

K-mean Alignment for Curve Clustering

Laura M. Sangalli, Piercesare Secchi, Simone Vantini, Valeria Vitelli

*MOX - Dipartimento di Matematica, Politecnico di Milano,
Piazza Leonardo da Vinci 32, 20133 Milano, Italy*

Abstract

The problem of curve clustering when curves are misaligned is considered. A novel algorithm is described, which jointly clusters and aligns curves. The proposed procedure efficiently decouples amplitude and phase variability; in particular, it is able to detect amplitude clusters while simultaneously disclosing clustering structures in the phase, pointing out features that can neither be captured by simple curve clustering nor by simple curve alignment. The procedure is illustrated via simulation studies and applications to real data.

Key words: Functional data analysis, curve alignment, curve clustering, k-mean algorithm

A problem, often encountered in functional data analysis, is misalignment of the data. A typical example, considered by a number of authors, is given by the children growth curves (see for example Ramsay and Li [18], Sheehy et al. [24,25], Ramsay and Silverman [19], James [9], Telesca and Inoue [29] and Gervini and Gasser [6]). Figure 1 shows the growth curves of 93 children (39 boys and 54 girls) from Berkeley Growth Study data (see Tuddenham and Snyder [30]). Looking at the corresponding growth velocities, also displayed

* Corresponding author: Piercesare Secchi.

Email address: piercesare.secchi@polimi.it.

Postal address: MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy.

Telephone number: +39 02 2399 4592.

Fax number: +39 02 2399 4606.

Email address: laura.sangalli@polimi.it, piercesare.secchi@polimi.it, simone.vantini@polimi.it, valeria.vitelli@mail.polimi.it (Laura M. Sangalli, Piercesare Secchi, Simone Vantini, Valeria Vitelli).

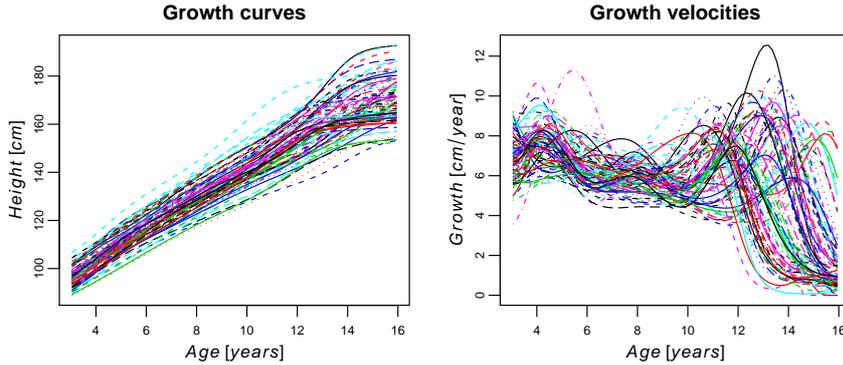


Figure 1. Growth curves of 93 children from Berkeley Growth Study data (left) and corresponding growth velocities (right).

in Figure 1, it is apparent that all curves follow a similar course; this is characterized by a sharp peak of growth velocity between 10 and 16 years, the pubertal spurt, and a minor velocity peak between 2 and 5 years, the mid-spurt. However, different children have their growth spurts at different times, some take more time in their spurts, others less, each child following his/her personal biological clock. Thus, to learn something about the common growth path, it is first necessary to align the biological clocks of the children, eliciting the variability due to the different timings.

Many methods for curve alignment (or curve registration) have been proposed in the literature. For example, Lawton et al. [13] and Altman and Villarreal [1] deal with this problem using self-modelling non-linear regression methods, Lindstrom and Bates [14] develop non-linear mixed-effects models, and Ke and Wang [11] merge the above approaches in the unifying framework of semiparametric non-linear mixed-effects models. A different line of research, advocated by J. O. Ramsay, is followed by Ramsay and Li [18], Ramsay and Silverman [19], James [9], Kaziska and Srivastava [10] and Sangalli et al. [22], who define suitable similarity indexes between curves and thus align the curves, maximizing their similarities by means of a Procrustes procedure.

The present paper is along the latter line of research, and moves forward from the problem of curve alignment, per se, focussing on the more complex problem of curve clustering when curves are misaligned. New questions arise and new answers are needed within this framework. Look for instance at Figure 2. Do the two clusters of curves in case *B* and case *C* represent two sets of curves with distinct shapes, or rather do they reflect a clustering in the phase, that could be eliminated if the curves were suitably aligned? How many set of curves with distinct shape are present in case *D*?

We describe a procedure that is able to efficiently cluster and align in k groups a set of curves. If the number of clusters k is set equal to 1, the algorithm

implements the Procrustes aligning procedure described in [22], whereas, if no alignment is allowed, it implements a functional k -mean clustering of curves (see Heckman and Zamar [8], Tarpey and Kinateder [27], Shimizu and Mizuta [26] and Cuesta-Albertos and Ricardo [4] for different implementations of the k -mean algorithm for curve clustering). For this reason, we will name this procedure *k-mean alignment*.

The paper is organized as follows. In Section 1 we formally describe the problem of curve alignment. In Section 2 we consider the problem of curve clustering when curves are misaligned. In Section 3 we describe the scheme of the k -mean alignment algorithm and in Section 4 we give the technical details for its implementation. Section 5 illustrates the efficiency of the algorithm via a simulation study. Section 6 shows the application to growth curves data, whilst Section 7 is devoted to the application to another real dataset, concerning three-dimensional vascular geometries. Finally, some conclusive considerations are drawn in Section 8. All simulations and analysis of real datasets are performed in R ([17]).

1 Defining phase and amplitude variabilities

The variability among two or more curves can be thought of as having two components: *phase variability* and *amplitude variability*. Heuristically, phase variability is the one that can be eliminated by suitably aligning the curves, and amplitude variability is the one that remains among the curves once they have been aligned. Consider a set \mathcal{C} of (possibly multidimensional) curves $\mathbf{c}(s) : \mathbb{R} \rightarrow \mathbb{R}^d$. Aligning $\mathbf{c}_1 \in \mathcal{C}$ to $\mathbf{c}_2 \in \mathcal{C}$ means finding a warping function $h(s) : \mathbb{R} \rightarrow \mathbb{R}$, of the abscissa parameter s , such that the two curves $\mathbf{c}_1 \circ h$ and \mathbf{c}_2 are the most similar (with $(\mathbf{c} \circ h)(s) := \mathbf{c}(h(s))$). It is thus necessary to specify a similarity index $\rho(\cdot, \cdot) : \mathcal{C} \times \mathcal{C} \rightarrow \mathbb{R}$ that measures the similarity between two curves, and a class W of warping functions h (such that $\mathbf{c} \circ h \in \mathcal{C}$, for all $\mathbf{c} \in \mathcal{C}$ and $h \in W$) indicating the allowed transformations for the abscissa. Aligning \mathbf{c}_1 to \mathbf{c}_2 , according to (ρ, W) , means finding $h^* \in W$ that maximizes $\rho(\mathbf{c}_1 \circ h, \mathbf{c}_2)$. This procedure decouples phase and amplitude variability without loss of information: phase variability is captured by the optimal warping function h^* , whilst amplitude variability is the remaining variability between $\mathbf{c}_1 \circ h^*$ and \mathbf{c}_2 . Note that the choice of the couple (ρ, W) defines what is meant by phase variability and by amplitude variability.

Many similarity indexes for measuring similarity between functions have been considered in the literature on functional data analysis; for a proficient mathematical introduction to the issue see the book by Ferraty and Vieu [5]. In this paper we shall consider the following bounded similarity index between two curves $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$, where $\mathcal{C} = \{\mathbf{c} : \mathbf{c} \in L^2(\mathbb{R}; \mathbb{R}^d), \mathbf{c}' \in L^2(\mathbb{R}; \mathbb{R}^d), \mathbf{c}' \neq \mathbf{0}\}$

(see also Sangalli et al. [22]):

$$\rho(\mathbf{c}_1, \mathbf{c}_2) = \frac{1}{d} \sum_{p=1}^d \frac{\int_{\mathbb{R}} c'_{1p}(s) c'_{2p}(s) ds}{\sqrt{\int_{\mathbb{R}} c'_{1p}(s)^2 ds} \sqrt{\int_{\mathbb{R}} c'_{2p}(s)^2 ds}}, \quad (1)$$

with c_{ip} indicating the p th component of \mathbf{c}_i , $\mathbf{c}_i = (c_{i1}, \dots, c_{id})$; geometrically, $\rho(\mathbf{c}_1, \mathbf{c}_2)$ is the average of the cosines of the angles between the derivatives of homologous components of \mathbf{c}_1 and \mathbf{c}_2 . The two curves are said to be *similar* when the index assumes its maximal value 1; for the similarity index defined in (1), this happens when the two curves are identical except for shifts and dilations of the components:

$$\rho(\mathbf{c}_1, \mathbf{c}_2) = 1 \quad \Leftrightarrow \quad \begin{array}{l} \text{for } p = 1, \dots, d, \exists A_p \in \mathbb{R}^+, \exists B_p \in \mathbb{R} : \\ c_{1p} = A_p c_{2p} + B_p. \end{array}$$

The choice of this similarity index comes along with the following choice for the class W of warping functions of the abscissa:

$$W = \{h : h(s) = ms + q \text{ with } m \in \mathbb{R}^+, q \in \mathbb{R}\}, \quad (2)$$

i.e., the group of strictly increasing affine transformations.

