MOX-Report No. 65/2018

Covariance based low-dimensional registration for function-on-function regression

Boschi, T.; Chiaromonte, F.; Secchi, P.; Li, B.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it http://mox.polimi.it

Covariance based low-dimensional registration for function-on-function regression

Tobia Boschi *
Department of Statistics, Penn State University.
Francesca Chiaromonte
Department of Statistics, Penn State University
and EMbeDS, Sant'Anna School of Advanced Studies.
Piercesare Secchi
Department of Mathematics, Politecnico di Milano,
and Center for Analysis, Decisions and Society, Human Technopole,
Milano, Italy.
Bing Li
Department of Statistics, Penn State University.

December 17, 2018

Abstract

We propose a new low-dimensional registration procedure that exploits the relationship between response and predictor in a function-on-function regression. In this context, Functional Covariance Components (FCC) provide a flexible and powerful

*We are grateful to the AneuRisk project – led by Alessandro Veneziani – for sharing the data, and to Laura Sangalli, Simone Vantini, and Alice Parodi for many useful discussions. Tobia Boschi and Francesca Chiaromonte were partially funded by NSF grant DMS-1407639 and the Huck Institutes of the Life Science of Penn State.

tool to represent the data in a low-dimensional space, capturing the most meaningful modes of dependency between the two set of curves. Based on this reduced representation, our procedure aligns simultaneously the two sets of curves, in a way that optimizes the subsequent regression analysis. To implement our procedure, we use both the Continuous Registration algorithm (CR) and a novel parallel algorithm coded in R. We then compare it to other common registration approaches via simulations and an application to the AneuRisk data.

Keywords: Function-on-Function regression; Functional Data Registration; Covariance Operator.

1 Introduction

Functional Data Analysis (FDA) has become a very active area of Statistics, offering sophisticated tools and methods applicable to scientific fields ranging from the geosciences to the social and biomedical sciences. A critical issue in FDA is *data registration*; that is, the problem of separating phase (horizontal) and amplitude (vertical) variation in a statistically meaningful way. Among the most common approaches to data registration are those based on warping functions. These are monotonic functions that transform the domain of the data, making them as similar as possible to an overall template function. Kneip and Ramsay (2008) and Wagner and Kneip (2018) recently introduced procedures that align the data using appropriate representations in reduced functional spaces. The standard choice to represent the data in a low-dimensional space is through a Functional Principal Components (FPC) basis.

In a function-on-function regression context, we propose a different choice of basis for low-dimensional registration; specifically, one that exploits the relationship between response and predictor – preserving information relevant to the regression. Functional Covariance Components (FCC) naturally provide such a basis: our novel procedure registers the two sets of curves (response and predictor) simultaneously based on the most significant modes of covariation between them.

To implement FCC registration, we develop an efficient new parallel algorithm in R, referred to as H1 throughout the article, which minimizes the H1 distance between each curve and its projection on the FCC basis. We benchmark our new algorithm against an algorithm proposed by Ramsay and Silverman (2005) and available in the fda R package, which we refer to as RS. For each curve, RS minimizes the smallest eigenvalue of the cross product matrix between the curve and a template – usually the mean curve. In our

adaptation, the template is not the mean, and it is not the same for all curves. Instead, for each curve, we use as template its projection on the FCC basis.

We also benchmark FCC registration, implemented with both H1 and RS, against commonly used methods such as mean-template registration (Ramsay and Silverman, 2005), elastic shape registration (Tucker et al., 2013) and principal components registration (Kneip and Ramsay, 2008). Our simulations show that FCC registration, especially with the novel H1 algorithm, performs on par with other approaches in terms of alignment, provides an effective low-dimensional representation of the data, and in fact improves performance in terms of regression analysis. Also our application to the AneuRisk data demonstrates the benefits of FCC registration in terms of regression results.

The remainder of the article is organized as follows. In Section 2 we give background on the low-dimensional registration problem. In Section 3 we introduce our FCC registration – showing how covariance components can guide the simultaneous alignment of two sets of curves. In Section 4 we describe the implementation of FCC registration based on RS and on our novel H1 algorithm. In Section 5 we report simulation results and an application to AneuRisk data on 52 patients, where we study the relationship between wall shear stress on the carotid walls and some of its geometrical features. In Section 6 we provide final remarks and discuss further developments and extensions of our work.

2 Background

Throughout the article, we consider a sample comprising two sets of n (paired) curves, indicated as $x(t) = x_1(t), \ldots, x_n(t)$ and $y(t) = y_1(t), \ldots, y_n(t)$. The domain is assumed to be a common interval $[a, b]$, though all our developments hold when domains differ for x and y. We also assume $x_i, y_i \in H^2([a, b])$; that is, we require curves to be twice differentiable

with both first and second derivatives belonging to $L^2([a,b])$.

In practice, to construct each curve x_i (or y_i) starting from any number, say S_i, of raw observations $x_{i1}^{(R)}, \ldots x_{iS_i}^{(R)}$, we use a linear combination of J B-spline functions of order b, ϕ_j^b, with $J-b+2$ equidistant knots. We write

$$x_i(t) = \sum_{j=1}^{J} c_{ij}\phi_j^b(t) \tag{1}$$

where the coefficients c_{ij} are chosen to minimize $\sum_{s=1}^{S_i}\left(x_{is}^{(R)} - x_i(s)\right)^2 + \lambda\int_a^b\left(\frac{d^2}{dt^2}x_i(t)\right)^2 dt$ (the first term is a sum of least squares distances, the integral is a smoothing penalization term, and λ is a smoothing parameter). Note that, for any $b > 3$, we do indeed obtain curves in H^2 – thus meeting our requirement. Of course, different basis systems or entirely different methods, e.g. free-knot regression splines as in Sangalli et al. (2009b), could be used to construct the curves. However, a basis system is necessary to define and implement our registration procedure.

Before proceeding, we introduce further notation similar to Horváth and Kokoszka (2012). The covariance function between x and y, is defined as $\sigma_{xy}(t,s) = \mathsf{E}\big[\big(x(s) - \mathsf{E}x(s)\big)\big(y(t) - \mathsf{E}y(t)\big)\big]$, and the variance functions of x and y correspond to σ_{xx} and σ_{yy}, respectively. The covariance operator between x and y is defined as $(\Sigma_{xy}v)(t) = \int_a^b \sigma_{xy}(t,s)v(s)ds$ for $v \in L^2([a,b])$. Similarly we can define Σ_{xx}, the variance operator of x, and Σ_{yy}, the variance operator of y. In terms of sample counterparts, the covariance function is $\hat{\sigma}_{xy}(s,t) = n^{-1}\sum_{i=1}^{n}\big(x_i(s) - \bar{x}(s)\big)\big(y_i(t) - \bar{y}(t)\big)$ and the covariance operator is $(\hat{\Sigma}_{xy}v)(t) = \int_a^b \hat{\sigma}_{xy}(t,s)v(s)ds$, where $\bar{x}(t) = n^{-1}\sum_{i=1}^{n} x_i(t)$ and $\bar{y}(t) = n^{-1}\sum_{i=1}^{n} y_i(t)$ are sample means.

2.1 Registration with warping functions

Functional data exhibit both *amplitude* and *phase* variation. The first pertains to the size of a curve's features, e.g. peaks and valleys, ignoring their locations, i.e. their position in the domain. The second pertains to the location of the features, ignoring their sizes. The distinction is critical because if the location of features varies among curves, they should not be compared at the same positions in the domain; for instance, curves may share a common pattern, but this pattern may be misaligned along the domain. In many applications ignoring phase variability can lead to an inability to capture important structure in the data, and to inefficiency in modeling. The problem of separating amplitude and phase variation in a statistically meaningful way is called *registration*.

Many registration techniques have been developed over the years. Here, we focus on registration based on *warping functions*. Following Ramsay and Silverman (2005), we define a warping function h as an element of the space $\mathcal{H} \subset H^2[(a,b)]$ of all continuous, strictly increasing functions such that $h(a) = a$ and $h(b) = b$. In the registration procedure, each curve x_i is associated to a specific warping function h_i which provides a (possibly non-linear) transformation of the domain capturing the phase variation of the curve. More specifically, h_i is chosen so that the registered function $x_i(h_i(t))$ is as similar as possible to a given template, with a definition of similarity that can vary among different algorithms. Since monotone transformations do not change shape features, the registered functions $x_1(h_1(t)), ..., x_n(h_n(t))$ will possess the same sequences of peaks and valleys as the original functions and, ideally, exhibit only amplitude variation. Note that the condition $h(a) = a$ and $h(b) = b$ excludes simple horizontal translations – imposing that all registered curves have common start and end points. This is fairly natural in many applications, see Ramsay and Silverman (2002), but this requirement can be modified in specific situations.

