treats them separately. Further, the L^2 metric is sensitive to phase differences. As a result, when δ_b is large, it captures the phase clustering and exhibits similar performance to the proposed method in that regard. However, it is unable to recover the true amplitude clusters. Fig. 8 shows the empirical and fitted trace-variograms for particular values of δ_a and δ_b, and demonstrates that spatial dependence is captured by all variograms, increasing the chance of grouping the subjects with stronger spatial correlation.

7. Real data analysis

7.1. Kriging of daily ozone data in north california

We apply the proposed amplitude-phase kriging method to U.S. daily ozone data, available on the air data website[1] of the United States Environmental Protection Agency. We focus on a small

---

[1] https://www.epa.gov/outdoor-air-quality-data.

Table 3

Leave-one-out cross-validation average prediction errors of amplitude-phase kriging (APK), two-stage kriging (TSK) and ordinary kriging (OK) for the ozone data in North California. All values were multiplied by 1000.

Method	$E1$	$E2$	$E3$	$E4$	$E5$
APK	4.88	**7.57**	**1.94**	49.02	7.17
TSK	4.67	8.79	2.21	**45.9**	6.6
OK	**4.17**	8.01	2.23	53.64	**6.44**
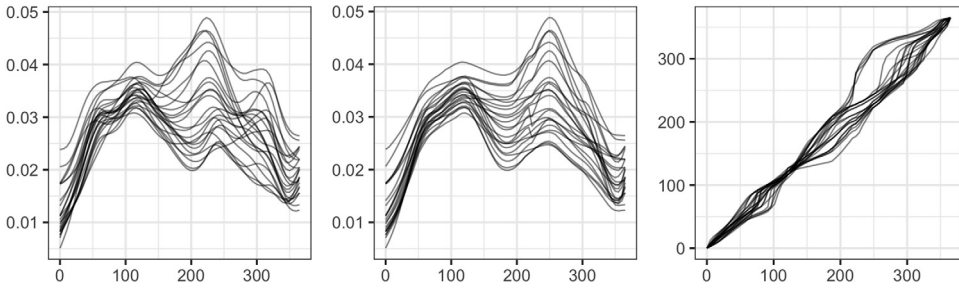
Fig. 9. After leaving out the observation at site 8, the remaining ozone concentration functions observed at 23 other sites (left) are aligned to the estimated amplitude prediction at site 8 using Algorithm 1, resulting in separate amplitude (middle) and phase (right) components.

area in North California ($35° \sim 39°$ N, $120 \sim 123°$ W) with 24 observation stations. Each station recorded daily average ozone concentration (parts per million) for the year 2018. We smooth the data using splines with smoothing parameter $\iota = 3 \times 10^{-4}$. We evaluate the effect of smoothing on kriging performance in Appendix C in the supplement.

7.1.1. Results

We use leave-one-out cross-validation on the 24 smoothed observations and report the mean of the five error metrics, $E1$–$E5$, for ordinary kriging (OK), two-stage kriging (TSK) and amplitude-phase kriging (APK) in Table 3; best performance is highlighted in bold. We do not compare to universal kriging here since this approach focuses on kriging residual functions after accounting for a spatially varying mean function. The proposed method outperforms ordinary kriging in terms of the reported amplitude/phase error metrics $E2$, $E3$ and $E4$. The amplitude and phase mean squared errors ($E3$ and $E4$) of amplitude-phase kriging are 13% and 8.6% smaller, respectively, compared to ordinary kriging. This shows that combining separate amplitude and phase predictions has a clear advantage in real data scenarios. Compared to two-stage kriging, amplitude-phase kriging generates more accurate amplitude predictions as evidenced by smaller $E2$ and $E3$ errors. This is most likely due to moderate shape variation among the spatial functional data. Two-stage kriging outperforms amplitude-phase kriging in terms of the phase error $E4$.

