Test
DOI 10.1007/s11749-011-0268-9

RESEARCH PAPER

On the definition of phase and amplitude variability in functional data analysis

Simone Vantini

Accepted: 16 September 2011
© Sociedad de Estadística e Investigación Operativa 2011

Abstract We introduce a modeling and mathematical framework in which the problem of registering a functional data set can be consistently set. In detail, we show that the introduction, in a functional data analysis, of a metric/semi-metric and of a group of warping functions, with respect to which the metric/semi-metric is invariant, enables a sound and not ambiguous definition of phase and amplitude variability. Indeed, in this framework, we prove that the analysis of a registered functional data set can be re-interpreted as the analysis of a set of suitable equivalence classes associated to original functions and induced by the group of the warping functions. Moreover, an amplitude-to-total variability index is proposed. This index turns out to be useful in practical situations for measuring to what extent phase variability affects the data and for comparing the effectiveness of different registration methods.

Keywords Functional data analysis · Phase variability · Amplitude variability · Registration · Alignment · Synchronization · Warping

Mathematics Subject Classification (2000) 62H05 · 62H35 · 62H99 · 62A01

1 Introduction

The problem of data registration (also known as curve alignment or synchronization) is often encountered in the recent statistical literature. Generally speaking, it is encountered when the ith subject can be thought associated to an unknown function $y_i(x)$ for which some point-wise (and maybe noisy) evaluations are available and the

Communicated by Domingo Morales.

S. Vantini (✉)
MOX Laboratory, Department of Mathematics "Francesco Brioschi", Politecnico di Milano, Milano, Italy
e-mail: simone.vantini@polimi.it

Published online: 14 October 2011

variability between subjects is assumed to be related not only to the dependent variable y but also to the independent one x; in a very wide sense, registering data means identifying this second source of variability and removing it by means of subject-dependent suitable transformations of the independent variable.

Statistical papers devoted to this issue can be parted in two groups: the *longitudinal data analysis* (LDA) inspired works and the *functional data analysis* (FDA) inspired ones. Some examples relevant to the former approach can be found in Lawton et al. (1972), Lindstrom and Bates (1990), Ke and Wang (2001), Altman and Villarreal (2004), and Brumback and Lindstrom (2004). Instead, some examples relevant to the latter approach—which the present work belongs to—can be found in Ramsay and Li (1998), Kneip et al. (2000), Liu and Müller (2004), Ramsay and Silverman (2005), James (2007), Kaziska and Srivastava (2007), Kneip and Ramsay (2008), Sangalli et al. (2009), Tang and Muller (2009), and Sangalli et al. (2010).

Even though many real-world problems can indifferently be tackled by means of both approaches, some differences occur between data sets that are typically analyzed by means of LDA techniques or FDA techniques: indeed, in LDA-inspired works, data typically show low within-subject signal-to-noise ratio and small within-subject sample size while, in FDA-inspired works, they typically show high within-subject signal-to-noise ratio and large within-subject sample size. Maybe also because of these differences, the two approaches are remarkably different also from a modeling perspective: LDA-inspired works usually consider data as realizations of random vectors whose distributions are fully modeled in a parametric way; FDA-inspired works are instead typically non-parametric, very few modeling assumption are postulated and observations are considered as discrete point-wise evaluations of smooth underlying random functions. Consequently, in LDA-inspired works, model characterization is achieved by means of likelihood maximization (for instance, the problem of data registration is simply managed by introducing and estimating subject-specific random effects), while in FDA-inspired works, model characterization is achieved by means of minimization or maximization of functionals defined in a suitable ∞-dimensional functional space which data are assumed to belong to. The similarities and differences between LDA and FDA are widely discussed in Ke and Wang (2001), Davidian et al. (2004), and Valderrama (2007).

FDA is a late but quickly growing branch of statistics that considers data sets made of curves or surfaces as realizations of random functions (i.e., random variables whose image is an ∞-dimensional functional space). First theoretical studies about ∞-dimensional random variables are dated to the beginning of the twentieth century while first applications to real-world data are found in the last decade of the same century. This gap is probably mainly due to the technology development: indeed, on the one hand, technology has made the acquisition of almost time and/or space-continuous measures (representable as curves or surfaces in general) more and more frequent in engineering, physics, economy, medicine, climatology, and in many other fields; and, on the other hand, the technological progress has also provided the statistical community with computational facilities for dealing with these kind of data.

Since the 1990s, the research activity in FDA has been continuously growing. As a result, a number of monographs entirely devoted to FDA and focusing on both theory and applications have been published (e.g., Ramsay and Silverman 2005; Ferraty and Vieu 2006; Ferraty and Romain 2011) and also some top level journals have recently dedicated special issues to FDA: *Statistica Sinica* edited by Davidian et al. (2004), *Computational Statistics and Data Analysis* by González Manteiga and Vieu (2007), *Computational Statistics* by Valderrama (2007), and *Journal of Multivariate Analysis* by Ferraty (2010).

Most of the FDA tools that are commonly used in applications are derived from extensions to ∞-dimensional separable Hilbert spaces (typically L^2) of corresponding multivariate analysis tools previously developed for finite-dimensional data. Because of the general non-existence of the probability density function for random functions, these extensions can be quite troublesome, and thus, nowadays, a lot of effort has been put in this direction. For instance, Delaigle and Hall (2010) defined a concept of log-density for functional data and the consequent concept of modal function; Ferraty et al. (2010) proved the uniform consistency of kernel estimators in functional data analysis; Cuevas et al. (2007) compared five different notions of depth and the following concepts of median function and trimmed-mean function.

These recent works and many others rely on the concept of *small ball probability* (Ferraty and Vieu 2006) which, in some sense, replaces the concept of probability density in functional data analysis at a certain resolution scale. The concept of small ball probability is intrinsically related to the metric (or semi-metric) nature of the functional space which functional data belong to, and not to its Hilbertian nature. The idea that metric (or semi-metric) spaces are the natural setting for a theoretical and practical investigation of many FDA tools first appeared in a pioneering work by Ferraty and Vieu (2002); in that work, in the framework of non-parametric functional regression, the authors show, in a real application and in a simulation study, how the use of semi-metrics different from the one induced by the L^2-norm can dramatically improve the predictive power of a regression model. More recently, Ferraty et al. (2010) explicitly stated that “in fact, as a statistician, an important task consists in building a semi-metric *adapted* to the functional variable”, i.e., a semi-metric satisfying some requested properties giving soundness and theoretical support to the analysis. The present work shares this idea with the latter work: indeed, we will propose a property of invariance—with respect to a group of transformations—that, we think, should be requested by the semi-metrics used in functional data analysis if a functional data registration is thought to be needed (i.e., if part of the variability is considered due to the independent variable).

Functional data registration is often a necessary step to achieve a successful functional data analysis. Naively speaking, a registration of a functional data set is considered to be any procedure that aims at making the n observed functions as similar as possible by means of n suitable transformations of the abscissas. These transformations are commonly named warping functions, and the variability of the functional data set imputable to them is usually named *phase variability*; finally, the residual variability observed among the aligned functions is named *amplitude variability*. Many basic tools of functional data analysis exclusively focus on amplitude variability (e.g., mean curve or functional principal component analysis), indeed, the evaluations of the functions at the same abscissa are implicitly assumed as realizations of independent and identically distributed random functions. If phase variability is present in the data, which is not necessarily true, and thus, if its presence is neglected, it can act as a confounding factor jeopardizing the whole statistical analysis because of wrong matchings occurring across subjects or, more technically—as we will show—because of the wrong metric (or semi-metric) used.

Although many successful methods have already been proposed in the literature, a clear theoretical analysis about the soundness and the meaningfulness of the problem of functional data registration is still missing: Davidian et al. (2004) mentions the problem of curve registration as a "domain which has attracted a certain level of attention but could still benefit from further study". This work is one of the first attempts to put in a coherent mathematical framework this key problem of functional data analysis with the aim of providing statisticians with a clear mind-set and a practical tool through which setting and comparing different registration methods. Recently, also in Kneip and Ramsay (2008), some effort has been done in this direction. Though that work and the present one differ in many aspects (indeed, the driving idea of Kneip and Ramsay (2008) is the concept of amplitude convex space of the aligned functions, while here the driving idea is the concept of phase equivalence classes for the non-aligned functions), the basic assumptions which each approach relies on do not appear incompatible, hopefully leaving space for a possible future integration of the two. It is worth mentioning also a less recent work by Liu and Müller (2004): this work—even if of no practical interest since it does not suggest any idea about how to register a functional data set in practice—has the peculiarity of being the first paper, to our knowledge, to propose a metric that explicitly takes into account amplitude and phase variability.