Among all possible warping representations, we adopt the one used by Kneip and Ram-

say (2008) and consider functions of the form

$$h(t; w_i) = a + (b-a)\frac{\int_a^t \mathrm{e}^{w_i(u)}du}{\int_a^b \mathrm{e}^{w_i(u)}du}, \tag{2}$$

where $w_i \in H^2([a,b])$. It is immediate to note that for any scalar $c \in \mathbb{R}$ the functions w and $w + c$ lead to the same warping h. Thus, following Kneip and Ramsay (2008), we only consider *standardized* warping functions by imposing that $\int_a^b w_i(t)dt = 0 \ \forall i$ and $n^{-1}\sum_{i=1}^n w_i(u) = 0 \ \forall u \in [a,b]$.

2.2 Registration to low dimensional spaces

Many FDA methods are based on identifying low dimensional linear subspaces of functions that are able to provide accurate approximations of the observed functional data. For this reason, some recent approaches such as Kneip and Ramsay (2008) and Wagner and Kneip (2018) consider registration not simply as an alignment problem, but also as a dimension reduction one. Warping functions are selected in such a way that the resulting registered curves span a low dimensional linear subspace. Hence, if we have n curves $x_1(t), \ldots, x_n(t) \in H^2([a,b])$ the registration problem can be described as in Wagner and Kneip (2018); namely, identifying a set of warping functions $h_1(t), \ldots, h_n(t)$, a linear subspace $\mathcal{L}_K \subset H^2([a,b])$ and a dimension K for the latter as to have

$$x_i(h_i(t)) = \sum_{j=1}^{K} a_{ij}\gamma_j(t), \qquad i = 1, \ldots, n \tag{3}$$

for a suitable set of continuous basis functions $\gamma_1, \ldots, \gamma_K$ with $span\{\gamma_1, \ldots, \gamma_K\} = \mathcal{L}_K$ and for some coefficients $a_{i1}, \ldots, a_{iK} \in \mathbb{R}$. Of course many choices are possible for the basis functions. Kneip and Ramsay (2008) and Wagner and Kneip (2018) use as basis the Functional Principal Components (FPC) identified by the eigendecomposition of the

variance operator Σ_{xx}. The basic idea behind this choice is to align the curves using as template a basis expansion that describes the leading modes of variation in the data. Note that, as pointed out in Wagner and Kneip (2018), the registration problem described in (3) poses an identifiability issue. Specifically, uniqueness of the $\mathcal{L}_K$ space is not guaranteed.

3 FCC registration

We now focus on a function-on-function regression context. Assume again to have two sets of paired curves $x(t) = x_1(t), \ldots, x_n(t)$ and $y(t) = y_1(t), \ldots, y_n(t)$, with $x_i, y_i \in H^2([a,b])$ $\forall i$. Following Horváth and Kokoszka (2012), the function-on-function regression model with a single predictor is defined as

$$y_i(s) = \alpha(s) + \int_a^b x_i(s)\beta(s,t)dt + \epsilon_i(s) \qquad i = 1, \ldots, n, \tag{4}$$

where y is the functional response, x the functional predictor, α the functional intercept in $H^2([a,b])$, $\beta(s,t)$ the regression coefficient surface, and the ϵ's are i.i.d. Gaussian random elements in $H^2([a,b])$, independent of the x's, with mean function 0 and common variance operator. Again, without loss of generality, we assume x and y to have the same domain. In this framework, we develop a registration procedure that, exploiting the relationship between x and y, is able to optimize predictive performance in the function-on-function regression. This registration driven by regression can be compared to the one driven by clustering introduced in Sangalli et al. (2010): the aim is not only the alignment itself, but also the performance of a specific statistical analysis. To achieve our goal, we propose FCC registration. Unlike FPC registration, which utilizes the leading modes of variation within one set of curves, FCC registration utilizes the leading modes of covariation between two sets of response and predictor curves. In the remainder of this Section, we describe our FCC

registration approach, and show how the covariance operator can guide a low-dimensional registration procedure.

3.1 Functional Covariance Components

To capture the linear associations between two sets of curves, we identify the pair of linear combinations of the curves having the largest covariance, then (orthogonally) the pair with the second largest covariance, etc. In the following, we use the terms *covariance functions* or *Functional Covariance Components* (FCC) to refer to these pairs. These are similar in spirit to the pairs produced by the popular Functional Canonical Correlation Analysis, first introduced by Leurgans et al. (1993), which maximizes the correlation instead of the covariance. Notably, whether the components are produced on the basis of covariance or correlation, they allow a low-dimensional representation of an infinite-dimensional relationship between two sets of curves. But they are not just an effective tool for dimension reduction; investigating linear combinations that maximize covariance has proven to be useful for many statistical problems, including regression. He et al. (2010) proposed a method to solve functional linear regression using FCC. More generally, the relationship between classical multivariate Canonical Correlation Analysis and regression was explored in Borga et al. (1997). Given the two sets of paired curves, the first *covariance weights* (or scores) are defined as

$$\langle \xi_1, x_i \rangle = \int_a^b \xi_1(t)\big(x_i(t) - \bar{x}(t)\big)dt, \quad i = 1, \ldots, n$$
$$\langle \eta_1, y_i \rangle = \int_a^b \eta_1(t)\big(y_i(t) - \bar{y}(t)\big)dt, \quad i = 1, \ldots, n$$

where the covariance functions ξ_1 and η_1 are chosen to maximize the covariance between $u_1 = (\langle \xi_1, x_1 \rangle, \ldots, \langle \xi_1, x_n \rangle)'$ and $v_1 = (\langle \eta_1, y_1 \rangle, \ldots, \langle \eta_1, y_n \rangle)'$. The process is then iterated:

to select the j-th covariance functions ξ_j and η_j and form the j-th vectors u_j and v_j, one again maximizes covariance – adding the requirement that each covariance function is orthogonal to prior ones; that is, $\text{cov}(u_\ell, u_j) = \text{cov}(u_\ell, v_j) = \text{cov}(v_\ell, v_j) = 0 \; \forall \ell < j$. $\text{cov}(u_j, v_j)$ declines at each iteration, until subsequent modes of covariation become negligible.

The identification of FCCs can be seen as an eigen-decomposition problem, and more specifically as as a *Singular Value Decomposition* (SVD) problem. Indeed, the pairs of covariance functions and covariance weights can be obtained from the SVD of the covariance operator $\Sigma_{xy} = UDV^T$. The ξ's and the η's are provided by the columns of U and V, respectively; that is, they are the left and right eigenfunctions of Σ_{xy}. D is a diagonal matrix such that $(D)_{jj} = \text{cov}(u_j, v_j)$.

To further underscore the similarity between FCCs and Functional Canonical Correlation Analysis, note that also the latter can be seen as a SVD problem, where instead of the covariance operator we consider $R_{xy} = \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2}$ (He et al., 2004). However FCCs has two main advantages. First, R_{xy} requires inversion of the functional operators Σ_{xx} and Σ_{yy}, which may be singular or almost singular in some applications, while Σ_{xy} does not require any operator inversion. Second, the functional bases generated by the SVD of R_{xy} do not allow direct reconstruction of the original curves x and y, while the functional bases generated by the SVD of Σ_{xy} do allow such reconstruction.