The couple (ρ, W) defined in (1) and (2) satisfies the following properties (a)-(c) that we deem to be minimal requirements for coherence:

- (a) The similarity index ρ is bounded, with maximum value equal to 1, so that two curves \mathbf{c}_1 and \mathbf{c}_2 are similar when $\rho(\mathbf{c}_1, \mathbf{c}_2) = 1$. Moreover, ρ is:
 - reflexive: $\rho(\mathbf{c}, \mathbf{c}) = 1, \forall \mathbf{c} \in \mathcal{C}$;
 - symmetric: $\rho(\mathbf{c}_1, \mathbf{c}_2) = \rho(\mathbf{c}_2, \mathbf{c}_1), \forall \mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$;
 - transitive: $[\rho(\mathbf{c}_1, \mathbf{c}_2) = 1, \rho(\mathbf{c}_2, \mathbf{c}_3) = 1] \Rightarrow \rho(\mathbf{c}_1, \mathbf{c}_3) = 1, \forall \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \in \mathcal{C}$.
- (b) The class of warping functions W is a convex vector space and has a group structure with respect to function composition \circ .
- (c) The index ρ and the class W are consistent in the sense that, if two curves \mathbf{c}_1 and \mathbf{c}_2 are simultaneously warped along the same warping function $h \in W$, their similarity does not change:

$$\rho(\mathbf{c}_1, \mathbf{c}_2) = \rho(\mathbf{c}_1 \circ h, \mathbf{c}_2 \circ h), \quad \forall h \in W.$$

This guarantees that it is not possible to obtain a fictitious increment of the similarity between two curves \mathbf{c}_1 and \mathbf{c}_2 simply by warping them simultaneously to $\mathbf{c}_1 \circ h$ and $\mathbf{c}_2 \circ h$.

It should be stressed that these minimal requirements concern ρ and W jointly. In particular, property (c) highlights the importance of a careful and consistent choice of the couple (ρ, W) . This is mandatory to ensure that the registration

problem is well-posed and that the registration procedure is coherent. Note also that properties (b) and (c), together, imply the following:

(d) For all h_1 and $h_2 \in W$,

$$\rho(\mathbf{c}_1 \circ h_1, \mathbf{c}_2 \circ h_2) = \rho(\mathbf{c}_1 \circ h_1 \circ h_2^{-1}, \mathbf{c}_2) = \rho(\mathbf{c}_1, \mathbf{c}_2 \circ h_2 \circ h_1^{-1}).$$

This means that a change in similarity between \mathbf{c}_1 and \mathbf{c}_2 obtained by warping simultaneously \mathbf{c}_1 and \mathbf{c}_2 , can also be obtained by warping the sole \mathbf{c}_1 or the sole \mathbf{c}_2 . Hence, the similarity that can be obtained by aligning \mathbf{c}_1 to \mathbf{c}_2 is the same as the one that can be obtained by aligning \mathbf{c}_2 to \mathbf{c}_1 .

Finally, the similarity index and class of warping functions defined in (1) and (2) have the following specific feature:

(e) Let W^d be the set of all transformations $\mathbf{r} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that:

$$\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d \mapsto \mathbf{r}(\mathbf{x}) = (r_1(x_1), \dots, r_d(x_d)) \in \mathbb{R}^d,$$

with $r_1, \dots, r_d \in W$. Then, for all \mathbf{r}_1 and $\mathbf{r}_2 \in W^d$,

$$\rho(\mathbf{r}_1(\mathbf{c}_1), \mathbf{r}_2(\mathbf{c}_2)) = \rho(\mathbf{c}_1, \mathbf{c}_2).$$

In words, the similarity index between two curves is unaffected by strictly increasing affine transformations of one or more components of the curves.

2 Curve clustering when curves are misaligned

Consider the problem of clustering and aligning a set of N curves $\{\mathbf{c}_1, \dots, \mathbf{c}_N\}$ with respect to a set of k template curves $\underline{\varphi} = \{\varphi_1, \dots, \varphi_k\}$ (with $\{\mathbf{c}_1, \dots, \mathbf{c}_N\} \subset \mathcal{C}$ and $\underline{\varphi} \subset \mathcal{C}$). For each template curve φ_j in $\underline{\varphi}$, define the domain of attraction

$$\Delta_j(\underline{\varphi}) = \{\mathbf{c} \in \mathcal{C} : \sup_{h \in W} \rho(\varphi_j, \mathbf{c} \circ h) \geq \sup_{h \in W} \rho(\varphi_r, \mathbf{c} \circ h), \forall r \neq j\}, \quad j = 1, \dots, k. \quad (3)$$

Moreover, define the labelling function

$$\lambda(\underline{\varphi}, \mathbf{c}) = \min\{r : \mathbf{c} \in \Delta_r(\underline{\varphi})\}.$$

Note that $\lambda(\underline{\varphi}, \mathbf{c}) = j$ means that the similarity index obtained by aligning \mathbf{c} to φ_j is at least as large as the similarity index obtained by aligning \mathbf{c} to any other template φ_r , with $r \neq j$. Thus $\varphi_{\lambda(\underline{\varphi}, \mathbf{c})}$ indicates a template the curve

\mathbf{c} can be best aligned to, and hence $\lambda(\underline{\varphi}, \mathbf{c})$ a cluster the curve \mathbf{c} should be assigned to.

Case of known templates. If the k templates $\underline{\varphi} = \{\varphi_1, \dots, \varphi_k\}$ were known, then clustering and aligning the set of N curves $\{\mathbf{c}_1, \dots, \mathbf{c}_N\}$ with respect to $\underline{\varphi}$ would simply mean to assign \mathbf{c}_i to the cluster $\lambda(\underline{\varphi}, \mathbf{c}_i)$ and align it to the corresponding template $\varphi_{\lambda(\underline{\varphi}, \mathbf{c}_i)}$, for $i = 1, \dots, N$.

Here we are interested in the more complex case where the k templates are unknown.

Case of unknown templates. Ideally, in order to cluster and align the set of N curves $\{\mathbf{c}_1, \dots, \mathbf{c}_N\}$ with respect to k unknown templates we should first solve the following optimization problem:

- (i) find $\underline{\varphi} = \{\varphi_1, \dots, \varphi_k\} \subset \mathcal{C}$ and $\underline{\mathbf{h}} = \{h_1, \dots, h_N\} \subset W$ such that

$$\frac{1}{N} \sum_{i=1}^N \rho(\varphi_{\lambda(\underline{\varphi}, \mathbf{c}_i)}, \mathbf{c}_i \circ h_i) \geq \frac{1}{N} \sum_{i=1}^N \rho(\psi_{\lambda(\underline{\psi}, \mathbf{c}_i)}, \mathbf{c}_i \circ g_i),$$

for any other set of k templates $\underline{\psi} = \{\psi_1, \dots, \psi_k\} \subset \mathcal{C}$ and any other set of N warping functions $\underline{\mathbf{g}} = \{g_1, \dots, g_N\} \subset W$;

and then, for $i = 1, \dots, N$,

- (ii) assign \mathbf{c}_i to the cluster $\lambda(\underline{\varphi}, \mathbf{c}_i)$, and align it to the corresponding template, $\varphi_{\lambda(\underline{\varphi}, \mathbf{c}_i)}$, using the warping function h_i .

The optimization problem (i) describes a search both for the set of optimal k templates and for the set of optimal N warping functions. Note that the solution $(\underline{\varphi}, \underline{\mathbf{h}})$ to (i) has mean similarity $\frac{1}{N} \sum_{i=1}^N \rho(\varphi_{\lambda(\underline{\varphi}, \mathbf{c}_i)}, \mathbf{c}_i \circ h_i)$ equal to 1 if and only if it is possible to perfectly align and cluster in k groups the set of N curves, i.e., if and only if there exist $\underline{\mathbf{h}}$ and a partition $\mathcal{P} = \{P_1, \dots, P_k\}$ of $\{1, \dots, N\}$ in k elements, such that $\rho(\mathbf{c}_i \circ h_i, \mathbf{c}_j \circ h_j) = 1$ for all i and j belonging to the same element of \mathcal{P} .

It should also be noted that, thanks to property (c) and to definition (3), if $(\underline{\varphi}, \underline{\mathbf{h}})$ is a solution to (i) then also $(\{\varphi_1 \circ g_1, \dots, \varphi_k \circ g_k\}, \{h_1 \circ g_{\lambda(\underline{\varphi}, \mathbf{c}_1)}, \dots, h_N \circ g_{\lambda(\underline{\varphi}, \mathbf{c}_N)}\})$ is a solution to (i), for any $\{g_1, \dots, g_k\} \subset W$; moreover this solution identifies the same clusters (i.e., is associated to the same partition $\mathcal{P} = \{P_1, \dots, P_k\}$ of $\{1, \dots, N\}$).

The optimization problem (i) is not analytically solvable. For this reason, we propose to simultaneously deal with (i) and (ii) via a k -mean alignment algorithm that iteratively alternates *template identification steps* and *assignment and alignment steps*. In the template identification step we estimate the set of

k templates associated to the k clusters identified at the previous assignment and alignment step; in the assignment and alignment step, we align the N curves to the set of the k templates obtained in the previous template identification step, and we assign each of the curves to one of the k clusters. As we shall see, the k -mean alignment algorithm also considers the problem of non-uniqueness of the solution, by targeting a specific solution via a *normalization step*.

3 k -mean alignment algorithm

We describe here the general scheme of the k -mean alignment algorithm. In the following section we will give the technical details for its practical implementation.

Let $\underline{\varphi}_{[q-1]} = \{\varphi_{1[q-1]}, \dots, \varphi_{k[q-1]}\}$ be the set of templates after iteration $q-1$, and $\{\mathbf{c}_{1[q-1]}, \dots, \mathbf{c}_{N[q-1]}\}$ be the N curves aligned and clustered to $\underline{\varphi}_{[q-1]}$. At the q th iteration the algorithm performs the following steps.