Focusing on site 8, we present more detailed alignment and kriging results based on the proposed approach. We present the results of amplitude-phase separation, computed via Algorithm 1, in Fig. 9. The given spatial functional data (except for the datum observed at site 8) is given in the left panel. It is clear that phase variation is present in the sample. The estimated warping functions, with respect to the amplitude kriging predictor at site 8, are shown in the right panel; phase variation in the ozone concentration functions is mainly due to local delays/advances in the timeline, which represent significant deviations from identity warping. The middle panel displays the aligned data. The right panel in Fig. 10 highlights the advantage of amplitude-phase kriging as compared to ordinary kriging: between days 200 and 300, where significant phase variation is present, amplitude-phase kriging is much more effective at predicting the shape of the function at site 8. In particular, ordinary kriging underestimates the magnitude of the second peak of ozone concentration. Accurate prediction of the phase component is difficult in practice since its definition
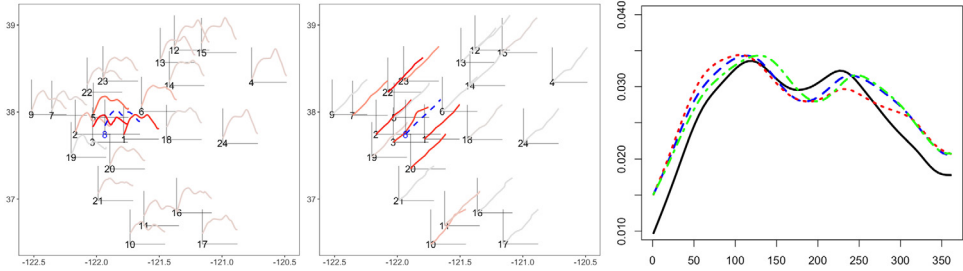
Fig. 10. Amplitude-phase kriging of amplitude (left) and phase (middle) components at site 8. The solid curves are estimated amplitude and phase components at observed sites; dashed blue lines are the predicted amplitude and phase components. The color shading shows the contribution (from 0 to 1) from each site to the prediction. Right: The prediction for site 8 with the true function (black) and predictions obtained via amplitude phase-kriging (blue), ordinary kriging (red) and two-stage kriging (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

depends on the shape of functional data. From the left and middle panels in Fig. 10, we can see that amplitude kriging generally borrows information from neighboring sites since we only use the spatial coordinates (distance) to model the dependency in this case. On the other hand, in phase kriging, we include both the spatial locations and the shape of the observed functions to model the dependency. Thus, the highest contribution into the final kriging estimate is a combination of phase functions that are spatially nearby, and those that correspond to observed functions that have a similar shape to the predicted amplitude. Furthermore, the spatial dependency in the phase component is generally fairly weak. This is why many previous studies prefer to treat phase variability as noise. However, in this real data analysis, we have found that even if the phase signal is not as strong as the amplitude signal, separate amplitude and phase prediction is still beneficial as evidenced in Table 3 and the right panel in Fig. 10.

7.2. Clustering of canadian weather data

We apply the proposed amplitude-phase clustering method to the Canadian weather data (Ramsay and Silverman, 2005). The data can be found in the R package 'fda' (Ramsay et al., 2020). In this paper, we analyze the daily temperature data averaged over 1960–1994, collected at 35 stations in Canada. Because the 35 stations cover a large area, we first filter out the longitudinal and latitudinal trends by fitting a functional linear regression model where longitude and latitude are included as covariates; the same approach was taken in Giraldo et al. (2012). The resulting functional residuals are then smoothed using splines (with low smoothing parameter $\iota = 5 \times 10^{-5}$) and used as the data for clustering.

7.2.1. Results

We use the clustering method described in Section 5 and compare the results to the L^2 metric-based clustering of Giraldo et al. (2012). The empirical and fitted L^2, amplitude and phase trace-variograms are shown in Fig. 11. There is evidence of spatial correlations in each of them and the amplitude trace-variogram has a smaller range than the L^2 one. We further observe that the Matérn model fits the empirical amplitude and phase trace-variograms better than the L^2 one, since some quadratic patterns are present in the latter. Values from these fitted variograms are plugged into the dissimilarity measures as weights for clustering.