The present paper is structured as follows: in Sect. 2, the issue of registering a function with respect to another one is tackled. In detail: in Sect. 2.1, the necessary mathematical framework, which will also be used in the following sections, is introduced; in Sect. 2.2, the problem of registering a function with respect to another is declined in the introduced mathematical framework and an amplitude-to-total variability ratio α^2 is proposed; in Sect. 2.3, a discussion about the number of equivalence classes related to phase variability is undertaken. In Sects. 3 and 4, the theory developed in Sect. 2 is generalized to the problem of registering a set of functions in the presence and absence of a reference function, respectively. In Sect. 5, the theory is further extended to deal with the registration of functions when a semi-metric is used in place of a metric. In Sect. 6, a couple of real applications presented in the literature are presented and discussed in the light of the theory here presented. Finally, in Sect. 7, the links between our approach and other recent directions of research also related to the concept of semi-metric are discussed and some hints for future research are proposed.

2 Registration of a pair of functions

In this section, we deal with the easiest case of functional data registration: this is the problem of registering two functions one with respect to the other one.

2.1 Mathematical framework

In order for a functional data registration problem to be meaningful and mathematically consistent according to our theory, some basic properties of the set F which the functional data belong to and of the set W of warping functions are demanded:

(a) $F = \{f : \Omega \subseteq \mathbb{R}^p \to \Psi \subseteq \mathbb{R}^q\}$ is a metric space equipped with a metric $d : F \times F \to \mathbb{R}_0^+$;
(b) W is a subgroup—with respect to ordinary composition $\circ$—of the group of the continuous automorphisms: $\Omega \subseteq \mathbb{R}^p \to \Omega \subseteq \mathbb{R}^p$;
(c) $\forall f \in F$ and $\forall h \in W$ we have that $f \circ h \in F$;
(d) Given any couple of elements $f_1, f_2 \in F$ and an element $h \in W$, the distance between f_1 and f_2 is invariant under the composition of f_1 and f_2 with h, i.e.,

$$d(f_1, f_2) = d(f_1 \circ h, f_2 \circ h); \tag{1}$$

we will refer to this property as W-*invariance* of d.

Thanks to properties (a)–(d), it is possible to define a semi-metric $d_W : F \times F \to \mathbb{R}_0^+$ that is jointly determined by the metric d and the group W (a proof can be found in Appendices A and B):

Theorem 1 (Definition of the semi-metric d_W) *Under properties* (*a*)–(*d*),

$$d_W(f_1, f_2) := \min_{h_1, h_2 \in W} d(f_1 \circ h_1, f_2 \circ h_2),$$

when defined, *is a semi-metric*.

Sufficient conditions for the existence of the minimum are reported in Appendices A and B. Like any other semi-metric, d_W induces a partition of the space F in to a quotient set that we will indicate as $\mathcal{F}$ (i.e., two functions belong to the same equivalence class of $\mathcal{F}$ if their semi-distance is zero). The W-invariance of the original metric d provides a bijective correspondence between the equivalence classes of the quotient set $\mathcal{F}$ and the orbits of the action of the group W on the set F. Thus, we can define a metric $d_{\mathcal{F}} : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_0^+$ on $\mathcal{F}$ that is consistent with the original metric d on F (a proof can be found in Appendices A and B); let $[f]$ indicate the equivalence class of $\mathcal{F}$ which f belongs to:

Theorem 2 (Definition of the metric $d_{\mathcal{F}}$) *Under properties* (*a*)–(*d*),

$$d_{\mathcal{F}}\big([f_1], [f_2]\big) := d_W(f_1, f_2)$$

is a metric.

In Sect. 2.2, we will show how the notions introduced above can be used to define a sound notion of phase and amplitude variability. Before that, it is useful to report some bounding properties linking $d_{\mathcal{F}}([f_1],[f_2])$ with $d(f_1, f_2)$ (proofs are in Appendices A and B):

$$0 \leq d_{\mathcal{F}}\big([f_1],[f_2]\big) \leq d(f_1, f_2), \tag{2}$$

with bounds characterized as follows:

$$d_W(f_1, f_2) = 0 \Leftrightarrow \exists h_1, h_2 \in W \quad \text{such that } f_1 \circ h_1 = f_2 \circ h_2, \tag{3}$$

$$d_W(f_1, f_2) = d(f_1, f_2) \Leftrightarrow h_1 = h_2 = \mathbf{1} \quad \text{is a minimizing couple.} \tag{4}$$

Remarks The minimizing couple, if it exists, is never unique (except for $W = \{\mathbf{1}\}$). Indeed, because of the W-invariance of d, if (h_1, h_2) is a minimizing couple, any other couple of the form $(h_1 \circ h, h_2 \circ h)$ with $h \in W$ is still a minimizing couple. Thus, without loss of generality, h_1 (or h_2) can be fixed equal to a convenient element of W—for instance, $\mathbf{1}$—and h_2 (or h_1) taken accordingly.

2.2 The problem of registration revisited

We are now ready to revisit the problem of functional data registration in the light of the results shown in the previous section. We will start from the problem of registering a pair of functions f_1 and f_2. In order to help the reader, in Fig. 1, a naive representation of this revisit is reported.

Definition 3 Functions $\tilde{f}_1 \in [f_1]$ and $\tilde{f}_2 \in [f_2]$ are said to be *mutually-registered representatives* of equivalence classes $[f_1]$ and $[f_2]$ (or in more familiar terms, simply *mutually-registered*) if and only if $d(\tilde{f}_1, \tilde{f}_2) = d_{\mathcal{F}}([f_1],[f_2])$.

In other words, two functions are mutually-registered representatives of their equivalence classes if and only if the distance between the two functions coincides with the distance between their respective equivalence classes. By Theorems 1 and 2, we have the following equivalent definition of *mutually-registered representatives* of $[f_1]$ and $[f_2]$:

Definition 4 Given f_1 and $f_2 \in F$ and a minimizing couple h_1 and $h_2 \in W$ (i.e., h_1 and h_2 such that $d(f_1 \circ h_1, f_2 \circ h_2) = d_{\mathcal{F}}([f_1],[f_2])$), $\tilde{f}_1 = f_1 \circ h_1$ and $\tilde{f}_2 = f_2 \circ h_2$ are said to be *mutually-registered representatives* of $[f_1]$ and $[f_2]$.

Note that, even if both f_1 and $\tilde{f}_1 \in [f_1]$, and both f_2 and $\tilde{f}_2 \in [f_2]$, only $d(\tilde{f}_1, \tilde{f}_2) = d_{\mathcal{F}}([f_1],[f_2])$ while $d(f_1, f_2) \geq d_{\mathcal{F}}([f_1],[f_2])$. Moreover, since given a couple of elements f_1 and $f_2 \in F$ there is not a unique minimizing couple h_1 and h_2, there is not a unique couple $\tilde{f}_1$ and $\tilde{f}_2$ of mutually-registered representatives of $[f_1]$ and $[f_2]$. It is worth mentioning two special couples of mutually-registered representatives of $[f_1]$ and $[f_2]$: the one corresponding to $h_1 = \mathbf{1}$ and the other corresponding to $h_2 = \mathbf{1}$. In the former case, $\tilde{f}_1 = f_1$, while in the latter case $\tilde{f}_2 = f_2$.