Template identification step. For $j = 1, \dots, k$, the template of the j th cluster, $\varphi_{j[q]}$, is estimated using all curves assigned to cluster j at iteration $q-1$, i.e. all curves $\mathbf{c}_{i[q-1]}$ such that $\lambda(\underline{\varphi}_{[q-1]}, \mathbf{c}_{i[q-1]}) = j$. Ideally, the template $\varphi_{j[q]}$ should be estimated as the curve $\varphi \in \mathcal{C}$ that maximizes the total similarity:

$$\sum_{i:\lambda(\underline{\varphi}_{[q-1]}, \mathbf{c}_{i[q-1]})=j} \rho(\varphi, \mathbf{c}_{i[q-1]}). \quad (4)$$

Assignment and alignment step. The set of curves $\{\mathbf{c}_{1[q-1]}, \dots, \mathbf{c}_{N[q-1]}\}$ is clustered and aligned to the set of templates $\underline{\varphi}_{[q]} = \{\varphi_{1[q]}, \dots, \varphi_{k[q]}\}$: for $i = 1, \dots, N$, the i -th curve $\mathbf{c}_{i[q-1]}$ is aligned to $\varphi_{\lambda(\underline{\varphi}_{[q]}, \mathbf{c}_{i[q-1]})}$ and the aligned curve $\tilde{\mathbf{c}}_{i[q]} = \mathbf{c}_{i[q-1]} \circ h_{i[q]}$ is assigned to cluster $\lambda(\underline{\varphi}_{[q]}, \mathbf{c}_{i[q-1]}) \equiv \lambda(\underline{\varphi}_{[q]}, \tilde{\mathbf{c}}_{i[q]})$.

Normalization step. After each assignment and alignment step, we also perform a normalization step. In detail, for $j = 1, \dots, k$, all the $N_j^{[q]}$ curves $\tilde{\mathbf{c}}_{i[q]}$ assigned to cluster j are warped along the warping function $(\bar{h}_{j[q]})^{-1}$, where

$$\bar{h}_{j[q]} = \frac{1}{N_j^{[q]}} \sum_{i:\lambda(\underline{\varphi}_{[q]}, \tilde{\mathbf{c}}_{i[q]})=j} h_{i[q]},$$

thus obtaining $\mathbf{c}_{i[q]} = \tilde{\mathbf{c}}_{i[q]} \circ (\bar{h}_{j[q]})^{-1} = \mathbf{c}_{i[q-1]} \circ h_{i[q]} \circ (\bar{h}_{j[q]})^{-1}$. In this way, at each iteration, the average warping undergone by curves assigned to cluster j

is the identity transformation $h(s) = s$. Indeed:

$$\frac{1}{N_{j^{[q]}}} \sum_{i: \lambda(\underline{\varphi}_{[q]}, \mathbf{c}_{i^{[q]}}) = j} \left(h_{i^{[q]}} \circ (\bar{h}_{j^{[q]}})^{-1} \right) (s) = s, \quad j = 1, \dots, k.$$

The normalization step is thus used to select, among all candidate solutions to the optimization problem, the one that leaves the average locations of the clusters unchanged, thus avoiding the drifting apart of clusters or the global drifting of the overall set of curves. Note that the normalization step preserves the clustering structure chosen in the assignment and alignment step, i.e., $\lambda(\underline{\varphi}_{[q]}, \tilde{\mathbf{c}}_{i^{[q]}}) = \lambda(\underline{\varphi}_{[q]}, \mathbf{c}_{i^{[q]}})$ for all i .

The algorithm is initialized with a set of initial templates $\underline{\varphi}_{[0]} = \{\varphi_{1^{[0]}}, \dots, \varphi_{k^{[0]}}\} \subset \mathcal{C}$, and with $\{\mathbf{c}_{1^{[0]}}, \dots, \mathbf{c}_{N^{[0]}}\} = \{\mathbf{c}_1, \dots, \mathbf{c}_N\}$, and stopped when, in the assignment and alignment step, the increments of the similarity indexes are all lower than a fixed threshold.

4 Technical details for the implementation of the k -mean alignment algorithm

We give the details for the practical implementation of the k -mean alignment algorithm that we used in the simulation studies and applications illustrated in the following sections. This implementation is based on the couple (ρ, W) defined in (1) and (2); by appropriate modifications of the technical details here described, it can be adapted to any choice of (ρ, W) satisfying (a), (b), and (c).

When implementing the k -mean alignment algorithm, it should be first noticed that in most practical situations the N curves to be clustered and aligned are defined on finite intervals and that the curve domains may also differ from curve to curve. For this reason, in the assignment and alignment step we estimate $\rho(\underline{\varphi}_{j^{[q]}}, \mathbf{c}_{i^{[q-1]}} \circ h_{i^{[q]}})$ by replacing the integrals on \mathbb{R} with the corresponding integrals on the intersection of the domains of $\underline{\varphi}_{j^{[q]}}$ and $\mathbf{c}_{i^{[q-1]}} \circ h_{i^{[q]}}$. Moreover, in order to prevent the algorithm from targeting fictitious and non-admissible solutions (with aligned curves having non-intersecting domains), we set restrictions on the warping functions of the class W , by fixing maximal and minimal shifts and dilations allowed within each assignment and alignment step. The solution found by k -mean alignment algorithm is pretty robust to these constraints, provided that they are reasonable, i.e. they are neither so narrow that the optimal warping functions are found in correspondence of the maximal/minimal shifts or dilations, nor so broad that the solution is non-admissible (with aligned curves having non-intersecting domains). In particular, in the simulation studies and applications reported in the follow-

ing sections, the allowed shift in each assignment and alignment step was restricted within $\pm 10\%$ of the minimal length of the original curve domains, and the allowed dilation was restricted within 0.9 and 1.1 (these constraints worked well in all the situations we have considered).

In the template identification step, we estimate the template $\varphi_{j^{[q]}}$ over the domain identified by the union of the domains of the curves assigned to the j th cluster at the previous iteration. We do not attempt to estimate $\varphi_{j^{[q]}}$ as the curve that maximizes the total similarity (4), but we estimate it via Loess (see for example Cleveland and Grosse [3]); this adaptive fitting method keeps the variance of the estimator of the template as constant as possible along the abscissa (see for example Hastie and Tibshirani [7]), given that the domains of the curves are different. In particular, since only first derivatives enter the definition of ρ given in (1), for the couple (ρ, W) here considered it is sufficient to estimate the first derivative $\varphi'_{j^{[q]}}$ of the template $\varphi_{j^{[q]}}$. In the present implementation of the algorithm we thus estimate $\varphi'_{j^{[q]}}$ by means of Loess, with Gaussian kernel and appropriate smoothness parameter, applied to the first derivatives of the curves assigned to the j th cluster at the previous iteration.

We choose the k initial templates $\{\varphi_{1^{[0]}}, \dots, \varphi_{k^{[0]}}\}$ at random among the N original curves, with the only requirement that the k chosen curves, even when aligned, are not similar (i.e., do not have similarity equal to 1). Finally, the algorithm is stopped when in the assignment and alignment step the increments of the similarity indexes are all lower than 0.01 (i.e., 1% of the achievable maximum).

5 Simulation studies

In this section we illustrate the potential of the k -mean alignment algorithm described in Sections 3 and 4, through a four-case simulation study.

5.1 Data generation

Consider the prototype curve:

$$c(s) = 1 * \sin(s) + 1 * \sin\left(\frac{s^2}{2\pi}\right), \quad 0 \leq s \leq 2\pi. \quad (5)$$

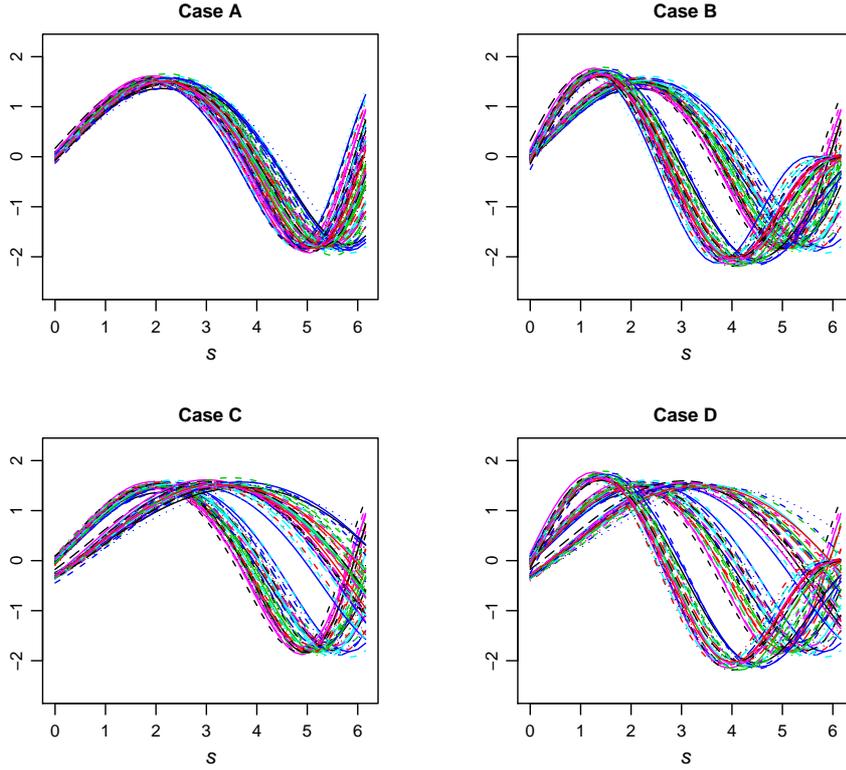


Figure 2. Curves simulated in case *A*, case *B*, case *C*, and case *D*. Different colors and curve styles are used to improve the visualization of the curves, but are neither related to the curve generation nor to any curve clustering.