3.2 Low-dimensional registration through FCC

Here, we describe how to develop a low-dimensional registration based on FCCs. Let $\xi_1, \ldots, \xi_{K_x}$ and $\eta_1, \ldots, \eta_{K_y}$ be the first K_x and K_y left and right eigenfunctions of Σ_{xy}, respectively. We take these as basis systems to define the templates functions in a low-dimensional registration procedure. For each curve x_i (y_i), the template x_{0i} (y_{0i}) is given by projection of x_i (y_i) on the covariance components $\xi_1, \ldots \xi_{K_x}$ ($\eta_1, \ldots, \eta_{K_y}$). Thus, our

FCC low-dimensional registration is described by the following system of equations:

$$\begin{aligned} x_{0i}(t) &= \mathsf{PRO}_{\Sigma_{xy}}(x_i, K_x)(t) = \sum_{j=1}^{K_x} a_{ij}\xi_j(t) + \bar{x}(t), \qquad i = 1, \ldots, n \\ y_{0i}(t) &= \mathsf{PRO}_{\Sigma_{xy}}(y_i, K_y)(t) = \sum_{j=1}^{K_y} b_{ij}\eta_j(t) + \bar{y}(t), \qquad i = 1, \ldots, n. \end{aligned} \tag{5}$$

The coefficients a_{ij} and b_{ij} are given by:

$$\begin{aligned} a_{ij} &= \langle \xi_j, x_i(t) - \bar{x}(t) \rangle = \int_a^b \xi_j(t)\big(x_i(t) - \bar{x}(t)\big)dt, \\ b_{ij} &= \langle \eta_j, y_i(t) - \bar{y}(t) \rangle = \int_a^b \eta_j(t)\big(y_i(t) - \bar{y}(t)\big)dt. \end{aligned}$$

We note that, unlike in (3), the goal here is not to find a subspace $\mathcal{L}_K$ through which to reconstruct the curves once they have been registered. Instead, we only use the subspace associated to Σ_{xy} to define the registration templates. Therefore, our procedure is not affected by identifiability issues for $\mathcal{L}_K$.

This novel procedure aligns simultaneously x and y exploiting as templates projections of maximal covariance. Note that, even though x and y are registered concurrently, the procedure identifies two different sets of warping functions h_x and h_y (in the following we will sometimes omit the subscripts in h_x and h_y for simplicity) and allows one to set two different dimension K_x and K_y for the basis expansion of x and y. Theoretically, x and y live in infinite dimensional functional spaces. However, in practice, the rank of the operator Σ_{xy} (and thus the number of possible pairs of covariance components) is finite and bound above by the minimum between the number of basis functions used for expressing x and the number of basis functions used for expressing y. In applications, K_x and K_y can be chosen empirically – according to the complexity of each set of curves and to the number of components needed to capture their most meaningful modes of variability. Importantly,

for the registration to be effective, K_x and K_y should not be very large. If they are, $\mathsf{PRO}_{\Sigma_{xy}}(x_i, K_x)$ and $\mathsf{PRO}_{\Sigma_{xy}}(y_i, K_y)$ capture almost all the variability and resemble very closely the original unregistered curves x and y – so the procedure ineffectively aligns curves to templates very similar to the curves themselves. When selecting K_x and K_y we make sure that $\mathsf{PRO}_{\Sigma_{xy}}(x_i, K_x)$ and $\mathsf{PRO}_{\Sigma_{xy}}(y_i, K_y)$ remain substantially different from x and y, respectively. Rigorous criteria to select K_x and K_y should be investigated in future work.

4 Algorithms

In this Section, we describe how to implement FCC registration with two different algorithms: RS, proposed by Ramsay and Silverman (2005), and our new H1 algorithm. They both find warpings making the registered curves as similar as possible to the given templates, but they consider two different definitions of similarity.

As in (2), each warping function h_i is defined starting from the function w_i, which in turn is expressed as a linear combination of B-splines with $J = J_w$ (see(1)). Therefore

$$h_i(t; c_i) = a + (b-a)\frac{\int_a^t \mathrm{e}^{\sum_{j=1}^{J_w} c_{ij}\phi_j^b(u)}du}{\int_a^b \mathrm{e}^{\sum_{j=1}^{J_w} c_{ij}\phi_j^b(u)}du}. \tag{6}$$

For both algorithms, the target functions for the i-th curves are $x_{0i} = \mathsf{PRO}_{\Sigma_{xy}}(x_i, K_x)$ and $y_{0i} = \mathsf{PRO}_{\Sigma_{xy}}(y_i, K_y)$, as defined in (5).

For each x_i (or y_i, separately), **RS** (coded in the function register.fd of the fda R package) selects the c_{ij} coefficients in (6) to minimize $\mu_2\left(T(h_i)\right) + \lambda \int_a^b \left(\frac{d^2}{dt^2}h_i(t)\right)^2 dt$, where μ_2 is the smallest eigenvalue of the symmetric operator

$$T(h_i) = \begin{pmatrix} \int_a^b x_{0i}^2(t)dt & \int_a^b x_{0i}(t)x_i(h_i(t))dt \\ \int_a^b x_{0i}(t)x(h_i(t))dt & \int_a^b x_i^2(h_i(t))dt \end{pmatrix}.$$

A second derivative roughness penalty is applied to h_i for regularization purposes, and the minimization problem is carried out using a gradient line search method. A strong limitation of the RS algorithm is pointed out by Vantini (2012); if the L2 norm of x_i is close to 0, then μ_2 is also close to 0. Therefore, in some cases, the minimization algorithm could just reduce the L2 norm of x_i, without forcing the curve to be similar to the template.

For each x_i (or y_i), our **H1** selects the coefficients c_{ij} to minimize the H1-squared distance to the target

$$d_{H1}(x_i, x_{0i})^2 = \int_a^b \left(x_i(h_i(c_i)) - x_{0i}\right)^2 dt + \int_a^b \left(\tfrac{d}{dt}x_i(h_i(c_i)) - \tfrac{d}{dt}x_{0i}\right)^2 dt.$$

Once the c_{ij}'s are computed for $i = 1, \ldots, n$, the functions $w_i, \ldots, w_n$ are standardized so that $\mu_w(u) = \frac{1}{n}\sum_{i=1}^n w_i(u) = 0 \ \forall u \in [a, b]$ and $\int_a^b w_i(u)du = 0 \ \forall i$. The minimization problem is coded in the function newuoa of the minqua R package . For each curve, newuoa implements a minimization in J_w dimensions. Note that, unlike RS, we do not introduce a roughness penalty. However, the warpings are curves of order at most J_w – so choosing a small J_w limits their complexity and can be seen as a form of regularization. Furthermore, small J_w values reduce computation time. Note though that the smallest viable J_w is 3; for $J_w = 2$ the warpings are linear, and we cannot require them to have the same starting and ending point (or the registration would not affect the curves at all). In the simulations in Section 5, $J_w = 3$ turns out to give the best results both in terms of alignment and in terms of subsequent regression performance.

Finally, for both RS and H1, each registered $x_i(h_{x_i})$ (or $y_i(h_{y_i})$ with its own h_y) is evaluated as follows. First, we evaluate $x_i(t)$ and $h_i(t)$ on a fine grid t_{fine} to obtain $x_{i,fine}$ and $h_{i,fine}$. Then, we build $h_i^{-1}(t)$ considering $h_{i,fine}$ on the abscissa and t_{fine} on the ordinate axis, producing $h_{i,fine}^{-1}$. Lastly, we evaluate $x_i(h_i(t))$ considering $h_{i,fine}^{-1}$ on the abscissa and $x_{i,fine}$ on the ordinate axis.

For initialization, we could set all c_{ij}'s to 0 in both algorithms. However, in H1 we increase efficiency using as initial values the c_{ij}'s produced by RS. Also, since the minimization occurs separately for each curve, we implemented a parallel version of the H1.

We conclude this section with a few remarks. First, note that instead of determining the target functions from the SVD of Σ_{xy}, one could consider the SVD of $\Sigma_{x(h)y(h)}$ and the projections of each $x_i(h_i)$ and $y_i(h_i)$ – the curves already aligned. However, the operator $\Sigma_{x(h)y(h)}$ is a function of all c_{ij} coefficients – so one would need to solve a minimization problem in $2n \cdot J_w$ dimensions, which is very complicated and computationally intensive. Second, note that both RS and H1 can also be used for FPC registration – replacing the SVD of Σ_{xy} by the eigen-decomposition of Σ_{xx} for x, and of Σ_{yy} for y. These determine the target functions $x_{0i} = \mathsf{PRO}_{\Sigma_{xx}}(x_i, K_x)$ and $y_{0i} = \mathsf{PRO}_{\Sigma_{yy}}(y_i, K_y)$. Third, and perhaps most important for our purposes in this article, note that registration could be improved applying H1 and RS iteratively (see final remarks in Section 6). In Section 5, we consider only one iteration for both algorithms, since on our simulated data repeating the procedure does not lead to a sizable improvement of the results. Code for the H1 algorithm is provided as a supplementary file. All other codes and scripts, including those for data simulation (see the next Section), are available upon request.