Definition 5 Given a couple f_1 and f_2, and a couple of mutually-registered representatives $\tilde{f}_1$ and $\tilde{f}_2$ such that $\tilde{f}_1 = f_1$ and $h_1 = \mathbf{1}$, $\tilde{f}_2$ is said to be a f_1-registered

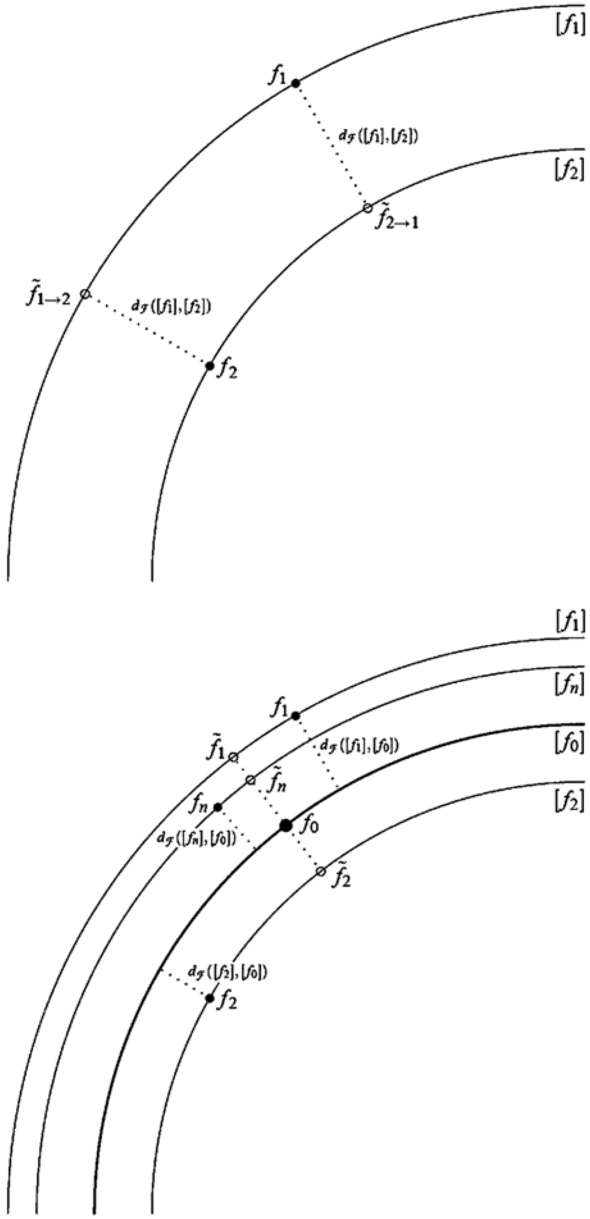

Fig. 1 Naive representation of the mathematical framework introduced: registration of a couple of functions (*top*), and registration of a set of functions with respect to a target function f_0 (*bottom*). Full dots refer to non-registered functions, empty dots to registered functions, and circumference arcs to equivalence classes

representative of $[f_2]$ (or in less formal but more familiar terms, $\tilde{f}_2$ is said to be a registered version of f_2 with respect to f_1). We will refer to it as $\tilde{f}_{2\to 1}$ and to the corresponding warping function as $h_{2\to 1}$. The definitions of $\tilde{f}_{1\to 2}$ and $h_{1\to 2}$ are analogous.

Note that the uniqueness of $\tilde{f}_{2\to 1}$ and $\tilde{f}_{1\to 2}$ cannot be guaranteed, in general. In particular, if $\tilde{f}_{2\to 1}$ and $\tilde{f}_{1\to 2}$ are unique (like in most practical cases), their definition can be made more explicit. Indeed, under the assumption of uniqueness, since the f_1-registered representative of $[f_2]$ is the element $\in [f_2]$ minimizing the distance with f_1, we have that

$$\tilde{f}_{2\to 1} = \arg\min_{f\in[f_2]} d(f, f_1),$$

$$\tilde{f}_{1\to 2} = \arg\min_{f\in[f_1]} d(f, f_2),$$

with $h_{1\to 2} = (h_{2\to 1})^{-1}$.

According to this framework, registering a function $f_1 \in F$ with respect to a function $f_2 \in F$—according to a metric d and a class of warping functions W—simply means replacing f_1 with $\tilde{f}_{1\to 2}$. Just to keep the notation as simple as possible, in the rest of the paper, we will assume, without loss of generality, $\tilde{f}_{2\to 1}$ and $\tilde{f}_{1\to 2}$ to be unique. Note that we are not talking about the uniqueness of the minimizing couple $(\tilde{f}_1, \tilde{f}_2)$, that is instead intrinsically non-unique.

The introduction of a quotient set $\mathcal{F}$ over F (dependent on the choices for d and W) is the key to a clear and not ambiguous definition of *Phase Variability* and *Amplitude Variability*. We are quite sure to come across the heuristic sense of many authors, by defining the phase variability as the one that can occur between functions belonging to the same equivalence class, i.e., the *variability within equivalence classes*; note that if f_1 and f_2 belong to the same equivalence class, we have that $0 = d_{\mathcal{F}}([f_1], [f_2])$. Coherently, the amplitude variability is the variability between functions not belonging to the same equivalence class and not imputable to phase variability, i.e., the *variability between equivalence classes*; we can thus say that the difference between f_1 and f_2 is imputable only to amplitude variability in the case $d_{\mathcal{F}}([f_1], [f_2]) = d(f_1, f_2)$.

Given the fact that $0 \leq d_{\mathcal{F}}([f_1], [f_2]) \leq d(f_1, f_2)$, we can define an amplitude-to-total variability ratio bounded between 0 and 1, useful in practical situations and measuring to what extent phase and amplitude variability contribute to total variability:

$$\alpha^2 = \frac{d^2_{\mathcal{F}}([f_1], [f_2])}{d^2(f_1, f_2)};$$

and then we can simply characterize the two extreme situations as follows:

- presence of phase variability only, when $\alpha^2 = 0$, i.e., $d_{\mathcal{F}}([f_1], [f_2]) = 0$;
- presence of amplitude variability only, when $\alpha^2 = 1$, i.e., $d_{\mathcal{F}}([f_1], [f_2]) = d(f_1, f_2)$.

The two extreme situations can be equivalently characterized as follows:

- presence of phase variability only, when $\tilde{f}_{2\to 1} \equiv f_1$ (and thus also $\tilde{f}_{1\to 2} \equiv f_2$), i.e., f_1 and f_2 can be made identical by means of suitable warping functions;
- presence of amplitude variability only, when $\tilde{f}_{2\to 1} \equiv f_2$ (and thus also $\tilde{f}_{1\to 2} \equiv f_1$), i.e., warping f_1 and f_2 can only make them more distant.

2.3 How many equivalence classes?

Given a set F and a metric d that is invariant with respect to some groups, the quotient set $\mathcal{F}$ depends only on the group W; to emphasize this dependency, in this subsection we will use the notation $\mathcal{F}_W$ to indicate the quotient set associated to the group W, and the notation $\mathcal{P}(\mathcal{F}_W)$ to indicate its powerset.

It is immediate to prove that if W is replaced by a sub-group W', the number of equivalence classes can only increase, i.e.,

$$W' \subset W \Longrightarrow \mathcal{P}(\mathcal{F}_{W'}) \supseteq \mathcal{P}(\mathcal{F}_W).$$

Equivalently, if W is replaced by a sup-group W' (such that d is also W'-invariant), the number of equivalence classes can only decrease, i.e.,

$$W' \supset W \Longrightarrow \mathcal{P}(\mathcal{F}_{W'}) \subseteq \mathcal{P}(\mathcal{F}_W).$$

More generally, within a functional data analysis, the replacement of the group W with the group $W' \subset W$ might cause the partitioning of former equivalence classes (associated to W) into new classes (associated to W'). This kind of variability that occurs between new classes associated to W' being subsets of the same old class associated to W is exactly the variability that according to W' is considered as part of the amplitude variability while according to W is considered as part of phase variability. Of course, the opposite occurs if the group W is replaced by a group $W' \supset W$.

In other words, given d, choosing W is the same as defining phase variability. It is worth mentioning the two extreme situations for the choice of W:

- $W = \{\mathbf{1}\}$: in this case, each element of F is equivalent only to itself, i.e., $F \equiv \mathcal{F}$. We are thus assuming that no phase variability is present in the functional data;
- $W = F$: in this case, all elements of F are equivalent, i.e., only one equivalence class exists coinciding with the whole set F. We are thus assuming that no amplitude variability is present in the functional data.

Note that while the former case can always occur, the latter one can occur only if F is a subgroup of the group of the continuous automorphisms: $\Omega \subseteq \mathcal{R}^p \to \Omega \subseteq \mathcal{R}^p$.