The hierarchical clustering trees as well as the clustering results on the map of Canada are shown in Fig. 12. Based on separate clustering of amplitude and phase, we discover some interesting results. First, the amplitude and phase clusterings agree in Western and Central Canada. The cities located on the West Coast are further partitioned into South and North clusters, while the cities in the Central region are in a single cluster. Second, the difference between amplitude and phase
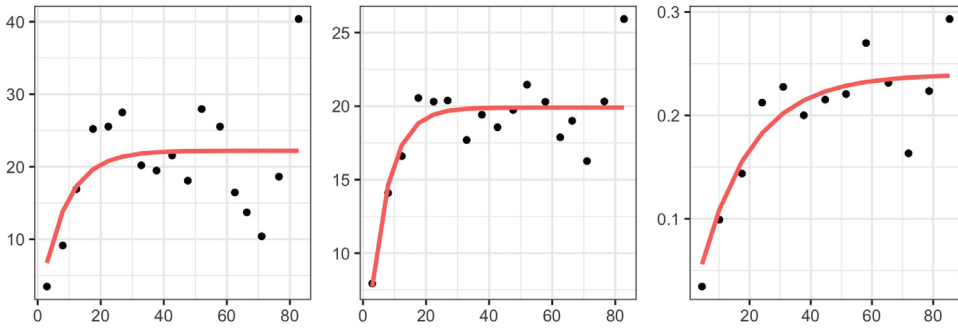
Fig. 11. Estimation of trace-variograms for clustering of the Canadian weather data. The dots represent the empirical L^2, amplitude and phase trace-variograms. Estimates of the trace-variograms (red curves) were obtained by fitting a Matérn variogram model to the empirical trace-variograms.

variation mainly appears in the clustering of Resolute, Iqaluit and St. Johns. Specifically, Resolute and St. Johns are clustered together based on amplitude due to similar magnitude (and shape) of the residual functions whereas Iqaluit is clustered separately. In terms of phase, Iqaluit and St. Johns are included in the large cluster in Southeast Canada, but Resolute forms its own cluster. This is due to a large phase distance between the Resolute residual function and the Iqaluit/St. Johns residual functions: the Resolute function reaches its peak earlier than the other two and has a much longer plateau. Third, compared to the L^2-based clustering method (bottom panel in Fig. 12), amplitude-phase clustering yields fewer clusters and the clusters tend to be more spatially connected. For example, the three cities in the Northwest are clustered together based on amplitude-phase clustering whereas L^2-based clustering separates them into three different clusters. Further, based on L^2-based clustering, we observe an unnatural result: Resolute, a station in the Arctic Circle, is clustered with the Vancouver and Victoria stations on the West Coast and St. Johns on the East Coast. We also applied hierarchical clustering without spatial weighting to the same dataset (see Appendix E in the supplement). It is clear that involving spatial dependency in the clustering helps preserve connectivity of adjacent sites, making the results more interpretable.

8. Discussion

It is difficult to verify the key assumptions of stationarity and isotropy for spatial functional data, especially when one decouples amplitude and phase components, which effectively results in two sets of functional data. Despite this, when deviation from stationarity is not too large, the amplitude and conditional phase trace-variograms provide useful summary statistics of spatial variation. Results from simulations and real data analyses offer corroboration. Although some of the presented empirical amplitude and (conditional) phase trace-variograms could indicate a spatial trend in the mean, it is difficult to assess whether the stationarity assumption has in fact been violated. If a spatial trend in the mean is of concern, one can adapt the proposed framework to a universal amplitude-phase kriging approach. This extension is non-trivial as it requires a regression framework for amplitude and phase. For amplitude, the specified regression model must be invariant to warping. For phase, considerable difficulties are posed by the non-Euclidean nature of its representation space. We will consider this extension in future work.