Case A. We simulate 90 curves from prototype (5), with small errors in amplitude and phase, i.e. for $i = 1, \dots, 90$ we generate

$$c_i^{[A]}(s) = (1 + \varepsilon_{1i}) * \sin(\varepsilon_{3i} + (1 + \varepsilon_{4i})s) + (1 + \varepsilon_{2i}) * \sin\left(\frac{(\varepsilon_{3i} + (1 + \varepsilon_{4i})s)^2}{2\pi}\right),$$

$$0 \leq s \leq 2\pi,$$

where the errors ε are all independent and normally distributed with mean 0 and standard deviation 0.05. The simulated 90 curves are displayed in case *A* of Figure 2.

Case B. The 90 curves displayed in case *B* of Figure 2, $c_1^{[B]}, \dots, c_{90}^{[B]}$, are obtained as follows:

- for $i = 1, \dots, 45$, $c_i^{[B]} = c_i^{[A]}$;
- for $i = 46, \dots, 90$, $c_i^{[B]}$ is obtained from $c_i^{[A]}$ by modifying its amplitude; in particular, instead of considering as prototype the curve in (5) we take as prototype

$$2 * \sin(s) - 1 * \sin\left(\frac{s^2}{2\pi}\right),$$

using the same amplitude and phase errors that were sampled for $c_{46}^{[A]}, \dots, c_{90}^{[A]}$, i.e. for $i = 46, \dots, 90$ we consider

$$c_i^{[B]}(s) = (2 + \varepsilon_{1i}) * \sin(\varepsilon_{3i} + (1 + \varepsilon_{4i})s) + (-1 + \varepsilon_{2i}) * \sin\left(\frac{(\varepsilon_{3i} + (1 + \varepsilon_{4i})s)^2}{2\pi}\right),$$

$$0 \leq s \leq 2\pi.$$

The two clusters of curves visible in case B of Figure 2 thus correspond to two prototype curves with distinct shapes.

Case C. The 90 curves displayed in case C of Figure 2, $c_1^{[C]}, \dots, c_{90}^{[C]}$, are obtained as follows:

- for $i = 1, \dots, 45$, $c_i^{[C]} = c_i^{[A]}$;
- for $i = 46, \dots, 90$, $c_i^{[C]}$ is obtained from $c_i^{[A]}$ by modifying its phase; in particular, we substitute each abscissa s with the modified abscissa

$$-\frac{1}{3} + \frac{3}{4}s,$$

using the same amplitude and phase errors that were sampled for $c_{46}^{[A]}, \dots, c_{90}^{[A]}$, i.e. for $i = 46, \dots, 90$ we consider

$$c_i^{[C]}(s) = (1 + \varepsilon_{1i}) * \sin\left(-\frac{1}{3} + \varepsilon_{3i} + \left(\frac{3}{4} + \varepsilon_{4i}\right)s\right) +$$

$$+ (1 + \varepsilon_{2i}) * \sin\left(\frac{\left(-\frac{1}{3} + \varepsilon_{3i} + \left(\frac{3}{4} + \varepsilon_{4i}\right)s\right)^2}{2\pi}\right), \quad 0 \leq s \leq 2\pi.$$

Hence, the two clusters of curves visible in case C of Figure 2 do not correspond to two prototype curves with distinct shapes, but are instead due to a clustering in the phase; all curves, if suitably aligned, belong to the same amplitude cluster.

Case D. The 90 curves displayed in case D of Figure 2, $c_1^{[D]}, \dots, c_{90}^{[D]}$, are obtained as follows:

- for $i = 1, \dots, 30$, $c_i^{[D]} = c_i^{[A]}$;
- for $i = 31, \dots, 45$, $c_i^{[D]}$ is obtained from $c_i^{[A]}$ via the modification in phase described for case C, and, for $i = 31, \dots, 60$, $c_i^{[D]} = c_i^{[C]}$ (modification in phase);
- for $i = 61, \dots, 90$, $c_i^{[D]} = c_i^{[B]}$ (modification in amplitude).

Thus, among the three clusters of curves visible in case D of Figure 2, only two amplitude clusters are present, and one of the two has associated a further clustering in the phase.

5.2 Data analysis with k -mean alignment

We shall now test the k -mean alignment algorithm on these four simulated cases. We will see that in all cases the algorithm successfully detects the true amplitude clusters, while also disclosing the true clustering structures in the phase. Moreover the procedure efficiently separates amplitude and phase variability.

In order to show the gain obtained by this joint clustering and alignment procedure with respect to a simple clustering procedure where no alignment is allowed, we shall also perform simple clustering of the curves by a k -mean algorithm without alignment, always using the similarity index defined in (1).

Case A. Figure 5 shows the aligned curves and warping functions resulting from 1-mean alignment of A curves. Figure 3 shows the boxplot of the similarity indexes between the original A curves and their mean estimated by Loess (" A , orig"), and the boxplots of the similarity indexes between the k -mean aligned curves and the associated estimated templates, for $k = 1, 2, 3$ (" A , $k=1$ ", " A , $k=2$ " and " A , $k=3$ ", respectively). Figure 4 displays the corresponding means of the similarity indexes (orange circles), together with the means of the similarity indexes that would be obtained by the simple k -mean algorithm without alignment (black circles); note that for $k=1$ without alignment, the plot shows the mean of the similarity indexes between the original A curves and their mean curve estimated by Loess.

Figures 3 and 4 show that the 1-mean alignment procedure leads to an evident increase of the similarity indexes, with respect to the similarities of the original curves, leaving not much scope for further improvement when k is set equal to 2 or 3 in the clustering and alignment procedure. Thus, this procedure correctly detects $k = 1$ amplitude cluster. Moreover, it efficiently elicits the phase variability of the curves, captured by the warping functions displayed in Figure 5, separating it from the the amplitude variability, which is the variability of the aligned curves.

Figure 4 illustrates the gain offered by this joint clustering and aligning procedure with respect to simple clustering without alignment. If phase variability is not elicited, and a simple k -mean algorithm without alignment is performed, this wrongly suggests the existence of multiple clusters. Furthermore, not even with three clusters, k -mean clustering without alignment is able to reach the similarities attained by 1-mean alignment.

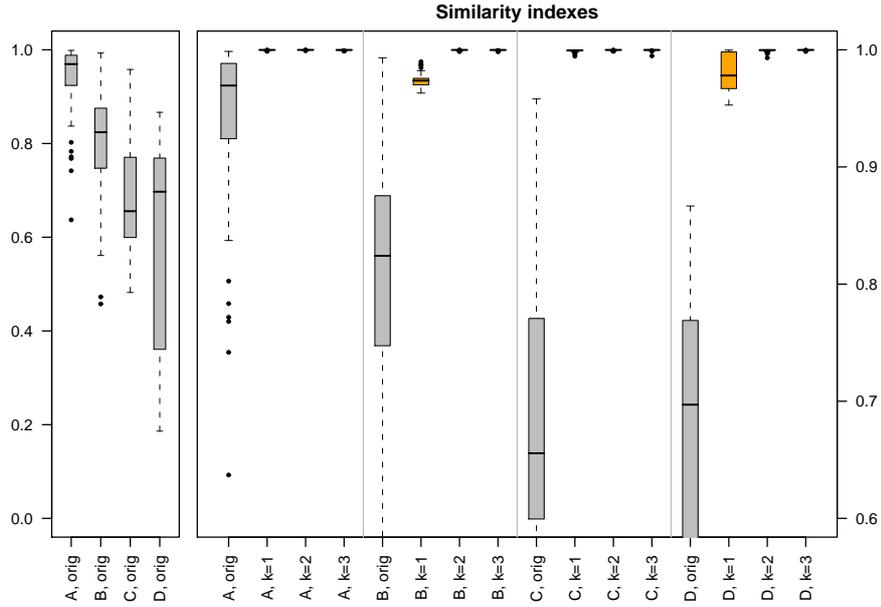


Figure 3. Left: boxplots of similarity indexes between the original curves and their mean curve estimated by Loess (the scale of the plot is from 0 to 1). Right: boxplots of similarity indexes of the original curves (which do not appear in full since the scale of the plot is now from 0.6 to 1) and boxplots of the similarity indexes between the k -mean aligned curves and their estimated templates, for $k = 1, 2, 3$ (cases A , B , C and D , respectively).

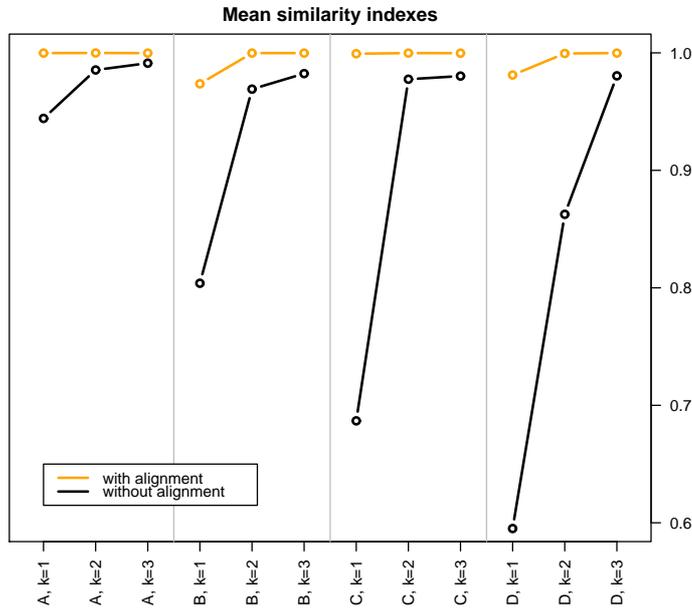


Figure 4. Means of similarity indexes obtained by k -mean alignment, for $k = 1, 2, 3$ (orange circles linked by orange lines) and by k -mean without alignment, for $k = 1, 2, 3$ (black circles linked by black lines); for $k = 1$ without alignment, the plot shows the mean of the similarity indexes between the original curves and their mean curve estimated by Loess.