3 Registration of a set of functions in presence of a target function

We have just shown that, under the introduced framework, registering f_2 with respect to f_1 means replacing f_2 with a function $\tilde{f}_{2\to 1} \in [f_2]$ whose distance to f_1 is minimal. In the same framework, it is straightforward to define the registration of a functional data set $\{f_i\}_{i=1,2,\ldots,n}$ with respect to a target function f_0. Indeed registering the set $\{f_i\}_{i=1,2,\ldots,n}$ with respect to f_0 means replacing the set $\{f_i\}_{i=1,2,\ldots,n}$ with the set $\{\tilde{f}_{i\to 0}\}_{i=1,2,\ldots,n}$ (or simply, $\{\tilde{f}_i\}_{i=1,2,\ldots,n}$) whose distances to f_0 are minimal over the relevant equivalence classes:

$$\{f_i\}_{i=1,2,\ldots,n} \longmapsto \left\{\tilde{f}_i = \arg\min_{f\in[f_i]} d(f_0, f)\right\}_{i=1,2,\ldots,n}.$$

In other words, registering the set $\{f_i\}_{i=1,2,\ldots,n}$ with respect to f_0 consists in finding in $[f_1]$, $[f_2]$, $\ldots$, $[f_n]$, the n functions that are the closest to f_0.

Also in this case, we can define an amplitude-to-total variability ratio:

$$\alpha^2 = \frac{\sum_{i=1}^{n} d_{\mathcal{F}}^2([f_i], [f_0])}{\sum_{i=1}^{n} d^2(f_i, f_0)};$$

we can still simply characterize the two extreme situations as follows:

– presence of phase variability only, when $\alpha^2 = 0$;
– presence of amplitude variability only, when $\alpha^2 = 1$.

The two extreme situations can be equivalently characterized as follows:

– presence of phase variability only, when for $i = 1, 2, \ldots, n$: $\tilde{f}_i \equiv f_0$;
– presence of amplitude variability only, when for $i = 1, 2, \ldots, n$: $\tilde{f}_i \equiv f_i$.

4 Registration of a set of functions in absence of a target function

In most practical problems, the focus is on registering a functional data set $\{f_i\}_{i=1,2,\ldots,n}$ with respect to "itself", since a target function f_0 is generally not available. Also in this case, it is still meaningful to talk about registration: roughly speaking, it is straightforward to assert that registering the set $\{f_i\}_{i=1,2,\ldots,n}$ would consist in replacing the set $\{f_i\}_{i=1,2,\ldots,n}$ with a set $\{\tilde{f}_i\}_{i=1,2,\ldots,n} \in [f_1] \times [f_2] \times \cdots \times [f_n]$ of functions that are "as close as possible".

A natural approach to formalize the notion of "as close as possible" is to introduce an auxiliary reference function $\hat{f}_0 \in F$ such that $\{\tilde{f}_i\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0\}$ is the solution of the following minimization problem:

$$\min_{\tilde{f}_i \in [f_i] \wedge \hat{f}_0 \in F} \left(\sum_{i=1}^{n} d^2(\tilde{f}_i, \hat{f}_0) \right). \tag{5}$$

In other words, registering a set $\{f_i\}_{i=1,2,\ldots,n}$ means registering each function of the set with respect to the sample Frechet mean of the registered set. Also in this case, sufficient conditions for the existence of a solution of the minimization problem (5) can be found (see Appendices A and B for details). Note that, also in this case, the solution is never unique (except for $W = \{\mathbf{1}\}$ where the solution might be unique). Indeed, because of the W-invariance of d, if $\{\tilde{f}_i\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0\}$ is a solution of the minimization problem (5) any other set of the form $\{\tilde{f}_i \circ h\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0 \circ h\}$, with $h \in W$, is still a solution.

Also in this case, we can define an amplitude-to-total variability ratio:

$$\alpha^2 = \frac{\sum_{i=1}^{n} d_{\mathcal{F}}^2([f_i], [\hat{f}_0])}{\sum_{i=1}^{n} d^2(f_i, \hat{f}_0)}.$$

Note that in this case (i.e., when a target function f_0 is not present but needs to be estimated) some care is needed to correctly compute the α^2 index:

- First, note that α^2 compares (numerator) the deviations of the registered functions from their Frechet mean with (denominator) the deviations of the original functions from the Frechet mean of the registered functions and not from their Frechet mean as one might expect! This mistake has been made—more or less explicitly—in many works dealing with the registration of functional data, providing, of course, an underestimation of the total variability and consequently an overestimation of the contribution of the amplitude variability to the total variability, even providing hardly interpretable situations in which the amplitude variability seems to be greater than the total one (e.g., Kneip and Ramsay 2008).
- Second, note that the α^2 ratio is not invariant under a joint warping of the solution set $\{\tilde{f}_i\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0\}$ along the same warping function h. Indeed, even if $\{\tilde{f}_i \circ h\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0 \circ h\}$ is still a solution of (5), in the computation of α^2, the numerator does not change while the denominator may change from $\sum_{i=1}^n d^2(f_i, \hat{f}_0)$ to $\sum_{i=1}^n d^2(f_i, \hat{f}_0 \circ h)$. It is natural to assume that, among all possible solutions of the minimization problem (5), the one that is "as close as possible" to the original situation is the natural candidate to be the "right one". Formally, it means that given a solution $\{\tilde{f}_i\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0\}$, the solution $\{\tilde{f}_i \circ h\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0 \circ h\}$ to be used to compute the α^2 ratio is the one minimizing the total variability, i.e., given $[\hat{f}_0]$, the following constraint on $\hat{f}_0$ has to be introduced in the minimization problem (5) in order to identify the correct solution:

$$\sum_{i=1}^n d^2(f_i, \hat{f}_0) = \min_{f \in [\hat{f}_0]} \left(\sum_{i=1}^n d^2(f_i, f) \right). \tag{6}$$

  Thus, the constrained minimization problem can be restated in a simpler way as follows: Given a set $\{f_i\}_{i=1,2,\ldots,n}$, find the reference class $[\hat{f}_0]$ such that

$$[\hat{f}_0] = \arg \min_{[f] \in \mathcal{F}} \left(\sum_{i=1}^n d_{\mathcal{F}}^2\big([f_i], [f]\big) \right)$$

  and take as representatives of the equivalence classes those functions $\{\tilde{f}_i\}_{i=1,2,\ldots,n}$ that are registered with respect to $\hat{f}_0$, that is, that function belonging the reference class $[\hat{f}_0]$ such that its average squared distance to the original functions is minimal.

  Neglecting constraint (6) causes, of course, an overestimation of the total variability and consequently an underestimation of the contribution of the amplitude variability to the total variability. This constraint essentially avoids the drifting apart of the registered functions from the original ones. Similar constraints have been used in the literature for the same purpose. For instance, in both Sangalli et al. (2009) and Kneip and Ramsay (2008) the constraint $\frac{1}{n}\sum_{i=1}^n h_i = \mathbf{1}$ was used, asserting that under this constraint the functions remain unwarped on average. Unfortunately, the latter constraint, even if heuristically equivalent to constraint (6), is not formally equivalent to it. Moreover, it requires the group W to be also a (convex) linear space.

If we take care of the points discussed above, also in this case, we can simply characterize the two extreme situations as follows:

- presence of phase variability only, when $\alpha^2 = 0$;
- presence of amplitude variability only, when $\alpha^2 = 1$.

The two extreme situations can be equivalently characterized as follows:

- presence of phase variability only, when for $i = 1, 2, \ldots, n$: $\tilde{f}_i \equiv \hat{f}_0$;
- presence of amplitude variability only, when for $i = 1, 2, \ldots, n$: $\tilde{f}_i \equiv f_i$.

Solving the minimization problem (5) might, of course, be not trivial. Even if any numerical minimization method can be used to approximate the solution, the proof of Lemma A.2 (see Appendices A and B) suggests all methods belonging to the family of the so-called *Procrustes fitting criteria* to be good candidates to solve this minimization problem. In particular, in the same way of Ramsay and Li (1998), Kneip et al. (2000), and Sangalli et al. (2009), an iterative search of a minimum can be performed, alternating minimization and expectation steps:

Minimization: $\{\tilde{f}_i^{[k+1]}\}_{i=1,2,\ldots,n} = \{\arg\min_{\tilde{f}_i \in [f_i]} (\sum_{i=1}^n d^2(\tilde{f}_i, \hat{f}_0^{[k]}))\}_{i=1,2,\ldots,n}$.
: In these steps, each function of the set $\{f_i\}_{i=1,2,\ldots,n}$ is registered with respect to the Frechet mean of the set $\{\tilde{f}_i^{[k]}\}_{i=1,2,\ldots,n}$;

Expectation: $\hat{f}_0^{[k+1]} = \arg\min_{\hat{f}_0 \in F} (\sum_{i=1}^n d^2(\tilde{f}_i^{[k+1]}, \hat{f}_0))$.
: In these steps, the Frechet mean of the set $\{\tilde{f}_i^{[k+1]}\}_{i=1,2,\ldots,n}$ is computed.