Based on simulations in Section 6, the phase trace-variogram is more informative when allowing for the plausibility that two nearby functions share similar shapes. Thus, a rigorous study of the conditional phase trace-variogram, when conditioned on the shape random field, will add further insight, but is beyond the scope of this paper. The main challenge will be to reconcile the variogram definition with the property that a phase functional random field will almost never satisfy stationarity since it is only interpreted in a *relative* sense.
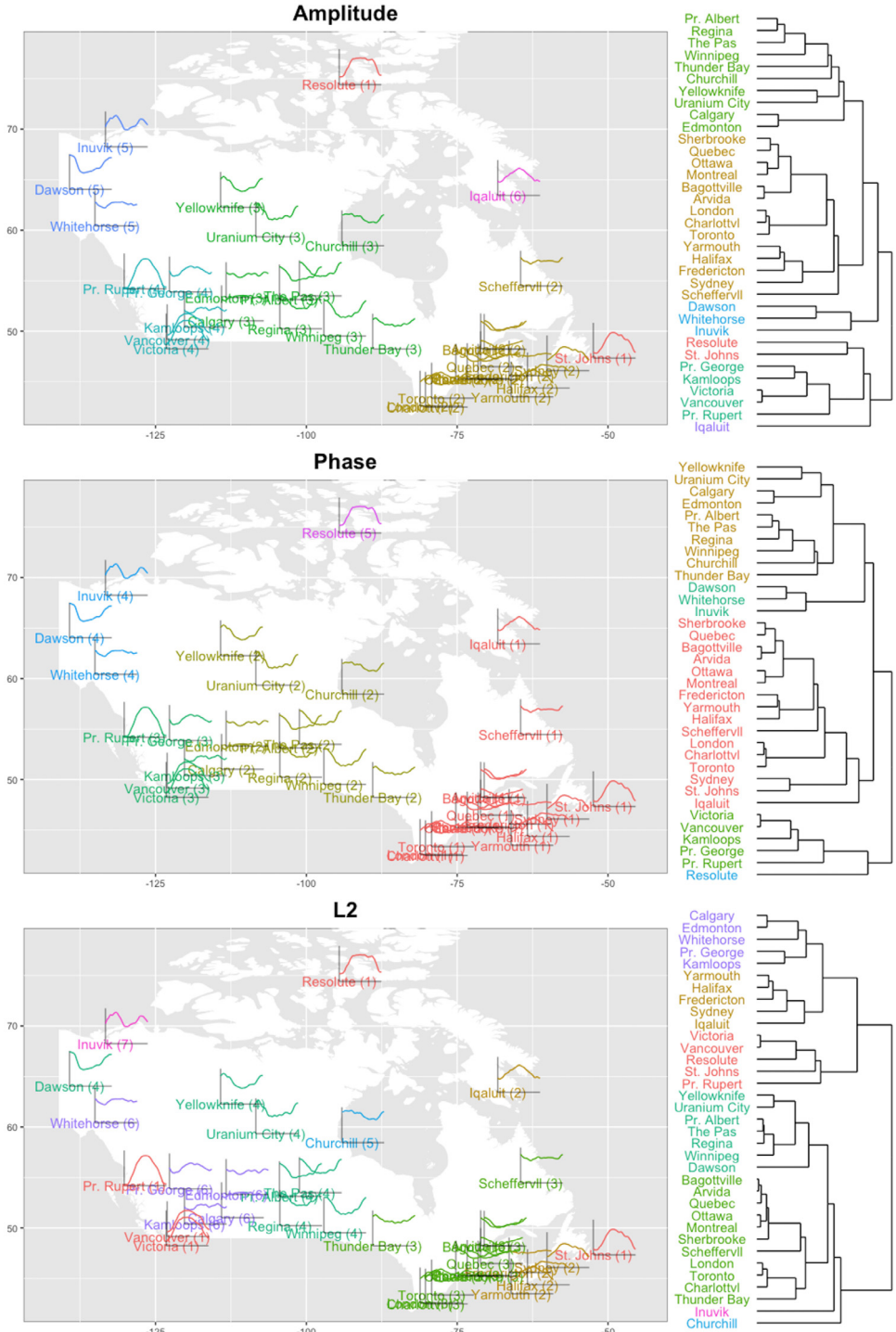
Fig. 12. Clustering (average linkage, clusters in different colors) of functional residuals, after adjusting for latitude and longitude effects, obtained from the Canadian weather data.