5 Simulation study and AneuRisk data

5.1 Simulation settings

We now present a simulation study to investigate the performance of FCC registration in comparison to other registration approaches. Specifically, we apply both the RS and the new H1 algorithm using different target functions: the projection of the data on the FCC

Scheme 1 Generation and Processing of Simulated Data

1: Generate the original predictor curves: $x_i^{org}(t)$, $i = 1, \ldots, n$

2: Generate the regression coefficient surface $\beta(t, s)$

3: Generate Gaussian error curves: $\epsilon_i(s)$, $i = 1, \ldots, n$

4: Compute the original response curves: $y_i^{org}(s) = \int_a^b x_i^{org}(t)\beta(t, s)dt + \epsilon_i(s)$, $i = 1, \ldots, n$

5: Generate the warpings $h_{x_i}(t)$ and $h_{y_i}(t)$, $i = 1, \ldots, n$

6: Misalign the curves, computing $x_i^{msl}(t) = x_i^{org}(h_{x_i}^{-1}(t))$ and $y_i^{msl}(t) = y^{org}(h_{y_i}^{-1}(t))$, $i = 1, \ldots, n$

7: Apply a registration procedure to the misaligned curves, producing $\hat{w}_{x_i}$ and $\hat{w}_{y_i}$, and thus $\hat{h}_{x_i}$ and $\hat{h}_{y_i}$, $i = 1, \ldots, n$

8: Compute the registered curves $x_i^{aln} = x_i^{msl}(\hat{h}_{x_i})$ and $y_i^{aln} = y_i^{msl}(\hat{h}_{y_i})$, $i = 1, \ldots, n$

9: Fit the regression of y^{aln} on x^{aln}, producing $\hat{\beta}$ and the fitted curves y_i^{reg}, $i = 1, \ldots, n$

basis, the projection of the data on the FPC basis, and for RS also the mean of the data. We refer to these registrations as cc_{RS}, cc_{H1}, pc_{RS}, pc_{H1}, and μ_{RS}, respectively. We also consider the elastic registration, *er*, proposed by Srivastava et al. (2011) and Tucker et al. (2013) which aligns the curves using their Karcher mean as template. We consider three different simulation settings as described in **Scheme 2**, and for each setting we generate and process the data following **Scheme 1**.

To evaluate the *alignment performance* of a registration procedure, we consider the total squared H1 distances between original and aligned curves $\sum_i^n d_{H1}(x_i^{org}, x_i^{aln})^2$ and $\sum_i^n d_{H1}(y_i^{org}, y_i^{aln})^2$. To evaluate the *regression performance*, we consider three different criteria: (i) the in-sample H1 prediction error SPE $= \sum_i^n d_{H1}(y_i^{org}, y_i^{reg})^2$; (ii) the leave-one-out H1 prediction error L1OPE $= \sum_i^n d_{H1}(y_i^{org}, y_i^{pred})^2$, where y_i^{pred} is the predicted curve obtained with the $\hat{\beta}_{(-i)}$ from a fit without the i-th curves pair; (iii) the Euclidean distance between β and $\hat{\beta}$, each evaluated on a fine grid.

In all simulations, we consider $n = 20$. To generate x^{org}, β, ϵ, y^{org}, h_x and h_y, we always start from $S = 100$ equidistant raw observations in $[0, 1]$. These are obtained as

Scheme 2 Simulation Scenarios (x and β)

1: **Simulation 1**

- $x_i(t) = \sum_{j=1}^{N_{peaks}} (-1)^{sgn} v_{ij} |f_N(\mu_{ij}, \sigma_{ij}^2)|$, where:
  - f_N Normal density $\cdot$ $N_{peaks} = 4$ $\cdot$ $v_{ij} \sim N(0.5, 0.2^2)$ $\cdot$ $\mu_{ij} \sim N(0.2j, 0.008^2)$
  - $\sigma_{ij} \sim N(0.1, 0.008^2)$ $\cdot$ $sgn = 2$ for $i = 1, \ldots 12$, $sgn = 1$ for $i = 13, \ldots 24$
- $\beta(s,t) = \sin(5st)$

2: **Simulation 2**

- $x_i(t)$ as in simulation 1, with:
  - $N_{peaks} = 9$ $\cdot$ $v_{ij} \sim N(0.8, 0.03^2)$ $\cdot$ $\mu_{ij} \sim N(0.1j, 0.004^2)$ $\cdot$ $\sigma_{ij} \sim N(0.048, 0.004^2)$
  - sgn as in simulation 1
- $\beta(s,t) = \sum_{j=1}^{9} f_{N_2}(\mu_j, sI_2)$, where:
  - f_{N_2} bivariate Normal density $\cdot$ I_2, 2×2 identity matrix $\cdot$ $s = 0.0004$
  - $\mu = \big((0.1, 0.1)', (0.9, 0.9)', (0.1, 0.9)', (0.9, 0.1)', (0.3, 0.3)', (0.7, 0.7)', (0.3, 0.7)', (0.7, 0.3)', (0.5, 0.5)'\big)$

3: **Simulation 3**

- $x_i^{org}(t) = \sin \frac{1}{10(t+0.1)} + noise_i$, where:
  - $\cdot$ $noise_i$, vector of 100 samples from $N(0, 0.01)$
- $\beta(s,t) = \sum_{j=1}^{5} f_{N_2}(\mu_j, sI_2) - \sum_{j=1}^{5} f_{N_2}(\mu_j, sI_2)$, with:
  - $(\mu_1, \ldots, \mu_5) = \big((0.1, 0.1)', (0.9, 0.9)', (0.1, 0.9)', (0.9, 0.1)', (0.5, 0.5)'\big)$
  - $(\mu_6, \ldots, \mu_9) = \big((0.7, 0.7)', (0.3, 0.7)', (0.7, 0.3)', (0.5, 0.5)'\big)$ $\cdot$ $s = 0.0004$

described in Scheme 2 for x^{org} and the two directions of β, and from Normal distributions with mean 0 and variance specific to simulations for ϵ ($\sigma = 1, 1.3, 0.005$, as to render errors commensurate to the size of the coefficients used in the expansion of the response). The raw observations for y^{org} are then obtained following the model in Scheme 1. Finally, the raw observations for h_x and h_y are obtained from the formula

$$h_i(t) = \frac{e^{z_i t} - 1}{e^{z_i} - 1} \qquad i = 1, \ldots, n,$$

where $z_i \sim N(0, 1)$. Based on the raw observations, we then use B-splines with $J = 20$, order 4 and smoothing parameter $\lambda = 10^{-7}$ (see (1)) for x^{org}, ϵ, y^{org}, h_x and h_y. We use

the same parameters for the B-splines of all target functions employed in the registration procedures, the FCC and FPC directions, and for the B-splines of x^{msl} and y^{msl}. The B-splines of both directions of β also have $J = 20$ and order 4, but λ (which is the same for both directions) is specific to simulations ($\lambda = 10^{-4}, 10^{-7}, 10^{-8}$, as to minimize the L1OPE of the regression performed on the original curves). The B-splines of w_x and w_y have order 4 and $\lambda = 10^{-7}$, but J_w is set to 3 as discussed in Section 4.

Note that in the first two simulation scenarios both predictor and response curves are designed to form two different groups. This allows us to better illustrate the advantages of low-dimensional registration procedures in comparison to μ_{RS} and er – which are not capable of retaining the separation between groups. Also, in all simulations, the responses are constructed without adding an intercept; both the original x's and y's have means very close to 0. For this reason, we do not include the intercept α when applying the linear model (5). For both FCC and FPC registration and in all three simulation scenarios we choose $K_x = K_y = 1$. As we can see in supplemental Figures A.5, A.6 and A.7, just one component is sufficient to capture the two groups while markedly reducing the complexity of the target curves. In contrast, $\mathsf{PRO}_{\Sigma_{xy}}(x, 2)$, $\mathsf{PRO}_{\Sigma_{xy}}(y, 2)$, $\mathsf{PRO}_{\Sigma_{xx}}(x, 2)$ and $\mathsf{PRO}_{\Sigma_{yy}}(y, 2)$ are already too similar to the unregistered curves. Importantly, we also note that in Simulations 2 and 3 $\mathsf{PRO}_{\Sigma_{xy}}(x, 1)$ differs markedly from $\mathsf{PRO}_{\Sigma_{xx}}(x, 1)$, and $\mathsf{PRO}_{\Sigma_{xy}}(y, 1)$ from $\mathsf{PRO}_{\Sigma_{yy}}(y, 1)$. In these scenarios, the FCC and FPC registrations are guided by different templates. We thus expect them to lead to different results.