The algorithm can be initialized identifying $\hat{f}_0^{[0]}$ with the Frechet mean of the initial set $\{f_i\}_{i=1,2,\ldots,n}$. Moreover, since $\sum_{i=1}^n d^2(\tilde{f}_i^{[k]}, \hat{f}_0^{[k]})$ can only decrease as k increases and it is lower bounded by 0, the algorithm can be stopped when

$$\sum_{i=1}^n d^2\big(\tilde{f}_i^{[k]}, \hat{f}_0^{[k]}\big) - \sum_{i=1}^n d^2\big(\tilde{f}_i^{[k+1]}, \hat{f}_0^{[k+1]}\big) < \epsilon.$$

Note that, at least theoretically, a small difference between $\sum_{i=1}^n d^2(\tilde{f}_i^{[k]}, \hat{f}_0^{[k]})$ and $\sum_{i=1}^n d^2(\tilde{f}_i^{[k+1]}, \hat{f}_0^{[k+1]})$ can be associated with big distances between $\tilde{f}_i^{[k]}$ and $\tilde{f}_i^{[k+1]}$. This is compatible with the W-invariance of d and the non-uniqueness of the solution of the minimization problem (5). Indeed, in some cases, the algorithm might approach the set of all possible solutions, targeting at each step a different solution. This should not create any concern since after any expectation step the function $\hat{f}_0^{[k+1]}$ has to be replaced by a suitable equivalent function $\hat{f}_0^{[k+1]} \circ h$ satisfying constraint (6).

5 Ancillary variability

In many (or maybe most) situations, d is not a metric but a semi-metric. In this frequent case, the presented theory still holds if F is replaced with $\bar{F}$, where $\bar{F}$ is the quotient set induced by the equivalence relation $d(f_1, f_2) = 0$ defined by the semi-metric d.

It is important to point out that if d is a semi-metric, a further kind of variability is evident: we named it *Ancillary Variability*. Thus, when d is a semi-metric, we can coherently define ancillary, phase and amplitude variability as follows:

– Ancillary variability is the one that can occur between functions belonging to the same equivalence class of $\bar{F}$;
– Phase variability is the one that can occur between equivalence classes of $\bar{F}$ belonging to the same equivalence class of $\mathcal{F}$;
– Amplitude variability is the one that can occur between different equivalence classes of $\mathcal{F}$.

Also in this case we can characterize some extreme situations:

– presence of ancillary variability only, when $d(f_1, f_2) = 0$;
– presence of phase and ancillary variability only, when $d_{\mathcal{F}}([f_1], [f_2]) = 0$;
– presence of amplitude and ancillary variability only, when $d_{\mathcal{F}}([f_1], [f_2]) = d(f_1, f_2)$.

Note that in this case, in the definition of the index α^2, the ancillary variability contributes neither to amplitude nor to total variability. Indeed, according to d, it is actually a non-variability. For this reason, the easiest approach to functional data registration in these cases should be that of setting the analysis, from the very beginning, in terms of elements of $\bar{F}$ and induced metric $d_{\bar{F}}$ rather than in terms of the original elements of F and the original semi-metric d. Moreover, note that the α^2 index is not defined if $d(f_i, f_j) = 0$ for all $i, j = 1, \dots, n$, that is the unrealistic situation of no total variability in the functional data set (e.g., if d was a metric, it would be the case of all identical functions).

6 Examples presented in the literature

The theory hereby developed is able to put in a unique theoretical framework many approaches to functional data registration that have already appeared in the literature. As particular implementations of our method (of course, if suitably re-interpreted), we illustrate two recent papers in which a registration of complex functional data is performed: Sangalli et al. (2009) and Kaziska and Srivastava (2007). Even if of sure interest, the aim of this section is not to compare the performances of different registration procedures compatible with our theory, but rather to show, from a practical point of view, how this theory can be used to clarify the concept of ancillary, phase, and amplitude variability standing behind a given registration procedure, and to quantify their relative importance.

In Sangalli et al. (2009), a registration of 65 three-dimensional curves representing 65 human Internal Carotid Artery centerlines $\subset \mathbb{R}^3$ is performed. In Fig. 2, the first derivatives of these centerlines before and after registration are reported. It is easy to identify in this work the set F, the group W, and the semi-metric d:

$$F = \left\{ f \in C^1\left(\mathbb{R}; \mathbb{R}^3\right) : f(s) \not\equiv c \quad \text{with } c \in \mathbb{R}^3 \right\},$$

$$W = \left\{ h \in C^1(\mathbb{R}; \mathbb{R}) : h(s) = ms + q \quad \text{with } m \in \mathbb{R}^+, q \in \mathbb{R} \right\},$$

$$d(f_1, f_2) = \sqrt{1 - \frac{1}{3} \sum_{k=x,y,z} \frac{\langle f'_{1k},\ f'_{2k} \rangle_{L^2(\Omega)}}{\|f'_{1k}\|_{L^2(\Omega)} \|f'_{2k}\|_{L^2(\Omega)}}}.$$

The corresponding notions of ancillary, phase, and amplitude variability are thus implicitly defined:

– Ancillary variability is the one that can occur between functions f_1 and f_2 that are equal up to an increasing affine transformation of homologous components, i.e.,

$$f_1, f_2 \in F : \exists A_k \in \mathbb{R}^+, \qquad B_k \in \mathbb{R} : f_{1k}(s) = A_k f_{2k}(s) + B_k.$$

– Phase variability is the one that can occur between functions f_1 and f_2 that are equal up to an increasing affine transformation of the abscissa, i.e.,

$$f_1, f_2 \in F : \exists m \in \mathbb{R}^+, \qquad q \in \mathbb{R} : f_1(s) = f_2(ms + q).$$

– Amplitude variability is the one that cannot been removed by the data by means of increasing affine transformations of neither the homologous components nor the abscissa.

Moreover, we are able to compute $\alpha^2 = 33\%$, i.e., the amplitude variability accounts for nearly just $1/3$ of the variability of the 65 Internal Carotid Artery centerlines, supporting the importance of having registered this functional data set before performing any further analysis (Fig. 2).

It is worth noticing that the group W used in Sangalli et al. (2009) is not compact, and thus in general their choice for W and d cannot guarantee the existence of a solution for the associated minimization problem. On the other hand, the lack of compactness of W should not raise any concern in practical situations. Indeed, in real applications, data are usually only slightly misaligned, and therefore a marked local minimum is usually present. So, by simply introducing some "reasonable" constraints on the warping functions (i.e., large enough to gather the local minimum and small enough not to obtain an artificial minimum at the boundary of the constraint) we can re-obtain a meaningful and well posed minimization problem.