Evidently, formulating phase variation as an isometric action by the group of warping functions plays a crucial role in this paper. This is enabled by adopting the square-root slope transform $f \to q$, which maps $f \circ \gamma \to (q, \gamma)$ under which $\|(q, \gamma)\| = \|q\|$. The property that warping or phase variation does not affect the function's norm drives our definitions of variograms V_a and V_p, their estimators, and is used when constructing, and examining properties of, kriging estimates (see proofs of Propositions 1 and 2 in Appendix B in the supplement as well as Proposition 3 and its proof in Appendix C in the supplement). Hence, although several methods for amplitude-phase separation are available, our developments in this paper complement the ubiquity of the L^2 metric in functional data analysis through the use of the square-root slope transform.

Extensions of developments in this paper to the setting of noisy, sparse spatial functional data constitute ongoing work. Variability due to considerable nonparametric or model-based smoothing will then need to be considered in addition to amplitude, phase and spatial variabilities. Promisingly, results here represent the first foray towards analyzing spatial complex functional data objects such as shapes of curves (Srivastava and Klassen, 2016) and surfaces (Jermyn et al., 2017) by decoupling spatial, shape and nuisance variations.

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spasta.2022.100687. Appendix A in the supplement includes further numerical assessments of the performance of Algorithm 1. Appendix B contains proofs of Propositions 1 and 2. Appendix C provides a theoretical discussion of the convergence of amplitude kriging under a one-dimensional model. Appendix D offers a discussion of invariance of amplitude-phase clustering to the global scale of the amplitude and (conditional) phase trace-variograms. Finally, Appendix E contains additional simulation results and visualizations

References

Abramowicz, K., Arnqvist, P., Secchi, P., De Luna, S.S., Vantini, S., Vitelli, V., 2017. Clustering misaligned dependent curves applied to varved lake sediment for climate reconstruction. Stoch. Environ. Res. Risk Assess. 31 (1), 71–85.

Adler, R.J., Taylor, J.E., 2007. Random Fields and Geometry. Springer.

Aguilera-Morillo, M.C., Durbán, M., Aguilera, A.M., 2017. Prediction of functional data with spatial dependence: a penalized approach. Stoch. Environ. Res. Risk Assess. 31 (1), 7–22.

Bernardi, M.S., Sangalli, L.M., Mazza, G., Ramsay, J.O., 2017. A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province. Stoch. Environ. Res. Risk Assess. 31 (1), 23–38.

Bourgault, G., Marcotte, D., Legendre, P., 1992. The multivariate (co) variogram as a spatial weighting function in classification methods. Math. Geol. 24 (5), 463–478.

Caballero, W., Giraldo, R., Mateu, J., 2013. A universal kriging approach for spatial functional data. Stoch. Environ. Res. Risk Assess. 27 (7), 1553–1563.

Chakraborty, A., Panaretos, V., 2021. Functional registration and local variations: Identifiability, rank, and tuning. Bernoulli 27 (2), 1103–1130.

Cressie, N., 2015. Statistics for Spatial Data. John Wiley & Sons.

Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-Temporal Data. John Wiley & Sons.

Delicado, P., Giraldo, R., Comas, C., Mateu, J., 2010. Statistics for spatial functional data: some recent contributions. Environmetrics 21 (3–4), 224–239.

Giraldo, R., Delicado, P., Mateu, J., 2011. Ordinary kriging for function-valued spatial data. Environ. Ecol. Stat. 18 (3), 411–426.

Giraldo, R., Delicado, P., Mateu, J., 2012. Hierarchical clustering of spatially correlated functional data. Stat. Neerl. 66 (4), 403–421.

Giraldo, R., Delicado, P., Mateu, J., 2020. Geofd: Spatial prediction for function value data. R package version 2.0 URL https://CRAN.R-project.org/package=geofd.

Goulard, M., Voltz, M., 1993. Geostatistical interpolation of curves: a case study in soil science. In: Geostatistics TrÓIa. Springer, pp. 805–816.

Haggarty, R., Miller, C., Scott, E., 2015. Spatially weighted functional clustering of river network data. J. R. Stat. Soc. Ser. C. Appl. Stat. 64 (3), 491–506.

Jermyn, I.H., S., K., Laga, H., Srivastava, A., 2017. Elastic Shape Analysis of Three-Dimensional Objects. Morgan & Claypool.