5.2 Simulation results

Salient results are summarized in Figures 1 and 2. Figure 1 shows performance in aligning predictor and response curves, and predicting the latter (out of sample) with different registration procedures. These performance values, together with the SPE and the distance

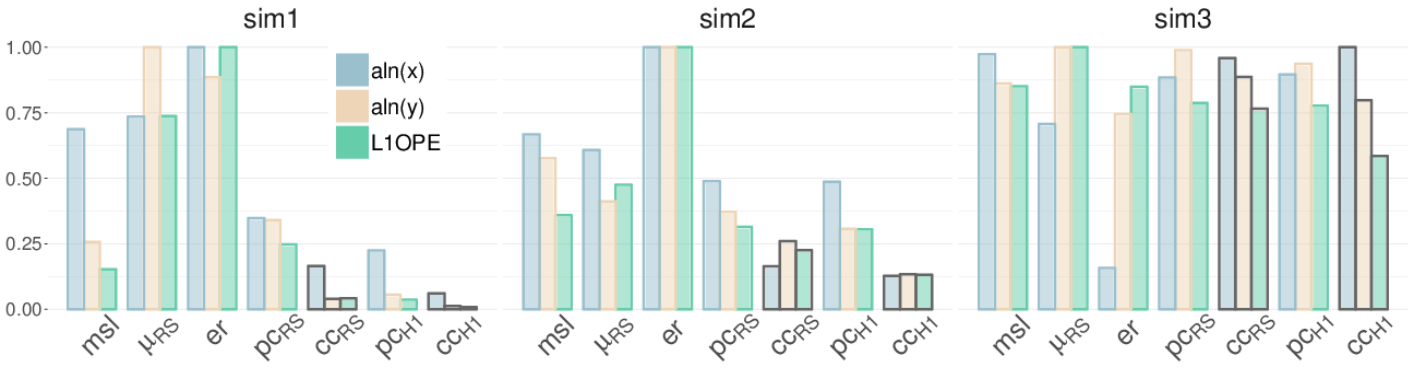

Figure 1: Simulation results. $aln(x)$ and $aln(y)$ are the $H1$ distances between aligned and original curves, and L1OPE is leave-one-out $H1$ prediction error. On the horizontal axis "msl" represents the misaligned curves, followed by curves registered with different procedures – as indicated. y-axis: Values on the vertical axis have been rescaled between 0 and 1, where 0 corresponds to results obtained with the original curves.

between β and $\hat{\beta}$, are reported in supplemental Table A.1. Figure 2 shows y^{pred}, the leave-one-out predicted response curves, from function-on-function regressions on the original data (which are perfectly aligned), the data after misalignment, and the data after the application of different registration procedures. The true β's of the 3 simulations are shown in supplemental Figure A.4.

In **Simulation 1** the x curves are characterized by two peaks and are easy to align. The surface β is smooth, so also the y curves are easy to align. In Figure 2 (upper panels) we see that the y^{pred} curves produced by μ_R do not exhibit the two peaks of the original responses, and those produced by er do not split in two groups. The y^{pred} curves obtained from low-dimensional registration with FPC and FCC appear very similar – which is not surprising given the easy nature of the registration problem in this scenario. However, in Figure 1 we see that FCC outperforms FPC in terms of both alignment and regression. Moreover, our new H1 algorithm outperforms RS. Regarding estimation of β (see supplemental Table A.1), all procedures have similar performance. Supplemental Figure A.8 shows how closely $\hat{\beta}$

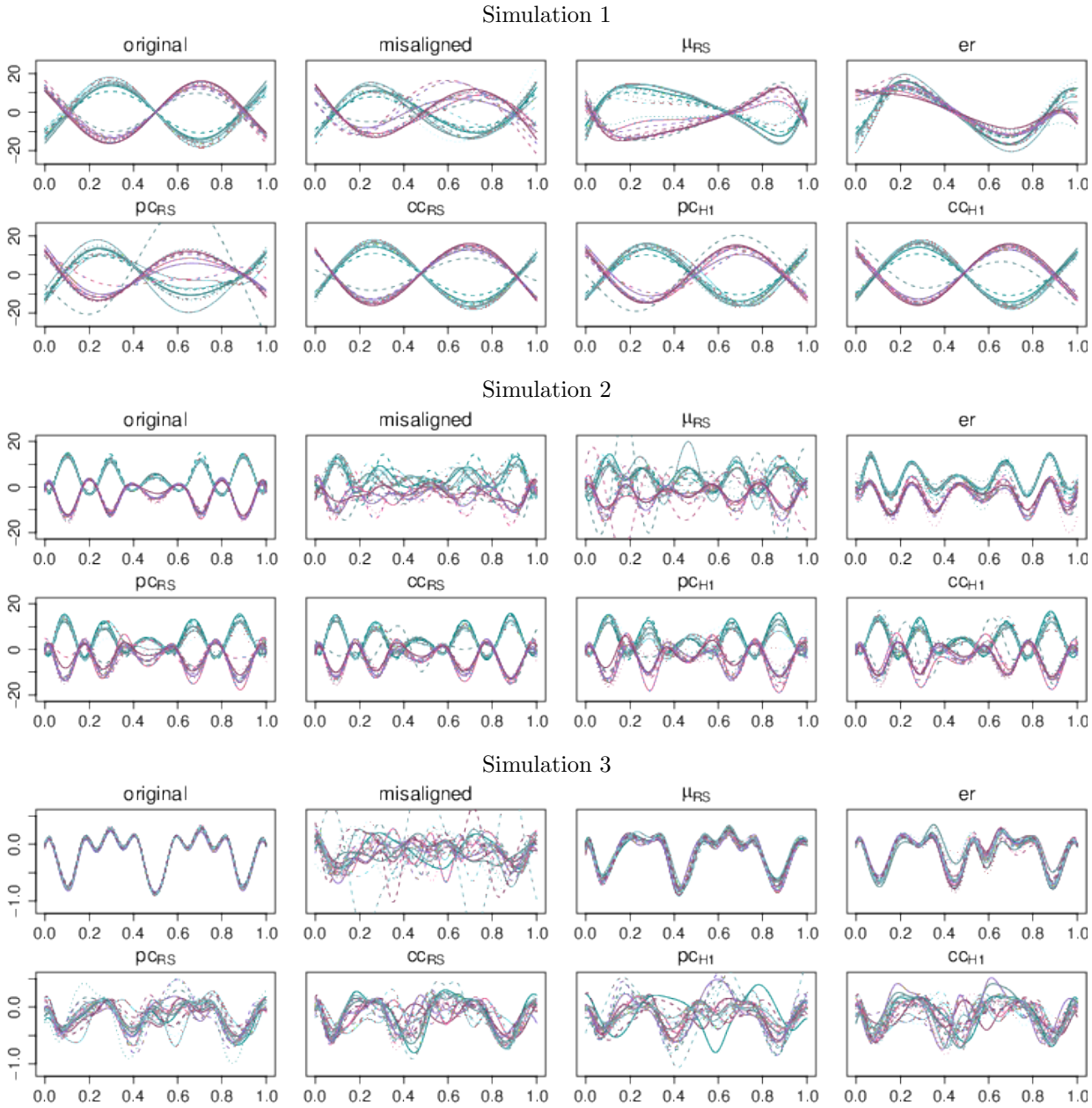

Figure 2: y^{pred} curves obtained with different regressions. "Original" indicates the regression performed on the original curves, and "misaligned" the one performed on the misaligned curves. In all other panels the regression is performed on curves registered with different procedures – as indicated. The two different colors indicate the two groups in each set of curves.

obtained from data aligned with our procedure resembles the surface estimated from x^{org} and y^{org} (prior to misalignment). In fact, as can be seen in supplemental Figure A.1, the registered curves are very similar to the original ones.