Kaziska and Srivastava (2007), in the first part of their work regarding the analysis of human shapes, deal with the registration of simple closed (i.e., periodic) bidimensional curves $\subset \mathbb{R}^2$. Also in this work, it is easy to identify the set F, the group W, and the metric d:

$$F = \mathcal{C},$$
$$W = \mathbb{S}^1 \times \mathcal{D},$$
$$d(f_1, f_2) = d_{\mathcal{C}}(f_1, f_2);$$

where $\mathcal{C}$ (the preshape space in the work) is the set of all continuous 2π-periodic functions mapping $[0, 2\pi]$ into a closed curve $\subset \mathbb{R}^2$ of length 2π and average direction π; $\mathbb{S}^1$ is the group of the translation of the abscissa; $\mathcal{D}$ is the group of the automorphisms $[0, 2\pi] \longmapsto [0, 2\pi]$; the distance $d_{\mathcal{C}}$, even if not clearly stated, appears to be the distance induced by the usual inner product $\langle \cdot, \cdot \rangle_{L^2([0,2\pi])}$, i.e.,

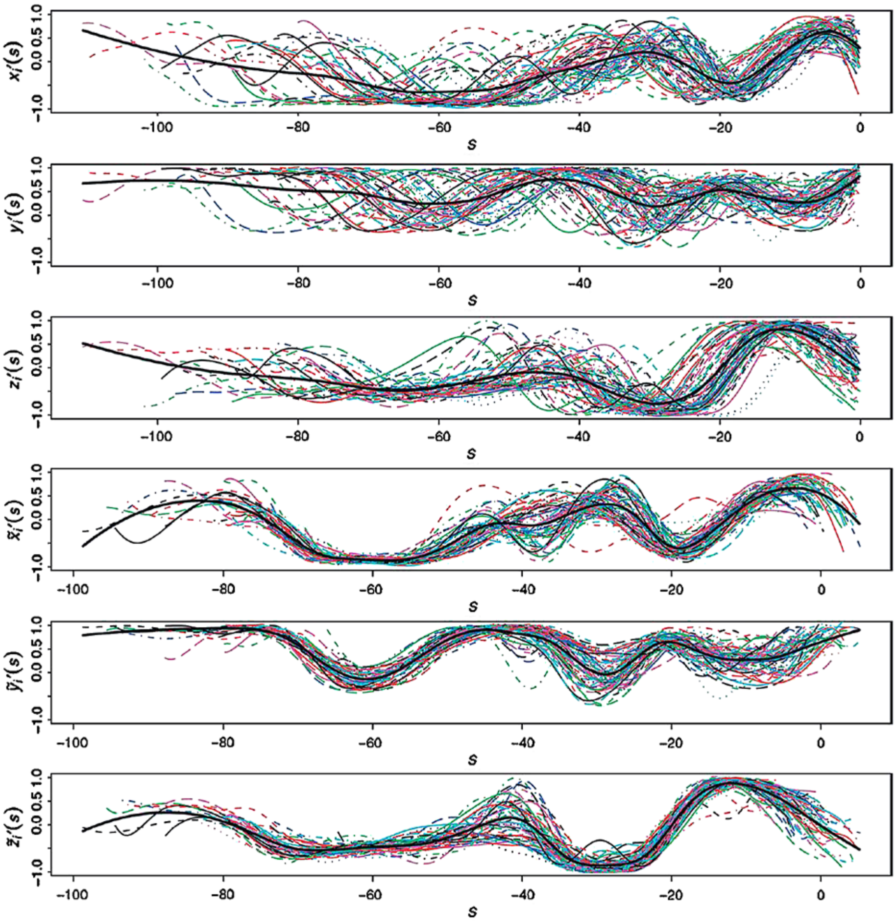

Fig. 2 The 65 first derivatives f'_{ix}, f'_{iy}, and f'_{iz} before registration (*top*) and the 65 first derivatives $\tilde{f}'_{ix}$, $\tilde{f}'_{iy}$, and $\tilde{f}'_{iz}$ after registration (*bottom*). The first derivatives of the estimated Frechet mean $\hat{f}_0$ are reported in black. Courtesy of Sangalli et al. (2009)

$d_{\mathcal{C}}(f_1, f_2) = \|f_1 - f_2\|_{L^2([0,2\pi])}$; for completeness, the quotient set $\mathcal{F}$ is indicated in the work as $\mathcal{S}$ (the shape space).

It is important to mention that, as declared by the authors, the original functions do not belong to $\mathcal{C}$; the original curves are indeed preprocessed—i.e., rotated, translated, and scaled in $\mathbb{R}^2$—such that their average length is 2π and their average direction is π. In this case, the ancillary variability (here the one imputable to rotation, translation, and scaling in $\mathbb{R}^2$) is explicitly removed before the analysis.

The corresponding notions of ancillary, phase, and amplitude variability come straightforward:

– As just pointed out, there is no ancillary variability between functions of $\mathcal{C}$ since this is removed before the analysis;

– Phase variability is the one that can occur between functions representing the same curve in $\mathbb{R}^2$ by means of a different parametrization, i.e., functions that are equal up to a change in the origin of the abscissa and an automorphism of the abscissa;
– Amplitude variability is the one that can occur between functions representing different curves in $\mathbb{R}^2$ after the phase variability has been removed, i.e., after the two abscissas have been matched by means of a joint translation and automorphism $h \in \mathbb{S}^1 \times \mathcal{D}$ minimizing $d_{\mathcal{C}}(f_1, f_2 \circ h)$.

Note that the choice of $d = d_{\mathcal{C}}$ and $W = \mathcal{D}$ does not completely agree with the theory here developed, in the sense that they do not satisfy (d), i.e., $d_{\mathcal{C}}$ is not $(\mathbb{S}^1 \times \mathcal{D})$-invariant (it is actually only $\mathbb{S}^1$-invariant, thanks to the periodicity of functions belonging to $\mathcal{C}$). This means that here two curves can be made arbitrarily more/less close simply by jointly changing the parametrization of the two curves. There are two main consequences due to the non-$\mathcal{D}$-invariance of $d_{\mathcal{C}}$: first, the quotient set $\mathcal{F}$ is not formally defined and thus the phase and amplitude variability are not soundly defined; second, if the registration procedure is performed anyway, one will likely find meaningless and unusable results from a practical point of view. This degeneracy problem (known as *the shape distortion problem*) occurring with the joint use of the group of automorphisms and of the L^2-norm has already been noticed in Brumback and Lindstrom (2004), Ramsay and Silverman (2005), and Kneip and Ramsay (2008).

Kaziska and Srivastava (2007), like Ramsay and Silverman (2005), get out of this degeneracy problem by introducing in the minimization process a penalization term. Indeed, even if in the declared theoretical framework the registration problem is introduced as the minimization over $h \in \mathbb{S}^1 \times \mathcal{D}$ of the functional $\|f_1 - f_2 \circ h\|_{L^2([0,2\pi])}$, in the search for the solution, the functional $\lambda\|f_1 - f_2 \circ h\|^2_{L^2([0,2\pi])} + (1-\lambda)\|h'\|^2_{L^2([0,2\pi])}$ is actually minimized.

In the light of the present work, we are convinced that the necessity of introducing a reasonable penalization term or reasonable constraints to make meaningful the results of a registration procedure is in general an evidence of a mismatch between the phase variability as actually defined by W and the phase variability as thought by the statistician.

For this reason, we think that the correct way to get out of degeneracy problems is not to introduce a penalization term or constraints (we are aware that, even if not theoretically sound, these solutions are, however, practically easy and efficient) but to replace W with a suitable group and to redefine phase and amplitude variability consequently. Of course, this second approach is definitely not trivial and poses new challenging questions for future research.

7 Discussion

The key idea of this work is that the choice for the metric (or semi-metric) d used to perform a functional data analysis should be strictly connected with the nature of the phase variability that the statistician expects to affect the data. In a very wide sense, we state that statisticians, while performing a functional data analysis, should deal with phase variability simply by using suitable semi-metric (i.e., d_W in the paper) able to take in to account the misalignment between curves imputable to phase variability. We also showed that W-invariant metrics (or semi-metrics) (i.e., d in the paper)—with W being the group of warping functions associated to phase variability—provide a natural way to build such semi-metrics (i.e., d_W in the paper).

On the whole, by introducing the semi-metric d_W, we managed to formalize the problem of registration by showing that performing an analysis of a functional data set using a semi-metric d_W is either equivalent to performing an analysis of suitable equivalence classes using the metric $d_{\mathcal{F}}$ (i.e., our theoretical abstraction) and equivalent to performing an analysis of the registered functions using the original metric d (i.e., the usual approach used in applications).

The search for suitable metrics (or semi-metrics) for dealing with functional data is a highly discussed topic also in other areas of FDA (e.g., Ferraty and Vieu 2006). This search is driven by both theoretical and practical reasons:

– from a theoretical point of view, the non-straightforward definition of the notion of probability density function for random functions (Delaigle and Hall 2010) has made the concept of small ball probability (which is, of course, of metric nature) fundamental for many important theoretical results of FDA: for example, the proof of the uniform consistency of kernel estimators in functional data analysis (Ferraty et al. 2010). Thus, in FDA, choosing a suitable metric could make or not make a functional data analysis theoretically sound;
– from a practical point of view, it is universally accepted that a good choice for the metric is the key point for performing a successful functional data analysis. See, for instance, Ferraty and Vieu (2002) where it is shown that the use of different semi-metrics different from the one induced by the L^2-norm can dramatically improve/worsen the predictive power of a non-parametric regression model.