Kurtek, S., Srivastava, A., 2011. Signal estimation under random time-warpings and nonlinear signal alignment. In: Advances in Neural Information Processing Systems. pp. 676–683.

Marron, J.S., Ramsay, J.O., Sangalli, L.M., Srivastava, A., 2015. Functional data analysis of amplitude and phase variation. Statist. Sci. 30 (4), 468–484.

Martínez-Hernádez, I., Genton, M.G., 2020. Recent developments in complex and spatially correlated functional data. arXiv preprint arXiv:2001.01166.

Mateu, J., Romano, E., 2017. Advances in spatial functional statistics. Stoch. Environ. Res. Risk Assess. 31 (1), 1–6.

Matuk, J., Bharath, K., Chkrebtii, O., Kurtek, S., 2021. Bayesian framework for simultaneous registration and estimation of noisy, sparse, and fragmented functional data. J. Amer. Statist. Assoc. 1–17.

Menafoglio, A., Petris, G., 2016. Kriging for Hilbert-space valued random fields: The operatorial point of view. J. Multivariate Anal. 146, 84–94.

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., Dalla Rosa, M., 2013. A universal kriging predictor for spatially dependent functional data of a Hilbert space. Electron. J. Stat. 7, 2209–2240.

Nerini, D., Monestiez, P., Manté, C., 2010. Cokriging for spatial functional data. J. Multivariate Anal. 101 (2), 409–418.

Oliver, M., Webster, R., 1989. A geostatistical basis for spatial weighting in multivariate classification. Math. Geol. 21 (1), 15–35.

Ramsay, J.O., Graves, S., Hooker, G., 2020. Fda: Functional data analysis. R package version 5.1.5 URL https://CRAN.R-project.org/package=fda.

Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis. Springer, New York, NY.

Rand, W.M., 1971. Objective criteria for the evaluation of clustering methods. J. Amer. Statist. Assoc. 66 (336), 846–850.

Reyes, A., Giraldo, R., Mateu, J., 2015. Residual kriging for functional spatial prediction of salinity curves. Comm. Statist. Theory Methods 44 (4), 798–809.

Ribeiro Jr, P.J., Diggle, P.J., Schlather, M., Bivand, R., Ripley, B., 2020. GeoR: Analysis of geostatistical data. R package version 1.8-1 URL https://CRAN.R-project.org/package=geoR.

Romano, E., Balzanella, A., Verde, R., 2010. Clustering spatio-functional data: a model based approach. In: Classification As a Tool for Research. Springer, pp. 167–175.

Romano, E., Balzanella, A., Verde, R., 2017. Spatial variability clustering for spatially dependent functional data. Stat. Comput. 27 (3), 645–658.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.

Sangalli, L.M., Secchi, P., Vantini, S., Vitelli, V., 2010. K-mean alignment for curve clustering. Comput. Statist. Data Anal. 54 (5), 1219–1233.

Schmidt, A.M., Guttorp, P., O'Hagan, A., 2011. Considering covariates in the covariance structure of spatial processes. Environmetrics 22 (4), 487–500.

Secchi, P., Vantini, S., Vitelli, V., 2013. Bagging Voronoi classifiers for clustering spatial functional data. Int. J. Appl. Earth Obs. Geoinf. 22, 53–64.

Srivastava, A., Klassen, E.P., 2016. Functional and Shape Data Analysis, Vol. 475. Springer.

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., Marron, J.S., 2011. Registration of functional data using Fisher-Rao metric. arXiv preprint arXiv:1103.3817.

Strait, J., Kurtek, S., Bartha, E., MacEachern, S.N., 2017. Landmark-constrained elastic shape analysis of planar curves. J. Amer. Statist. Assoc. 112 (518), 521–533.

Tucker, J.D., 2021. Fdasrvf: Elastic functional data analysis. R package version 1.9.7 URL https://CRAN.R-project.org/package=fdasrvf.

Wang, J.-L., Chiou, J.-M., Müller, H.-G., 2016. Functional data analysis. Annu. Rev. Stat. Appl. 3, 257–295.