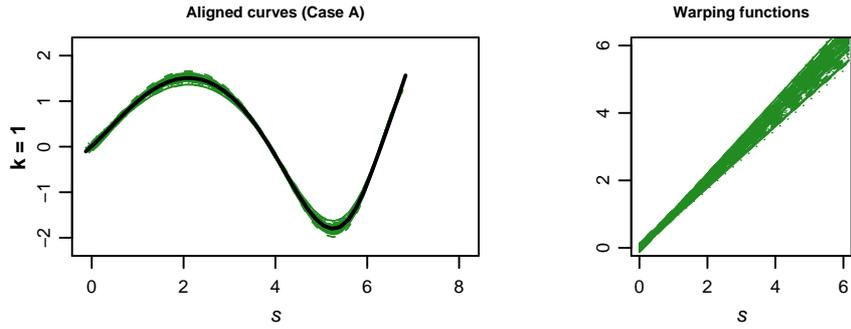


Figure 5. Case A: 1-mean aligned curves with superimposed estimated template (black line) and corresponding warping functions.

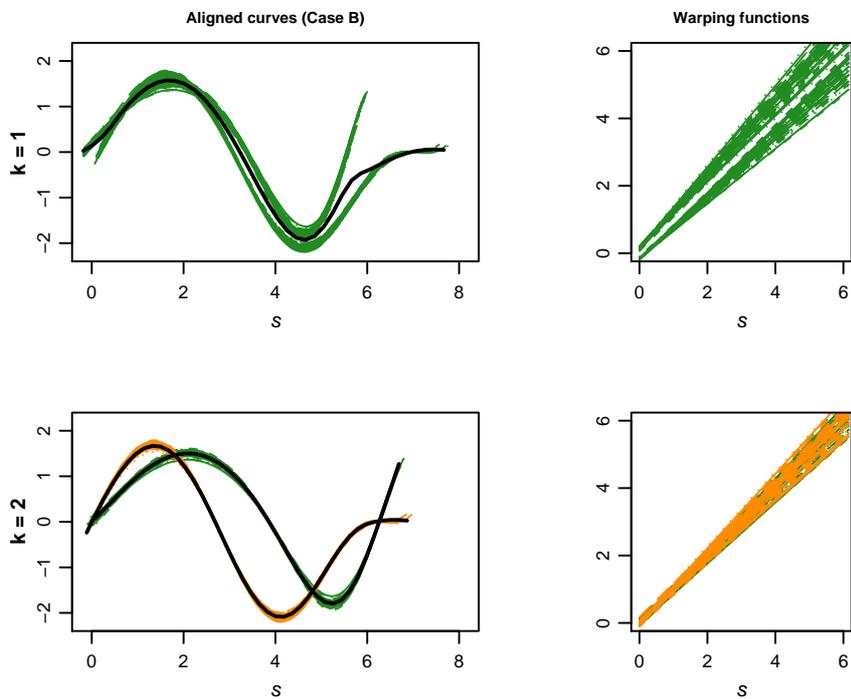


Figure 6. Case B: 1-mean and 2-mean aligned curves with superimposed estimated templates (black lines) and corresponding warping functions. The colors of aligned curves and warping functions depend on the amplitude cluster.

Case B. Figure 6 shows the aligned curves and warping functions resulting from 1-mean alignment and 2-mean alignment of B curves (top and bottom, respectively). The 1-mean alignment fails to give a clear picture of the single amplitude cluster that is looked for, since the aligned curves still appear to be separated in two groups; moreover two clusters are evident in the phase. A better picture is instead provided by the 2-mean alignment, with the 2 amplitude clusters neatly separated and no clustering in phase.

Figure 3 shows the boxplot of the similarity indexes between the original B curves and their mean estimated by Loess (" B , orig"), and the boxplots of the similarity indexes between the k -mean aligned curves and the associated estimated templates, for $k = 1, 2, 3$ (" B , $k=1$ ", " B , $k=2$ " and " B , $k=3$ " respectively). Figure 4 displays the corresponding mean similarities (orange circles). Note that 1-mean alignment leads to a large increase in the similarities, with respect to the similarities of the original curves, but a further considerable gain can be obtained by setting $k=2$ in the clustering and aligning procedure, whereas an eventual choice of $k=3$ is not justified by an additional increase in the similarities. Thus, the procedure correctly identifies 2 amplitude clusters.

We can compare the similarities attained by 2-mean alignment of B curves to the ones attained by 1-mean alignment of A curves. In fact, since half of the B curves coincide with the corresponding A curves, and the other half is obtained from the corresponding A curves by a common modification of their amplitude, one expects that 2-mean alignment of B curves should lead to a comparable result, in terms of similarities, with respect to 1-mean alignment of A curves. This is confirmed by inspection of the boxplots.

Figure 4 also displays the mean of the similarity indexes that would be obtained by the simple k -mean algorithm without alignment (black circles; " B , $k=1$ ", " B , $k=2$ ", " B , $k=3$ "), illustrating the gain offered by k -mean alignment with respect to k -mean without alignment.

Case C. Figure 7 shows the aligned curves and warping functions resulting from 1-mean alignment and 2-mean alignment of C curves (top and bottom, respectively). In this case, the 1-mean alignment already seems to give good results, with the curves nicely aligned in one single group, and two clusters evidenced in phase. Also the 2-mean alignment visually gives good results, with two neatly separated amplitude clusters and no clustering in phase. But, can 2 amplitude clusters really capture the similarity of the 90 curves better than simply 1 cluster? Figure 3 shows the similarity indexes of the original C curves and of the k -mean aligned curves, for $k = 1, 2, 3$ (" C , orig", " C , $k=1$ ", " C , $k=2$ " and " C , $k=3$ " respectively). Figure 4 show the corresponding mean similarities (orange circles). Note that the similarities attained with $k = 1$ amplitude cluster are already very large and the use of $k = 2$ amplitude clusters is not paid off by a further reasonable gain in the similarities. Thus, the clustering and alignment procedure correctly suggests that $k=1$ amplitude cluster is sufficient to capture the similarity of the curves. Hence, the clustering observed in Figure 3, case C, is due to clustering in the phase, and it is captured by the warping functions relative to 1-mean alignment of the curves. Note that when $k=2$, the clustering and alignment procedure uses the unnecessary second amplitude cluster to explain a clustering that is instead present in the phase.

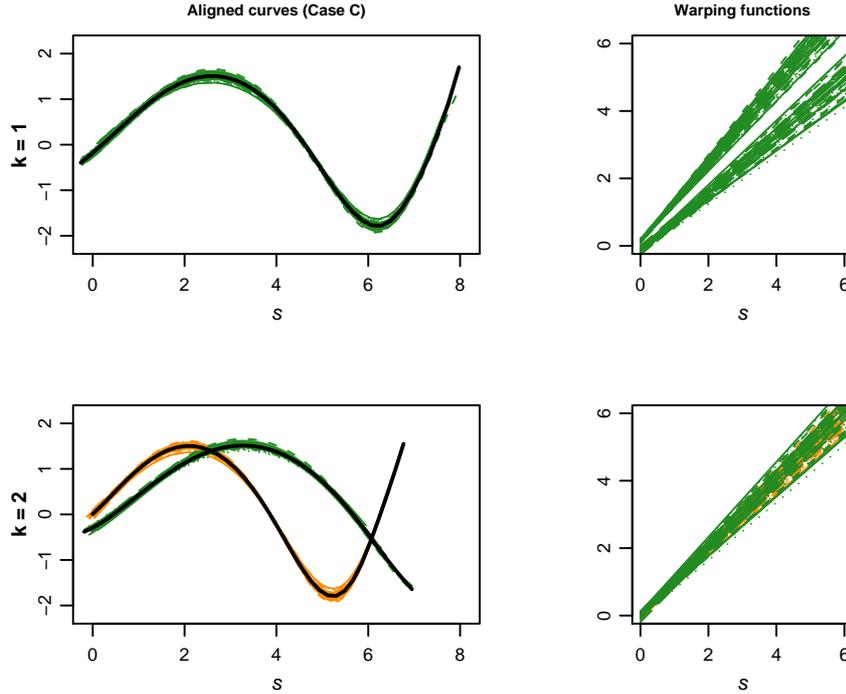


Figure 7. Case C: 1-mean and 2-mean aligned curves with superimposed estimated templates (black lines) and corresponding warping functions. The colors of aligned curves and warping functions depend on the amplitude cluster.

The fact that the clustering of the curves is in the phase, rather than in the amplitude, can not of course be detected by simple k -mean clustering without alignment. Moreover, as shown by the mean similarities displayed in Figure 4 (black circles, " $C, k = 1$ ", " $C, k = 2$ ", " $C, k = 3$ "), not even with three clusters k -mean clustering without alignment can reach the similarities attained by 1-mean alignment.