In **Simulation 2** the x curves have multiple peaks and are difficult to align. The surface β is also rather rough, leading to hard to align y curves. In this scenario FCC, and in particular cc_{H1}, gives the best alignment for both the predictor and the response, as we can see in Figure 1. Indeed, the registered curves produced by cc_{H1} (shown in supplemental Figure A.2) are very close to the original ones. Notably, cc_{H1} improves the performance by a factor of 10 with respect to er and μ_{RS} and by a factor of 3 with respect to pc_{H1} – and the latter fails in predicting the order of magnitude of some curves. In terms of regression results, H1 again dominates RS, and cc_{H1} outperforms all other registration procedures. This is also evident in Figure 2 (middle panels): y^{pred} obtained from data aligned with cc_{H1} and cc_{RS} are very similar to y^{pred} obtained from the original curves (prior to misalignment). Again, y^{pred} from μ_{RS} and er do not clearly split in two groups, and the curves registered with μ_{RS} and er resemble the original x^{org} and y^{org} even less than their misaligned versions x^{msl} and y^{msl}. Supplemental Figure A.9 shows the estimated β: none of the algorithms seems to capture the peaks that characterize the true β and the β estimated on the original curves. This is due to the difficulty of the registration problem: the aligned curves are not close enough to x^{org} and y^{org} and therefore the relationship that links x and y is described by a surface that resembles the true β.

In **Simulation 3** both the predictor and the response come from just one group. The x curves are hard to align, as in Simulation 2; here in addition to multiple peaks we have very different peak amplitudes. The surface β is again rough, resulting in hard to align y curves. In Figure 1 we see that all procedures, except perhaps for er, do rather poorly in terms of alignment. Poor alignment performance for our approach is also evident in

	msl	μ_{RS}	er	$pc_{RS(1)}$	$cc_{RS(1)}$	$pc_{RS(2)}$	$cc_{RS(2)}$	$pc_{H1(1)}$	$cc_{H1(1)}$	$pc_{H1(2)}$	$cc_{H1(2)}$
SPE	100	104	31	127	117	126	90	77	76	95	64
L1OPE	100	99	31	122	123	121	99	77	85	95	74

Table 1: AneuRisk data. SPE and L1OPE for the regressions performed on misaligned data, and on data produced by different registration procedures. For FCC and FPC, performed with H1 and RS, results are shown for both 1 *and* 2*-dimensional approximations. The values have been rescaled to have SPE and L1OPE equal to 100 on the misaligned data.*

supplemental Figure A.2, where the registered x and y curves produced by cc_{H1} are not very close to the original ones. In terms of L1OPE however, our approach seems to gain an edge – with the best performance once again achieved by cc_{H1}. Figure 2 (lower panels) shows the y^{pred} curves produced by all procedures. The predicted curves produced by er and μ_{RS} appear more "orderly" but they are, in a way, too smooth – notably, they lose one of the earlier peaks in the original curves. In contrast, while appearing more "chaotic", the predicted curves produced by the low-dimensional registration approaches better preserve the nature of the data; in particular, each predicted curve retains the same number of peaks of the corresponding original curve. Figure A.10 shows how, just like in Simulation 2, all registration procedures (including FCC) fail to produce an accurate estimate of β.

5.3 AneuRisk data

We apply our new registration procedure to the AneuRisk dataset described in Sangalli et al. (2014). This data was collected by image reconstruction of three-dimensional cerebral angiographies of 65 patients, with the aim of investigating the interplay between morphological properties of artery walls and hemodynamic factors, and shed light on the possible causes of aneurysmal pathology. Patients in the study belonged to two groups: patients with an aneurysm on the Willis circle, after the final bifurcation of the Internal Carotid Artery

(ICA), and patients with an aneurysm on the last tract of the ICA or without an aneurysm. In this article we analyze the relationship between the axial derivative of the local average of the ICA wall shear stress (WSS1), our functional response y, and the ICA curvature, our functional predictor x. y is a hemodynamic factor obtained via computational fluid dynamics in the ICA geometries (Passerini et al., 2012). x is a morphological feature of the ICA and is computed as described in Sangalli et al. (2009a). Since y is available only for 52 out of the 65 patients, the sample size for our analysis is $n = 52$. We will refer to these x and y as misaligned curves, even if they have already been registered following the procedure proposed by Sangalli et al. (2009a) and based on the three spatial coordinates of the ICA centerline. Here, we find new warping functions and perform an additional registration – with the goal of capturing the information that links the curvature to the WSS1. Importantly, if we can predict WSS1 accurately as a function of the curvature, we can avoid its computationally intense direct calculation. For comparison purposes, we consider all the procedures presented in the simulation study. For FCC and FPC we align the curves based on the projection on the first components ($K_x = K_y = 1$), as well as on the projections on first and second components ($K_x = K_y = 2$) – we denote the two variants adding subscripts "1" and "2" to the procedures' names. The reason to do this is that, unlike our simulation scenarios, the AneuRisk data has projections on the first two components which still differ markedly from the misaligned curves. This is the case for both x and y (see supplemental Figure A.11).

Regression performance is evaluated based on SPE and L1OPE. Since on real data we do not know the true original curves, the definition of these quantities is modified as SPE $= \sum_i^n d_{H1}(y_i^{aln}, y_i^{reg})^2$ and L1OPE $= \sum_i^n d_{H1}(y_i^{aln}, y_i^{pred})^2$, where y_i^{pred} is again the predicted curve obtained for i from a fit withholding the i-th curves pair.

For both x and y, we have 450 raw observations for each of the $n = 52$ subjects.

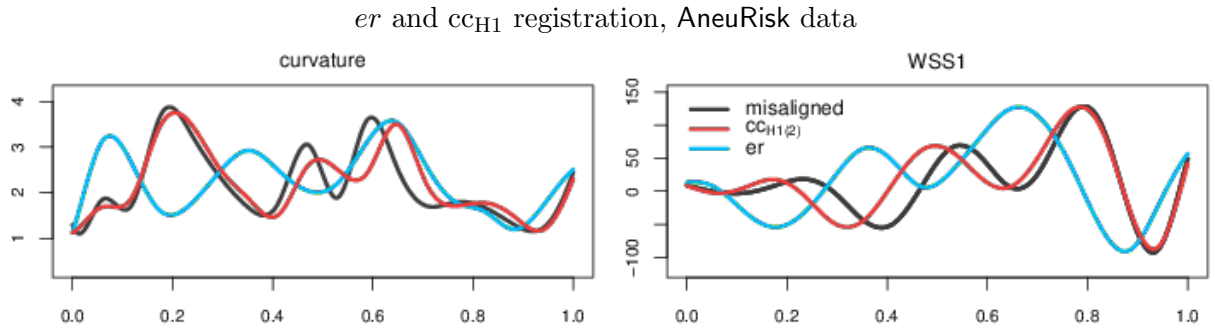

Figure 3: Curvature (x) and WSS1 (y) curves for an individual patient. Misaligned curves are drawn in grey, curves registered by $cc_{H1(2)}$ in light-blue line, and curves registered by er in red .

Their domain is transformed to be $[0, 1]$, and we use B-splines with $J = 20$, order 4 and smoothing parameter $\lambda = 10^{-7}$ to represent both x and y. Again, the same parameters are used for the basis of the FPC and FCC directions, the basis of the projections, and the basis of both the directions of β. The smoothing parameter for β, λ_β, varies for each registration and is selected to minimize the L1OPE. We do not include the intercept α in (4), since a linear model without intercept improves the L1OPE for all the procedures. Just as in the simulation study, the basis of the w's are B-splines with $J_w = 3$, order 4, and smoothing parameter $\lambda = 10^{-7}$ (increasing the number of basis elements here slows down the computation and does not appreciably improve performance).