The concept of metric is central also in the recent literature dealing with the problem of defining concepts alternative to the sample mean that are able to capture the main features of the functional data set even if not exactly aligned. For instance, Delaigle and Hall (2010), after having defined a concept of log-density for functional data, defined a consequent concept of modal function (related to the maximal value of the log-density of the first r sample functional principal components); Cuevas et al. (2007), after having compared five different notions of depth for functional data, defined a consequent concept of median function (i.e., the deepest function of the data set) and 0.25-trimmed mean function (i.e., the average of the 75% deepest functions of the data set).

Despite of their apparent distance, methods for registering functional data and methods for computing “meaningful mean functions” (like the ones mentioned above) are very close not only in the methods but also in the aims: indeed, they are both urged by the presence of misalignment between functions. With respect to our particular registration approach, these alternative ways of dealing with phase variability—even if still dependent on the particular choice for the metric—have the advantage of not requiring strong properties of the metric, unlike we do, and thus leaving space for a wide variety of possible metrics. As a drawback, they just provide a phase-variability-corrected “mean function”, leaving uncorrected the original functions. Our approach provides instead both a phase-variability-corrected “mean function" (i.e., the reference function $\hat{f}_0$) and the phase-variability-corrected functions (i.e., the registered functions $\tilde{f}_i$).

In the present paper, we also propose an amplitude-to-total variability ratio α^2 useful for quantifying both the importance of amplitude variability and the effectiveness of a registration procedure: it is zero when there is only phase variability and thus the registration procedure can perfectly align the original functions; and it is one when there is no phase variability and thus the registration procedure leave the original functions unaffected. The index α^2 is here proposed as a purely descriptive tool. It is of paramount interest to investigate in the future the possibility of using it as an inferential tool for testing the absence/presence of phase variability. Identifying the distribution a suitable test statistic derived from α^2 under the null hypothesis of the absence of phase variability—which, of course, will depend on the sample size n—is not straightforward. It will require the introduction of a probabilistic model for both amplitude and phase variability and, probably, also an "anova-inspired" decomposition of total variability in amplitude and phase variability. For this reason, our future research in this topic will focus also on the search for pairs of compact subgroups of the group of the automorphisms and invariant metrics/semi-metrics satisfying this latter requirement.

In our opinion, a clear definition of phase and amplitude variability should be a milestone of the future research activity in functional data analysis (FDA) for at least two reasons:

- first, these two concepts, distinctive of functional data analysis, support the characterization of functional data analysis as a statistical research area itself, not making it just a simple generalization of multivariate analysis;
- second, an obscure definition of phase and amplitude variability will sentence functional data registration to be considered just a preprocessing method for the actual statistical analysis, while—we think—it should be rather considered as a decomposition of the functional data variability in two equally worthy parts.

The present work is just a first attempt to provide a theoretically sound definition of phase and amplitude variability in the light of similar naive concepts used in the literature and in line with present research tendencies of FDA.

Acknowledgements We want to thank Laura Maria Sangalli and Piercesare Secchi for having stimulated this work: the idea of a coherence between the metric and the class of warping functions—which this work starts from—emerged while working with them within the AneuRisk Project. We also want to thank Marco Fuhrman for his comments. Thanks also to the associate editor and to the referees: their comments and their suggested references significantly improved the "statistical flavor" of the work, really pointing out its statistical novelty.

Appendix A: Auxiliary lemmas

Lemma A.1 (Sufficient condition for the existence of the minimizing couple) *If W is compact and $\forall f \in F$ the map $f\circ : W \to F$ is continuous, then $\forall f_1, f_2 \in F$, $\exists \min_{h_1,h_2 \in W} d(f_1 \circ h_1, f_2 \circ h_2)$.*

Lemma A.2 (Sufficient condition for the existence of a solution of the minimization problem (5)) *If W is compact and $\forall f \in F$ the map $f\circ : W \to F$ is continuous, then a solution of the problem (5) exists.*

Appendix B: Proofs

Proof of Lemma A.1 d is continuous since triangular inequality implies $|d(f_1, f_2) - d(f_1, f_3)| \leq d(f_2, f_3)$; the maps $f_1\circ$ and $f_2\circ$ are demanded to be continuous; thus $d(f_1 \circ h_1, f_2 \circ h_2)$ is continuous in h_1 and h_2. Moreover, $d(f_1 \circ h_1, f_2 \circ h_2)$ is lower bounded (≥ 0). Since W is compact, the extreme value theorem ensures the minimum to exist. $\square$

Proof of Lemma A.2 $[f_1], [f_2], \ldots, [f_n]$ are compact sets $\subseteq F$ since W is compact and for $i = 1, 2, \ldots, n$, $f_i\circ : h \in W \longmapsto (f_1 \circ h) \in F$ is continuous. The functional $\sum_{i=1}^{n} d^2(\tilde{f}_i, \hat{f}_0)$ is lower bounded, so $\inf_{\tilde{f}_i \in [f_i] \wedge \hat{f}_0 \in F} (\sum_{i=1}^{n} d^2(\tilde{f}_i, \hat{f}_0)) \geq 0$. The compactness of the set $[f_i]$ guarantees that the inferior limit occurs in correspondence of functions $\tilde{f}_i$ belonging to the set $[f_i]$. Moreover, in correspondence of the inferior limit, we have that $\hat{f}_0$ belongs to the sample Frechet mean set of a set of function belonging to a compact set, i.e., $\bigcup_{i=1,2,\ldots,n} [f_i]$. It is known that the sample Frechet mean set is non-empty and compact. So we can find an element $\hat{f}_0$ belonging to the sample Frechet mean set at which the functional takes value equal to its inferior limit. Thus the inferior limit is also a minimum. $\square$

Proof (Theorem 1: positive semi-definiteness of d_W) From the positive definiteness of d, $f_1 = f_2 \Rightarrow d(f_1, f_2) = 0$; from (2), $d(f_1, f_2) = 0 \Rightarrow d_W(f_1, f_2) = 0$. $\square$

Proof (Theorem 1: symmetry of d_W) The symmetry of d_W descends from the symmetry of d, indeed if $(\bar{h}_1, \bar{h}_2)$ is a minimizing couple for $d(f_1 \circ h_1, f_2 \circ h_2)$, then $(\bar{h}_2, \bar{h}_1)$ is a minimizing couple for $d(f_2 \circ h_2, f_1 \circ h_1)$, providing the same minimum. $\square$

Proof (Theorem 1: triangular inequality for d_W) The triangular inequality for d_W descends from the triangular inequality for d and from the W-invariance of d. Let $(\bar{h}_1, \bar{h}_2)$ and $(\bar{h}_2, \bar{h}_3)$ be minimizing couples for $d(f_1 \circ h_1, f_2 \circ h_2)$ and $d(f_2 \circ h_2, f_3 \circ h_3)$, respectively, i.e.,

$$d_W(f_1, f_2) = d(f_1 \circ \bar{h}_1, f_2 \circ \bar{h}_2), \tag{B.1}$$

$$d_W(f_2, f_3) = d(f_2 \circ \bar{h}_2, f_3 \circ \bar{h}_3). \tag{B.2}$$

As already stressed, because of the W-invariance of d and without loss of generality, $\bar{h}_2$ of the former couple can be fixed equal to $\bar{h}_2$ of the latter couple. The couple $(\bar{h}_1, \bar{h}_3)$ is not in general a minimizing couple for $d(f_1 \circ h_1, f_3 \circ h_3)$, thus,

$$d_W(f_1, f_3) \leq d(f_1 \circ \bar{h}_1, f_3 \circ \bar{h}_3). \tag{B.3}$$

The triangular inequality for d applied to $f_1 \circ \bar{h}_1$, $f_2 \circ \bar{h}_2$, and $f_3 \circ \bar{h}_3$ provides:

$$d(f_1 \circ \bar{h}_1, f_3 \circ \bar{h}_3) \leq d(f_1 \circ \bar{h}_1, f_2 \circ \bar{h}_2) + d(f_2 \circ \bar{h}_2, f_3 \circ \bar{h}_3). \tag{B.4}$$