Case D. Finally, Figure 8 shows the aligned curves and warping functions resulting from 1-mean, 2-mean and 3-mean alignment of D curves (top, center and bottom, respectively). The boxplots of the similarity indexes, shown in Figure 3 (" $D, orig$ ", " $D, k = 1$ ", " $D, k = 2$ " and " $D, k = 3$ "), and the corresponding mean similarities in Figure 4 (orange circles), correctly suggest to use 2 amplitude clusters. The 2-mean alignment procedure efficiently identifies the 2 amplitude clusters and evidences that one of the two clusters (the green one in Figure 8, $k = 2$) has associated a further clustering in the phase. Note that when $k = 1$, the procedure tries to explain the clustering of D curves by imputing it to the phase, where it finds three clusters of warping functions, but the procedure fails to give a clear picture of the single amplitude cluster. Whereas, when $k = 3$, the procedure uses the unnecessary third amplitude cluster to explain a clustering that is instead present in the phase, as noticed

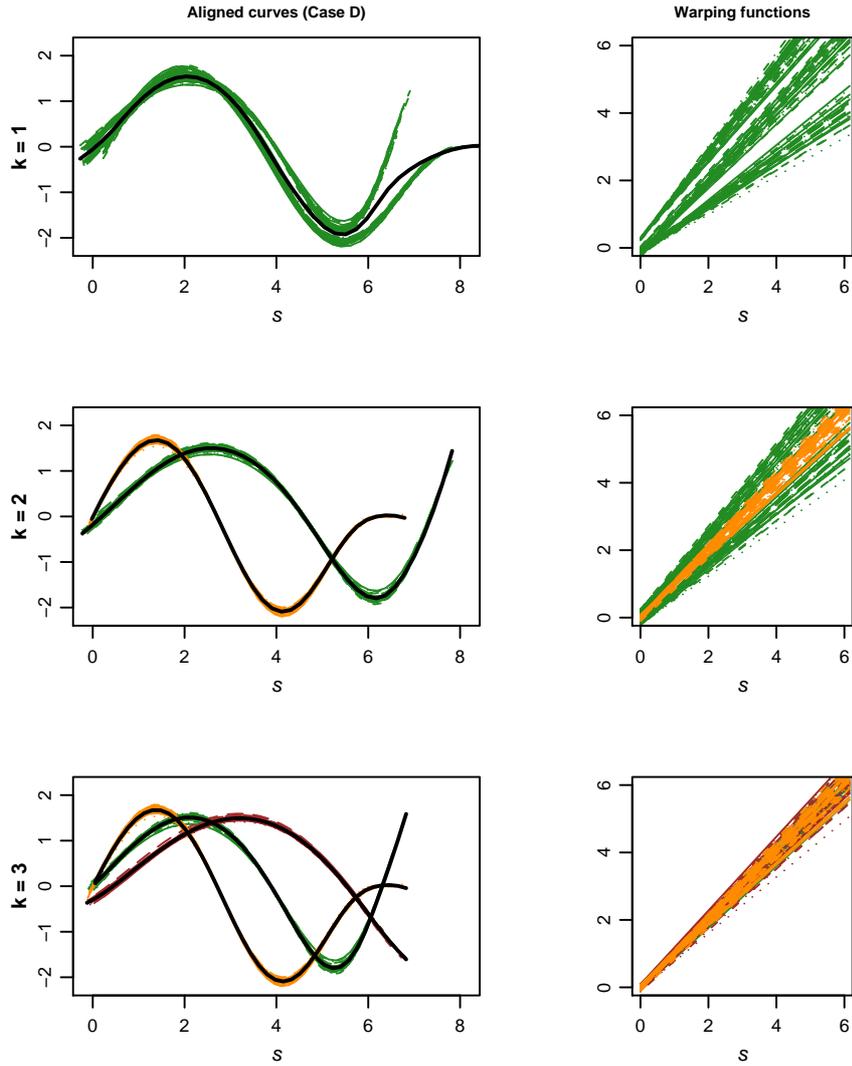


Figure 8. Case D: 1-mean, 2-mean and 3-mean aligned curves with superimposed estimated templates (black lines) and corresponding warping functions. The colors of aligned curves and warping functions depend on the amplitude cluster.

for case C .

Similarly to what has been noticed for case C , the simple k -mean algorithm without alignment can not distinguish clustering in amplitude and clustering in phase.

6 An application to the analysis of growth data

In this section we present the results obtained by applying the k -mean alignment algorithm to the Berkeley Growth Study data. This data set includes the

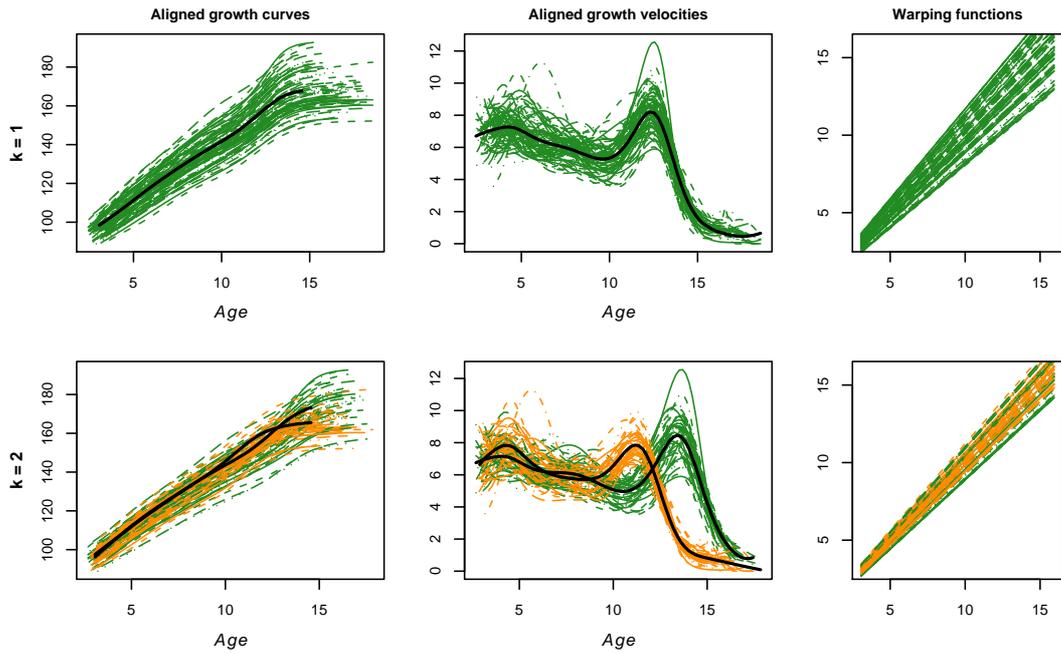


Figure 9. Results of k -mean alignment of growth curves, for $k=1, 2$: aligned growth curves (with superimposed estimated templates, in black) and corresponding growth velocities (with superimposed first derivatives of estimated templates, in black), together with warping functions. The colors of aligned curves and warping functions depend on the amplitude cluster.

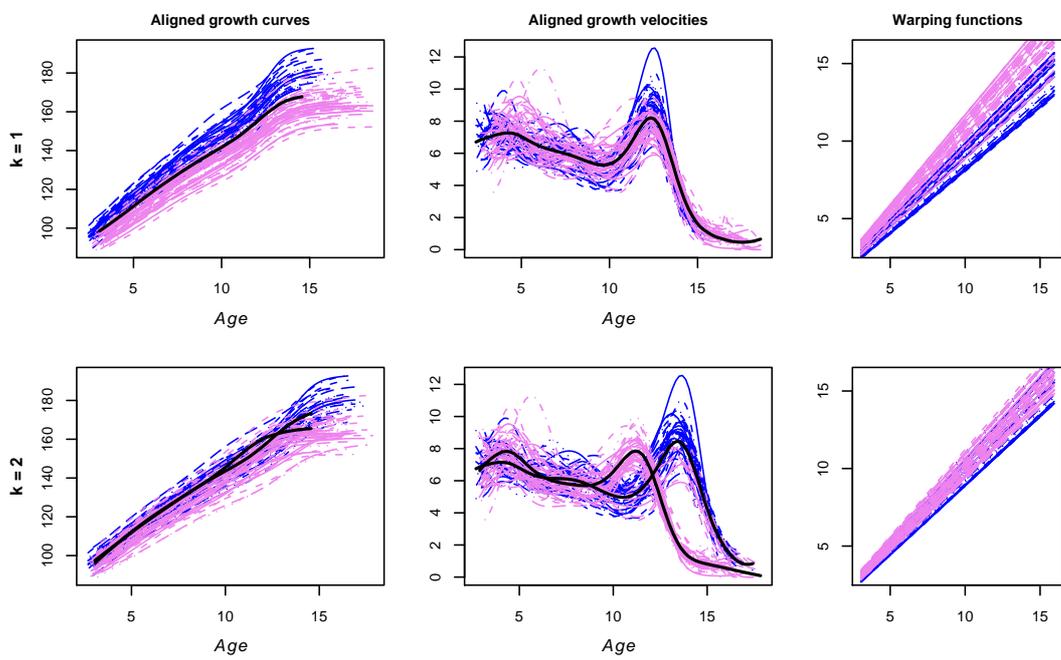


Figure 10. Figure obtained from Figure 9 displaying in blue the growth curves, growth velocities and warping functions of boys and in pink the ones of girls.

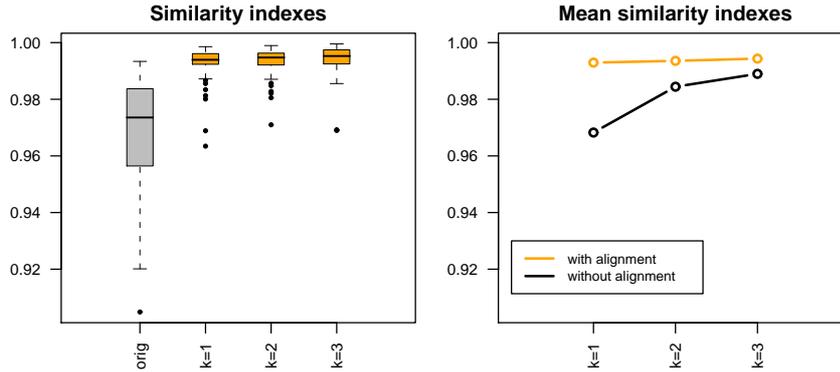


Figure 11. Left: boxplots of similarity indexes for original growth curves and k -mean aligned growth curves, $k = 1, 2, 3$. Right: means of similarity indexes obtained by k -mean alignment and by k -mean without alignment.

heights (in cm) of 93 children, measured quarterly from 1 to 2 years, annually from 2 to 8 years and biannually from 8 to 18 years. We estimate the growth curves by means of monotonic cubic regression splines (see Ramsay and Silverman [19]), implemented using the R function `smooth.monotone` available in the `fda` package [20]. We shall use the similarity index ρ and class of warping functions W defined in (1) and (2). It should be recalled that this similarity index ρ is unaffected by strictly increasing affine transformations of the curves (see (e)) (and thus also by strictly positive dilations of the curve first derivatives). Thus, when we cluster and align the curves, the focus is not on the absolute heights of the children or on their more or less pronounced growths, but rather in their growth patterns. Moreover, using as class of warping function the class W of strictly increasing affine transformations, we are allowing for constant modifications of the running speeds of the children biological clocks.