Table 1 contains SPE and L1OPE values achieved by different regressions. Here, SPE and L1OPE of each procedure are computed contrasting predicted curves to the aligned response curves produced by the procedure itself. For this reason, comparisons among procedures should be interpreted with caution. Registering with the RS algorithm, both in FPC and FCC, leads to poor regression performance – sometimes worse than that achievable on the misaligned curves. On the contrary, registering with the H1 algorithm improves regression performance (over the misaligned curves) in all cases. Specifically,

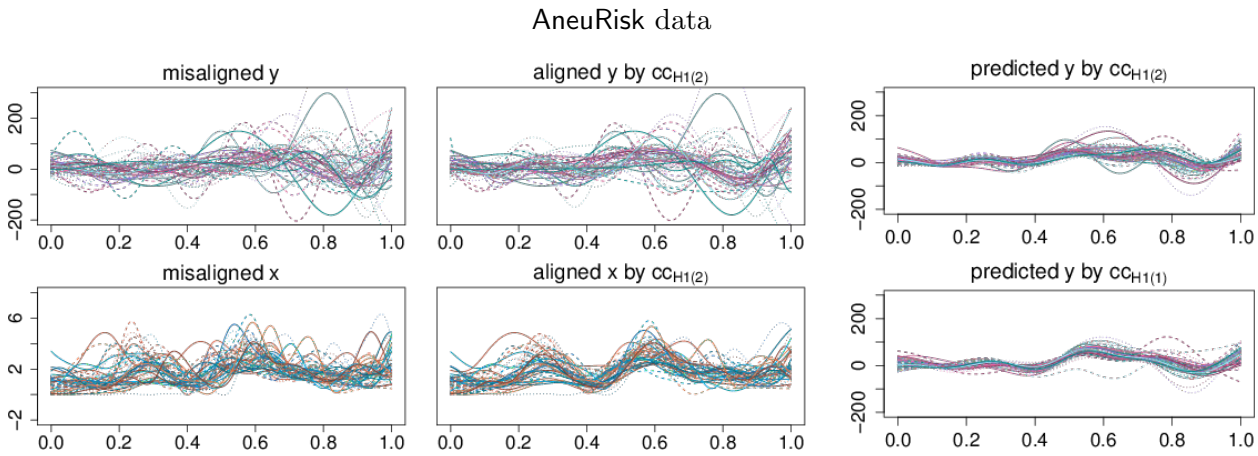

Figure 4: misaligned y (top left); y registered by $cc_{H1(1)}$ (top center); misaligned x (bottom left); x registered by $cc_{H1(1)}$ (bottom center); y^{pred} curves by $cc_{H1(1)}$ (top right) and by $cc_{H1(2)}$ (bottom right).

$cc_{H1(1)}$ and $cc_{H1(2)}$ give the second best performances both in SPE and in L1OPE.

Notably, *er* registration appears to outperform all other procedures. However, this is due to the fact that *er* produces very complex warping functions, which capture too much variability and modify the nature of the curves. To illustrate this behavior, consider the curvature (x) and WSS1 (y) of one patient reported in Figure 3. $cc_{H1(2)}$ moves and changes the amplitude of the curves' features, but does not effect their number and the general pattern of the curves themselves. On the other hand, *er* leads to over-smoothed curves for both x and y and does not preserve their features: x has 3 peaks instead of 5, and y has 2 peaks instead of 3. These over-smoothed x and y are more similar to each other, which explains why it is easier to predict y from x; regression performance increases, but this entails an unnatural change in the curves' shape and might lead to the loss of important information on the relationship between response and predictor. Predicted response curves based on all registration procedures are displayed in supplemental

Figure A.12, and estimated β's in supplemental Figure A.14. Supplemental Figure A.13 shows again predicted response curves based on all registration procedures – but focusing on just one patient. The original misaligned response curve of the patient is shown as well. Here one can clearly appreciate how only the curves predicted by FCC procedures retain all the peaks of the original curve, while other procedures tend to combine together the two highest peaks. Figure 4 shows the misaligned curves for curvature and WSS1, the curves registered by $cc_{H1(2)}$, and the predicted response curves based on $cc_{H1(2)}$ and $cc_{H1(1)}$. In terms of x registration, $cc_{H1(2)}$ combines all the curves' peaks in two big groups, one at about 0.3 and one at about 0.6. The registration of y, on the contrary, does not seem to reveal notable common landmarks or substantially modify the original curves – which have a more complex nature than the x's. In terms of predicted response curves, $cc_{H1(1)}$ captures the general pattern but also clearly misses some modes of variability present in the aligned y's. As expected, curves predicted based on $cc_{H1(2)}$ exhibit more variability (considering two components allows one to generate more complex predictions). However, they are still substantially different from the registered y's.

The results presented here, while interesting and informative, are still rather partial. Beyond the goal of approximating wall shear stress without the computational burden of its direct calculation, the broader purposes in analyzing this data concern medical diagnosis. For instance, it would be critical to accurately classify patients in two groups based on the position of their aneurysm. Further extensions of our FCC registration may provide relevant insight; in particular, it would be very useful to design a generalization of the procedure able to handle multiple functional predictors at once, and/or curves with different domains – for more on this, see Section 6.

6 Conclusions

In this article we introduced FCC registration, a new low-dimensional registration procedure based on the covariance operator between a functional response and a functional predictor. The procedure finds the most important modes of covariation between the two sets of curves, aligns them simultaneously, and at the same time performs a reduction of the data. We implement FCC registration using two different algorithms: the continuous registration algorithm introduced by Ramsay and Silverman (2005), and a new algorithm based on H1 distances. Both in our simulation study and in our application to the AneuRisk data, FCC registration improved regression performance in comparison to other registration approaches. This was particularly the case in its H1 implementation. Furthermore, FCC registration led to better alignment of the response y. Indeed, our approach leverages a low-dimensional representation built upon the dependence between functional response and functional predictor. Notably, in most cases analyzed, even for simulated curves that comprised two distinct groups, subspaces of dimension 1 were enough to capture significant modes of covariation.

FCC registration lends itself to numerous further developments. First, as mentioned in Section 4, the RS and H1 algorithms could be iterated. Iterative registration procedures have been proposed by Kneip and Ramsay (2008) and Sangalli et al. (2009a). Updating the covariance operator, and therefore the projections of x and y on the selected basis systems at each iteration, could substantially improve the registration procedure. Indeed, when performing just one iteration, we add to the projections in (5) the mean of the *misaligned* data. This may not resemble at all the mean of ideally aligned data, and thus introduce noise into the target functions. Iterating can mitigate the problem, since at each iteration one would add to the target functions the mean of the data as aligned in the previous iteration.

This may still be imperfect, but would become progressively closer to the "right" mean. We plan to explore iterations in future work, paying special attention to the definition of appropriate stopping criteria. A second avenue for further development is extending FCC registration to regressions with more than one functional predictor. This would greatly expand its applicability, since many contemporary data sets contain measurements on several potentially useful predictors. For instance, in the AneuRisk data of Section 5, considering several predictors could certainly help explain variation in the first derivative of wall sheer stress. Also interestingly, the ability to simultaneously register the response and multiple predictors would allow us to align the three spatial coordinates of the ICA centerline before computing the curvature, instead of starting directly from the curvature computed on coordinates already aligned by Sangalli et al. (2009a). Finally, extending our approach to warpings that can modify the domains of the curves will also broaden applicability – allowing one to handle a larger spectrum of real scenarios where curves are measured on different domains. This extension could proceed in several directions, e.g. allowing shifts, or pursuing local (as opposed to global) registration of the curves.

We conclude pointing out an important interpretation of FCC registration in the context of Sufficient Dimension Reduction (SDR). SDR is a set of techniques to handle regression problems with a large number of predictors (see Cook and Weisberg (2009), Adragni and Cook (2009) and Ma and Zhu (2013) for details). Unlike variable selection, which assumes that among all available predictors only a few are truly related to the response, SDR assumes that the response depends only a few linear combinations – potentially loading on many, or even all predictors. The goal of SDR is to estimate a subspace able to capture all the regression information, called the central space, or the effective dimension reduction (e.d.r) space (Li, 1991). Recently, SDR has been extended to the case of functional data, see Lee et al. (2013) and Li et al. (2017).

Starting from the work of Borga et al. (1997) and of Li and Duan (1989) in the finite-dimensional multivariate context, one can evince the link between covariance components and SDR. Borga et al. (1997) point out the relationship between Canonical Correlation Analysis and the direction identified by the OLS regression. Li and Duan (1989) prove that, under some assumptions on the distribution of the predictors, OLS can be considered as the very first and simplest SDR technique when the central space has dimension 1. In the infinite-dimensional functional framework, the reduction produced by our FCC registration can in fact be seen as an SDR procedure – estimating a subspace of dimension 1 (or larger) which captures the function-on-function regression information. Many interesting developments can be pursued based on this connection, such as finding a registration procedure based on other SDR techniques, e.g. Sliced Inverse Regression (SIR) (Li, 1991) and its several functional versions (fSIR) – see Ferré and Yao (2003), Ferré and Yao (2005), and Wang et al. (2015).