Finally—by chaining (B.3), (B.4), (B.1), and (B.2)—the triangular inequality for d_W is obtained:

$$\begin{aligned} d_W(f_1, f_3) &\leq d(f_1 \circ \bar{h}_1, f_3 \circ \bar{h}_3) \\ &\leq d(f_1 \circ \bar{h}_1, f_2 \circ \bar{h}_2) + d(f_2 \circ \bar{h}_2, f_3 \circ \bar{h}_3) \\ &= d_W(f_1, f_2) + d_W(f_2, f_3). \end{aligned}$$ □

Proof ((2): *lower bound*) $d(f_1 \circ h_1, f_2 \circ h_2) \geq 0 \; \forall h_1, h_2 \in W \Rightarrow \min_{h_1, h_2 \in W} d(f_1 \circ h_1, f_2 \circ h_2) \geq 0$. □

Proof ((2): *upper bound*) $\min_{h_1, h_2 \in W} d(f_1 \circ h_1, f_2 \circ h_2) \leq d(f_1, f_2)$ since $d(f_1, f_2) = d(f_1 \circ \mathbf{1}, f_2 \circ \mathbf{1})$. □

Proof ((3): $\Rightarrow$) By Theorem 1, $d_W(f_1, f_2) = 0$ implies that there exists a couple (h_1,h_2) such that $d(f_1 \circ h_1, f_2 \circ h_2) = 0$; the positive definiteness of d implies that $d(f_1 \circ h_1, f_2 \circ h_2) = 0 \Rightarrow f_1 \circ h_1 = f_2 \circ h_2$. □

Proof ((3): $\Leftarrow$) The positive definiteness of d implies that $f_1 \circ h_1 = f_2 \circ h_2 \Rightarrow d(f_1 \circ h_1, f_2 \circ h_2) = 0$; since 0 is also a lower bound, this couple is also a minimizing couple, then $d_W(f_1, f_2) = 0$. □

Proof ((4)) Proof is immediate by Theorem 1. □

Proof (Theorem 2: positive definiteness of $d_{\mathcal{F}}$) Let "$\doteq$" be the equivalence relation induced on F by the semi-metric d_W.

Sufficient condition: $[f_1] = [f_2]$ implies that $\forall \bar{f}_1 \in [f_1]$ and $\bar{f}_2 \in [f_2] \Rightarrow \bar{f}_1 \doteq \bar{f}_2$; moreover, $\bar{f}_1 \doteq \bar{f}_2 \Rightarrow d_W(\bar{f}_1, \bar{f}_2) = 0$; but, by definition, $d_W(\bar{f}_1, \bar{f}_2) = d_{\mathcal{F}}([f_1], [f_2])$, and thus $d_{\mathcal{F}}([f_1], [f_2]) = 0$. Necessary condition: $d_{\mathcal{F}}([f_1], [f_2]) = 0$ implies that $d_W(\bar{f}_1, \bar{f}_2) = 0 \; \forall \bar{f}_1 \in [f_1]$ and $\bar{f}_2 \in [f_2]$; thus $\forall \bar{f}_1 \in [f_1]$ and $\bar{f}_2 \in [f_2]$ we have that $\bar{f}_1 \doteq \bar{f}_2$; thus any $\bar{f}_1 \in [f_1]$ is equivalent to any $\bar{f}_2 \in [f_2]$ and vice versa, which means that $[f_1] = [f_2]$. □

Proof (Theorem 2: symmetry of $d_{\mathcal{F}}$) Let us take two elements $\bar{f}_1$ and $\bar{f}_2 \in F$ such that $\bar{f}_1 \in [f_1]$ and $\bar{f}_2 \in [f_2]$; by definition, $d_{\mathcal{F}}([f_1], [f_2]) = d_W(\bar{f}_1, \bar{f}_2)$ and $d_{\mathcal{F}}([f_2], [f_1]) = d_W(\bar{f}_2, \bar{f}_1)$; moreover, the symmetry of d_W ensures $d_W(\bar{f}_1, \bar{f}_2) = d_W(\bar{f}_2, \bar{f}_1)$, and thus the symmetry of $d_{\mathcal{F}}$ is proven. □

Proof (Theorem 2: triangular inequality for $d_{\mathcal{F}}$) Let us take three elements $\bar{f}_1$, $\bar{f}_2$, and $\bar{f}_3 \in F$ such that $\bar{f}_1 \in [f_1]$, $\bar{f}_2 \in [f_2]$, and $\bar{f}_3 \in [f_3]$; by definition, $d_{\mathcal{F}}([f_1], [f_3]) = d_W(\bar{f}_1, \bar{f}_3)$, $d_{\mathcal{F}}([f_1], [f_2]) = d_W(\bar{f}_1, \bar{f}_2)$, and $d_{\mathcal{F}}([f_2], [f_3]) = d_W(\bar{f}_2, \bar{f}_3)$; moreover, the triangular inequality for d_W ensures $d_W(\bar{f}_1, \bar{f}_3) \leq d_W(\bar{f}_1, \bar{f}_2) + d_W(\bar{f}_1, \bar{f}_3)$, and thus the triangular inequality for $d_{\mathcal{F}}$ is proven. □

References

Altman NS, Villarreal JC (2004) Self-modelling regression for longitudinal data with time-invariant covariates. Can J Stat 32(3):251–268

Brumback LC, Lindstrom MJ (2004) Self modeling with flexible, random time transformations. Biometrics 60(2):461–470

Cuevas A, Febrero M, Fraiman R (2007) Robust estimation and classification for functional data via projection-based depth notions. Comput Stat 22(3):481–496

Davidian M, Lin X, Wang JL (2004) Introduction (Emerging issues in longitudinal and functional data analysis). Stat Sin 14(3):613–614. Special issue

Delaigle A, Hall P (2010) Defining probability density for a distribution of random functions. Ann Stat 38(2):1171–1193

Ferraty F (2010) High-dimensional data: a fascinating statistical challenge. J Multivar Anal 101(2):305–306. Special issue

Ferraty F, Romain Y (2011) The Oxford Handbook of Functional Data Analysis. Oxford Handbook. Oxford University Press, Oxford

Ferraty F, Vieu P (2002) The functional nonparametric model and application to spectrometric data. Comput Stat 17(4):545–564

Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Springer Series in Statistics, Theory and Practice. Springer, New York

Ferraty F, Laksaci A, Tadj A, Vieu P (2010) Rate of uniform consistency for nonparametric estimates with functional variables. J Stat Plan Inference 140(2):335–352

González Manteiga W, Vieu P (2007) Statistics for functional data. Comput Stat Data Anal 51(10):4788–4792. Special issue

James GM (2007) Curve alignment by moments. Ann Appl Stat 1(2):480–501

Kaziska D, Srivastava A (2007) Gait-based human recognition by classification of cyclostationary processes on nonlinear shape manifolds. J Am Stat Assoc 102:1114–1124

Ke C, Wang Y (2001) Semiparametric nonlinear mixed-effects models and their applications. J Am Stat Assoc 96(456):1272–1298. With comments and a rejoinder by the authors

Kneip A, Ramsay JO (2008) Combining registration and fitting for functional models. J Am Stat Assoc 103(483):1155–1165

Kneip A, Li X, MacGibbon KB, Ramsay JO (2000) Curve registration by local regression. Can J Stat 28(1):19–29

Lawton WH, EA Sylvestre, Maggio MS (1972) Self modeling nonlinear regression. Technometrics 14(3):513–532

Lindstrom MJ, Bates DM (1990) Nonlinear mixed effects models for repeated measures data. Biometrics 46(3):673–687

Liu X, Müller HG (2004) Functional convex averaging and synchronization for time-warped random curves. J Am Stat Assoc 99(467):687–699

Ramsay JO, Li X (1998) Curve registration. J R Stat Soc, Ser B, Stat Methodol 60(2):351–363

Ramsay JO, Silverman BW (2005) Functional Data Analysis, 2nd edn. Springer, New York

Sangalli LM, Secchi P, Vantini S, Veneziani A (2009) A case study in exploratory functional data analysis: Geometrical features of the internal carotid artery. J Am Stat Assoc 104(485):37–48

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

Tang R, Muller HG (2009) Time-synchronized clustering of gene expression trajectories. Biostatistics 10(1):32–45

Valderrama MJ (2007) An overview to modelling functional data. Comput Stat 22(3):331–334. Special issue