Figure 9 shows the results obtained by 1-mean and 2-mean alignment of these curves. The boxplots of the similarity indexes and the corresponding mean similarities attained by k -mean alignment, for $k = 1, 2, 3$, shown in Figure 11, suggest that the correct number of amplitude clusters is just 1; in fact, the choice of $k = 2$, in the clustering and alignment procedure, is not payed off by a reasonable further gain in the similarities.

Since, out of the 93 children, 39 are boys and 54 are girls, we might wonder if the analysis points out some differences among them (notice that here we are not performing any supervised classification of boys and girls). Figure 10 is obtained from Figure 9 displaying in blue the growth curves, growth velocities and warping functions of boys and in pink the ones of girls. Let us focus on 1-mean alignment, which is the one suggested by the analysis above. The corresponding warping functions show a pretty neat separation of boys and girls in the phase. This highlights that the biological clocks of boys and girls

run at different speeds; that is to say, boys and girls follow a common growth path, but boys start to grow later, having warping functions with smaller intercepts, and grow more slowly, having warping function with smaller slopes. Another clearly expressed feature emerges from the corresponding 1-mean aligned growth curves: if the biological clocks of the children are registered by 1-mean alignment, then the height of boys stochastically dominates the one of girls for any registered biological age. Finally, boys seem also to have a more pronounced growth, especially during puberty, as highlighted by their more prominent growth velocity peak.

Note that if, for some reasons, 2-mean alignment is preferred, the resulting picture is not so easily interpretable and needs a more complex description; this is due to the fact that the information related to the growth curves and growth velocities, and to the running speeds of the biological clocks of boys and girls, is not so neatly decoupled. Indeed, the main difference between the two amplitude clusters, identified by 2-mean alignment, is relative to the early or late position of the main peak in growth velocity. Since all boys belong to one cluster and most of the girls to the other, one would tend to read this again as a difference in the timings of growth spurts; but this is not a fair interpretation, since it is also necessary to take into account the different running speeds of the personal biological clocks, represented by the warping functions. On the whole, the picture generated by 2-mean alignment is out of focus. This seems analogous to what happened in case C of the simulation study: when the clustering and alignment procedure was asked to identify 2 amplitude clusters, it used the unnecessary second cluster to explain a clustering structure that was instead present in the phase.

7 An application to the analysis of three-dimensional cerebral vascular geometries

In this section, k -mean alignment is used to improve upon the exploratory statistical analyses of the AneuRisk Project¹ dataset (previous analyses are detailed in Sangalli et al. [21,22]). The AneuRisk Project is a joint research program that aims at evaluating the role of vascular geometry and hemodynamics in the pathogenesis of cerebral aneurysms. The data considered in the analyses include the three spatial coordinates (in mm) of 65 Internal Carotid

¹ The project involves MOX Laboratory for Modeling and Scientific Computing (Dip. di Matematica, Politecnico di Milano), Laboratory of Biological Structure Mechanics (Dip. di Ingegneria Strutturale, Politecnico di Milano), Istituto Mario Negri (Ranica), Ospedale Niguarda Ca' Granda (Milano) and Ospedale Maggiore Policlinico (Milano), and is supported by Fondazione Politecnico di Milano and Siemens Medical Solutions Italia.

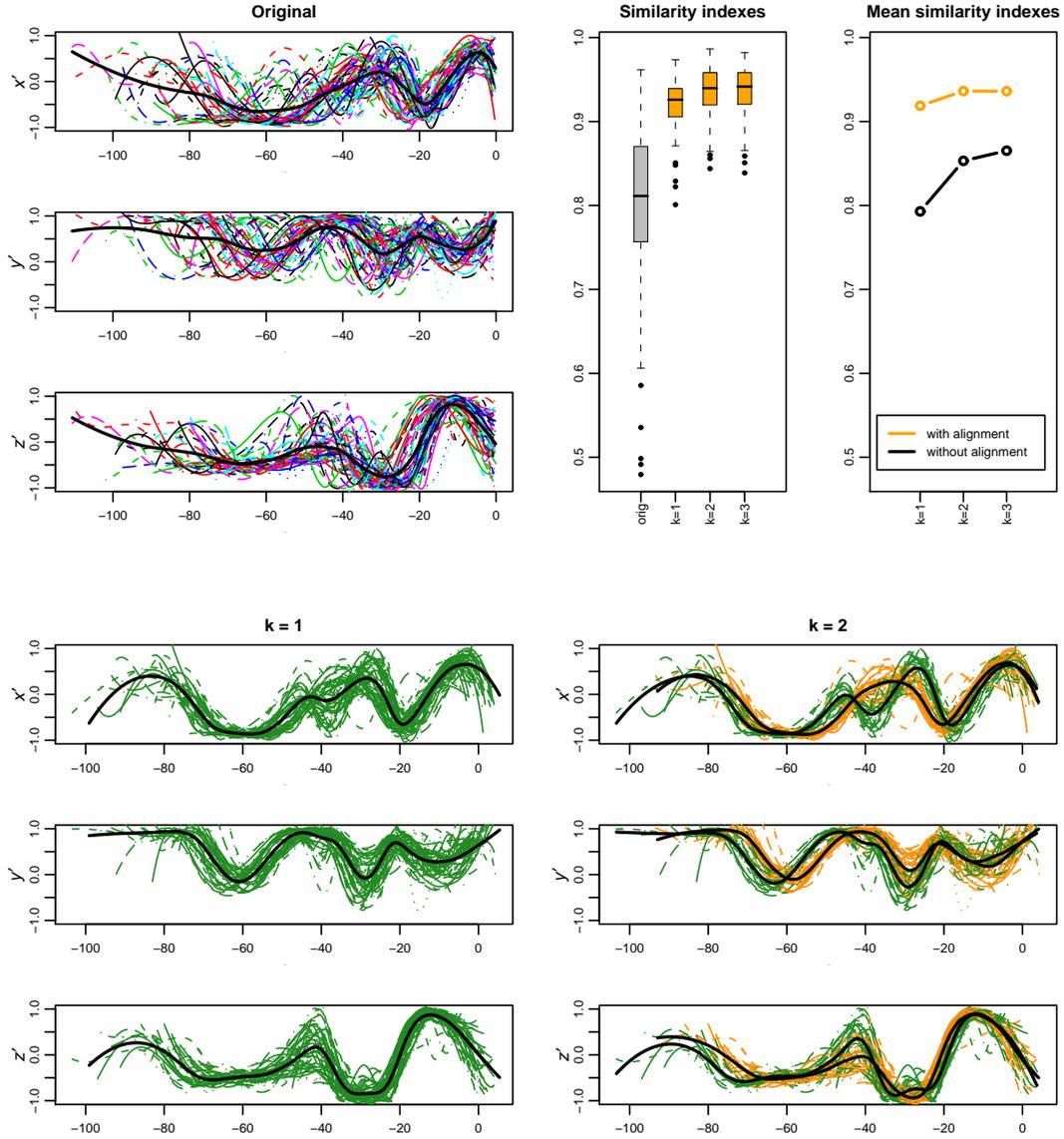


Figure 12. Top left: first derivative x' , y' , z' of the three spatial coordinates of ICA centerlines. Top center: boxplots of similarity indexes for original curves and k -mean aligned curves, $k = 1, 2, 3$. Top right: means of similarity indexes obtained by k -mean alignment and by k -mean without alignment. Bottom: first derivatives of 1-mean and 2-mean aligned curves (left and right respectively).

Artery (ICA) centerlines, measured on a fine grid of points along a curvilinear abscissa (in mm), decreasing from the terminal bifurcation of the ICA towards the heart. Estimates of these three-dimensional curves are obtained by means of three-dimensional free-knot regression splines, as described in Sangalli et al. [23]. The first derivatives, x' , y' , z' , of estimated curves are displayed in Figure 12, top left.

In Sangalli et al. [22] the 65 centerlines are aligned (or 1-mean aligned, as we

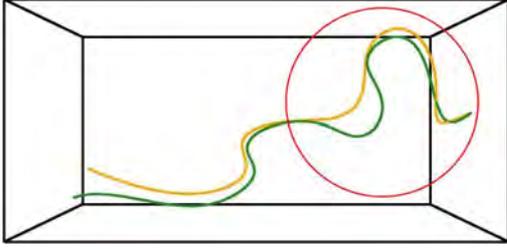


Figure 13. Three-dimensional image of the estimated templates of the 2 amplitude clusters, found by 2-mean alignment of ICA centerlines. The template of the orange cluster is the prototype of an Ω -shaped ICA (one siphon), whereas the one of the green cluster is the prototype of an S -shaped ICA (two siphons).

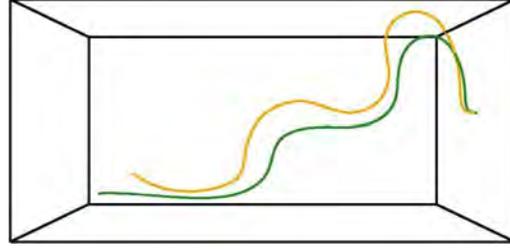


Figure 14. Three-dimensional image of the estimated templates of the 2 clusters found by simple 2-mean clustering without alignment of ICA centerlines. It should be noticed that the two templates appear to have almost the same morphological shape, and seem to differ mainly in their phase.

would now say) according to the couple (ρ, W) defined in (1) and (2). Figure 12, bottom left, displays the first derivatives of the aligned centerlines; the figure shows that the amplitude variability of the 1-mean aligned curves is mostly concentrated in values of the abscissa between -50 and -20. A possible explanation for this fact relates it to the difference in shape for ICA's, along a classification proposed by Kraysenbuehl [12]; this classification discriminates between Γ -shaped, Ω -shaped, and S -shaped ICA's, according to the form of their distal part, which may resemble the letters Γ , Ω or S (in presence of zero, one, or two siphons, respectively).

We want here to explore this explanation by means of k -mean alignment. Looking at the boxplots of the similarity indexes and the corresponding mean similarities attained by k -mean alignment, it can be argued that the use of $k = 2$ amplitude clusters leads in fact to a reasonable further gain in the similarities with respect to $k = 1$, whilst no additional improvement is obtained for $k = 3$. Thus the k -mean alignment algorithm suggests the presence of $k = 2$ amplitude clusters within the analyzed centerlines. Figure 12 compares the first derivatives of 1-mean aligned curves and the first derivatives of 2-mean aligned curves. Indeed, the two amplitude clusters identified by 2-mean alignment could be described as the Ω -shaped ICA's cluster (orange) and S -shaped ICA's cluster (green). This can be better appreciated in Figure 13 where a three-dimensional visualization of the estimated template curves of the two amplitude clusters is displayed. In the end, we can say that only two shape types are present in the AneuRisk dataset; among the 65 ICA's, 35 ICA's belong to the cluster whose template is Ω -shaped, while 30 ICA's belong to the cluster whose template is S -shaped.

Since the shape of the ICA influences the pathogenesis of cerebral aneurysms through its effects on the hemodynamics (as discussed in Piccinelli et al. [16], Sangalli et al. [21,22,23]), this classification of the ICA centerlines could be

helpful in the determination of the risk level of a given patient.

An analysis of these data by simple k -mean clustering without alignment does not instead give interesting insights about this problem. Figure 14 displays the two templates of the clusters obtained by 2-mean clustering without alignment. Notice that the two templates appear to have almost the same morphological shape, and seem to differ mainly in their phase. For these data, the simple k -mean clustering seems to be driven by phase variability, that has not been appropriately elicited, and fails to identify different morphological shapes.

8 Discussion

We described the problem of misaligned curve clustering and proposed an algorithm (k -mean alignment) that jointly clusters and aligns curves with respect to k unknown templates. We tested the novel procedure via both simulated and real cases. In all testing situations, k -mean alignment has been able to efficiently detect true amplitude clusters and also to disclose clustering structures in the phase, pointing out features that can neither be captured by simple k -mean clustering without alignment nor by simple curve alignment without clustering.

Only very recently Tang and Muller [28] have proposed another approach to face the problem of curve clustering when curves are misaligned. The main difference with respect to our approach is that the method proposed by the two authors aims first at aligning the curves, within what could be potential curve clusters, and only afterwards, in a subsequent and separate stage of the analysis, cluster the previously aligned curves, by means of classical clustering methods. Our procedure aims instead at jointly clustering and aligning the curves.

During the revision process of this paper a few other contributions appeared on the problem of clustering and alignment of functional data ([2], [15]), highlighting the growing interest for this topic.

Acknowledgements.

This work has been supported by Ministero dell'Istruzione dell'Università e della Ricerca (research project "Metodi numerici avanzati per il calcolo scientifico" PRIN2006). The dataset analyzed in Section 7 is provided by the Aneurisk Project.

We are grateful to the Associate Editor and two anonymous referees for their

constructive comments.

References

- [1] Altman, N. S. and Villarreal, J. C. (2004), “Self-modelling regression for longitudinal data with time-invariant covariates,” *Canad. J. Statist.*, 32, 251–268.
- [2] Boudaoud, S. and Rix, H. and Meste, O. (2010), “Core Shape modelling of a set of curves,” *Comput. Statist. Data Anal.*, 54, 308–325.
- [3] Cleveland, W. S., Grosse, E., and Shyu, W. M. (1992), *Local regression models. Chapter 8 of Statistical Models in S*, eds Chambers, J. M. and Hastie, T. J., Wadsworth & Brooks, Pacific Grove, California.
- [4] Cuesta-Albertos, J. A. and Fraiman, R. (2007), “Impartial trimmed k -means for functional data,” *Comput. Statist. Data Anal.*, 51, 4864–4877.
- [5] Ferraty, F. and Vieu, P. (2006), “Functional nonparametric statistics in action,” in *The art of semiparametrics*, Physica-Verlag/Springer, Heidelberg, Contrib. Statist., 112–129.
- [6] Gervini, D. and Gasser, T. (2004), “Self-modelling warping functions”, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66, 959–971.
- [7] Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized additive models*, vol. 43 of *Monographs on Statistics and Applied Probability*, London: Chapman and Hall Ltd.
- [8] Heckman, N. E. and Zamar, R. H. (2000), “Comparing the shapes of regression functions,” *Biometrika*, 87, 135–144.
- [9] James, G. M. (2007), “Curve alignment by moments,” *The Annals of Applied Statistics*, 1, 480–501.
- [10] Kaziska, D. and Srivastava, A. (2007), “Gait-Based Human Recognition by Classification of Cyclostationary Processes on Nonlinear Shape Manifolds,” *Journal of the American Statistical Association*, 102, 1114–1128.
- [11] Ke, C. and Wang, Y. (2001), “Semiparametric nonlinear mixed-effects models and their applications,” *J. Amer. Statist. Assoc.*, 96, 1272–1298, with comments and a rejoinder by the authors.
- [12] Krayenbuehl, H., Huber, P., and Yasargil, M. G. (1982), *Krayenbuehl/Yasargil Cerebral Angiography*, Thieme Medical Publishers, 2nd ed.
- [13] Lawton, W. H., Sylvestre, E. A., and Maggio, M. S. (1972), “Self Modeling Nonlinear Regression,” *Technometrics*, 14, 513–532.
- [14] Lindstrom, M. J. and Bates, D. M. (1990), “Nonlinear mixed effects models for repeated measures data,” *Biometrics*, 46, 673–687.

- [15] Liu, X. and Yang, M. C. K. (2009), “Simultaneous curve registration and clustering for functional data,” *Comput. Statist. Data Anal.*, 53, 1361–1376.
- [16] Piccinelli, M., Bacigaluppi, S., Boccardi, E., Ene-Iordache, B., Remuzzi, E., Veneziani, A., and Antiga, L. (2007), “Influence of internal carotid artery geometry on aneurism location and orientation: a computational geometry study,” Available at www.mathcs.emory.edu.
- [17] R Development Core Team (2006), “R: A Language and Environment for Statistical Computing”, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, <http://www.R-project.org>.
- [18] Ramsay, J. O. and Li, X. (1998), “Curve registration,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60, 351–363.
- [19] Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, Springer New York NY, 2nd ed.
- [20] Ramsay, J. O. and Wickham, H. (2007), *fda: Functional Data Analysis*, r package version 1.1.8.
- [21] Sangalli, L. M., Secchi, P., and Vantini, S. (2008), “Explorative Functional data analysis for 3D-geometries of the Inner Carotid Artery,” in *Functional and Operatorial Statistics*, ed. Dabo-Niang, Sophie; Ferraty, F., Springer, Contributions to Statistics, 289–296.
- [22] Sangalli, L. M., Secchi, P., Vantini, S., and Veneziani, A. (2009a), “A Case Study in Exploratory Functional Data Analysis: Geometrical Features of the Internal Carotid Artery,” *J. Amer. Statist. Assoc.*, 104, 37–48.
- [23] — (2009b), “Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines,” *Journal of the Royal Statistical Society Ser. C, Applied Statistics*, 58, Part 3, in press.
- [24] Sheehy, A., Gasser, T., Molinari, L., and Largo, R. H. (2000a), “An analysis of variance of the pubertal and midgrowth spurts for length and width,” *Annals of Human Biology*, 26, 309–331.
- [25] — (2000b), “Contribution of growth phases to adult size,” *Annals of Human Biology*, 27, 281–298.
- [26] Shimizu, N. and Mizuta, M. (2007), “Functional Clustering and Functional Principal Points,” in *Knowledge-Based Intelligent Information and Engineering Systems 11th International Conference, KES 2007, XVII Italian Workshop on Neural Networks, Vietri sul Mare, Italy, September 12-14, 2007. Proceedings, Part III*, eds. Apolloni, B., Howlett, R. J., and Jain, L. C., Berlin Heidelberg: Springer-Verlag, Lecture Notes in Artificial Intelligence 4693, 501–508.
- [27] Tarpey, T. and Kinader, K. K. J. (2003), “Clustering functional data,” *J. Classification*, 20, 93–114.

- [28] Tang, R. and Müller, H.-G. (2009), “Time-synchronized clustering of gene expression trajectories”, *Biostatistics*, 10, 32–45.
- [29] Telesca, T. and Inoue, L. (2008), “Bayesian Hierarchical Curve Registration,” *J. Amer. Statist. Assoc.*, 103, 328–339.
- [30] Tuddenham, R. D. and Snyder, M. M. (1954), “Physical growth of California boys and girls from birth to age 18,” Tech. Rep. 1, University of California Publications in Child Development.