References

Adragni, K. P. and R. D. Cook (2009). Sufficient dimension reduction and prediction in regression. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 367*(1906), 4385–4405.

Borga, M., T. Landelius, and H. Knutsson (1997). A unified approach to pca, pls, mlr and cca. *Linkping University, Department of Electrical Engineering*.

Cook, R. D. and S. Weisberg (2009). *Applied regression including computing and graphics*, Volume 488. John Wiley & Sons.

Ferré, L. and A.-F. Yao (2003). Functional sliced inverse regression analysis. *Statistics 37*(6), 475–488.

Ferré, L. and A.-F. Yao (2005). Smoothed functional inverse regression. *Statistica Sinica*, 665–683.

He, G., H.-G. Müller, and J.-L. Wang (2004). Methods of canonical analysis for functional data. *Journal of Statistical Planning and Inference 122*(1), 141–159.

He, G., H.-G. Müller, J.-L. Wang, and W. Yang (2010). Functional linear regression via canonical analysis. *Bernoulli*, 705–729.

Horváth, L. and P. Kokoszka (2012). *Inference for functional data with applications*, Volume 200. Springer Science & Business Media.

Kneip, A. and J. O. Ramsay (2008). Combining registration and fitting for functional models. *Journal of the American Statistical Association 103*(483), 1155–1165.

Lee, K.-Y., B. Li, F. Chiaromonte, et al. (2013). A general theory for nonlinear sufficient dimension reduction: Formulation and estimation. *The Annals of Statistics 41*(1), 221–249.

Leurgans, S. E., R. A. Moyeed, and B. W. Silverman (1993). Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, 725–740.

Li, B., J. Song, et al. (2017). Nonlinear sufficient dimension reduction for functional data. *The Annals of Statistics 45*(3), 1059–1095.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association 86*(414), 316–327.

Li, K.-C. and N. Duan (1989). Regression analysis under link violation. *The Annals of Statistics*, 1009–1052.

Ma, Y. and L. Zhu (2013). A review on dimension reduction. *International Statistical Review 81*(1), 134–150.

Passerini, T., L. M. Sangalli, S. Vantini, M. Piccinelli, S. Bacigaluppi, L. Antiga, E. Boccardi, P. Secchi, and A. Veneziani (2012). An integrated statistical investigation of internal carotid arteries of patients affected by cerebral aneurysms. *Cardiovascular Engineering and Technology 3*(1), 26–40.

Ramsay, J. O. and B. W. Silverman (2002). *Applied functional data analysis: methods and case studies*, Volume 77. Springer.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. Springer.

Sangalli, L. M., P. Secchi, S. Vantini, et al. (2014). Aneurisk65: A dataset of three-dimensional cerebral vascular geometries. *Electronic Journal of Statistics 8*(2), 1879–1890.

Sangalli, L. M., P. Secchi, S. Vantini, and A. Veneziani (2009a). A case study in exploratory functional data analysis: geometrical features of the internal carotid artery. *Journal of the American Statistical Association 104*(485), 37–48.

Sangalli, L. M., P. Secchi, S. Vantini, and A. Veneziani (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: Series C (Applied Statistics) 58*(3), 285–306.

Sangalli, L. M., P. Secchi, S. Vantini, and V. Vitelli (2010). K-mean alignment for curve clustering. *Computational Statistics & Data Analysis 54*(5), 1219–1233.

Srivastava, A., W. Wu, S. Kurtek, E. Klassen, and J. Marron (2011). Registration of functional data using fisher-rao metric. *arXiv preprint arXiv:1103.3817*.

Tucker, J. D., W. Wu, and A. Srivastava (2013). Generative models for functional data using phase and amplitude separation. *Computational Statistics & Data Analysis 61*, 50–66.

Vantini, S. (2012). On the definition of phase and amplitude variability in functional data analysis. *Test 21*(4), 676–696.

Wagner, H. and A. Kneip (2018). Nonparametric registration to low-dimensional function-spaces. pp. Manuscript. Universitt Bonn, Bonn, Germany.

Wang, G., J. Zhou, W. Wu, and M. Chen (2015). Robust functional sliced inverse regression. *Statistical Papers*, 1–19.

A SUPPLEMENTARY MATERIAL

	simulation 1					simulation 2					simulation 3				
	$aln(x)$	$aln(y)$	SPE	L1OPE	β	$aln(x)$	$aln(y)$	SPE	L1OPE	β	$aln(x)$	$aln(y)$	SPE	L1OPE	β
org	0	0	216	241	88	0	0	6822	12329	59	0	0	55	71	48
msl	51581	91981	26937	28903	165	299637	626904	362190	391828	252	7128	2737	1877	2421	142
μ_{RS}	55234	357792	138570	136232	86	272671	447316	287644	514350	144	5181	3180	2743	2832	52
er	75100	316671	189336	185892	85	449045	1087473	1057176	1069024	98	1154	2366	2546	2415	58
pc_{RS}	26140	121821	33285	46391	99	219185	404149	317832	344136	130	6480	3142	2160	2244	88
cc_{RS}	12397	14017	7245	8193	81	73634	282772	228127	250489	121	7022	2819	2116	2184	112
pc_{H1}	16858	19640	5950	7453	93	218500	334265	294909	334020	201	6564	2977	2020	2216	83
cc_{H1}	4533	4567	1580	1873	84	57104	144979	119381	151666	161	7326	2538	1599	1685	112

Table A.1: Simulation results. Columns: $H1$ distances between the curves in each row and the original x and y curves; in sample $H1$ prediction error; leave-one-out $H1$ prediction error; Euclidean distances between $\hat{\beta}$ and the true β. Rows: Org, the original curves; msl, the misaligned curves; followed by curves registered with different procedures – as indicated.

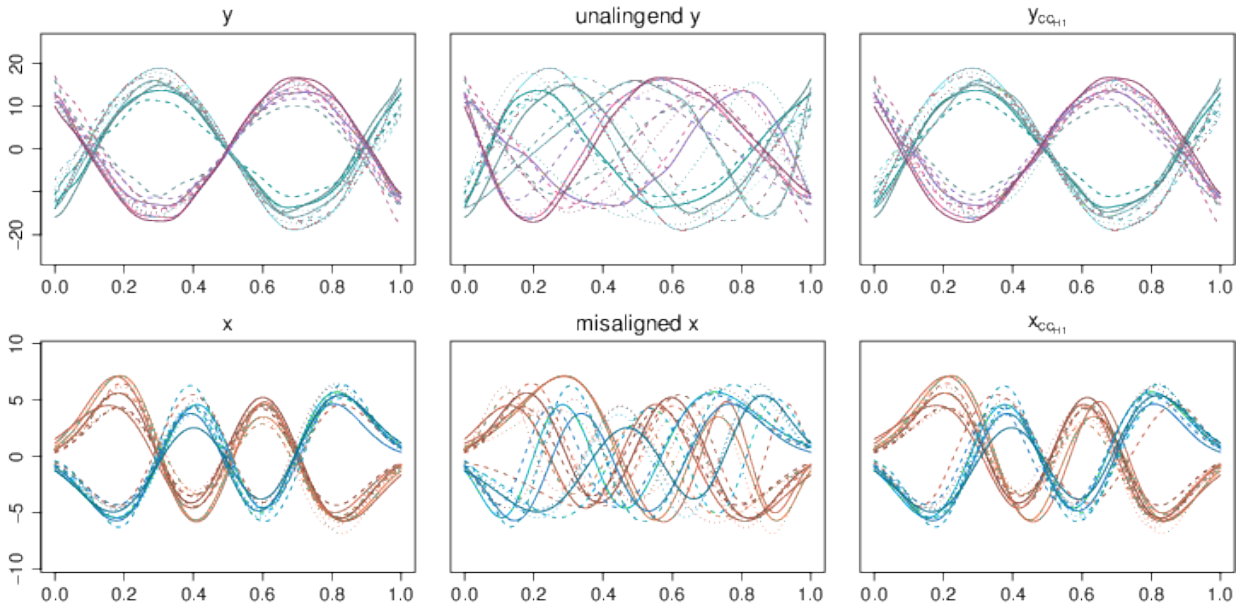

Figure A.1: Simulation 1 original curves (left), misaligned curves (center), registered curves by cc_{H1R} *(right) for* y *(top) and* x *(bottom).*

Original, misaligned and registered curves - simulation 2

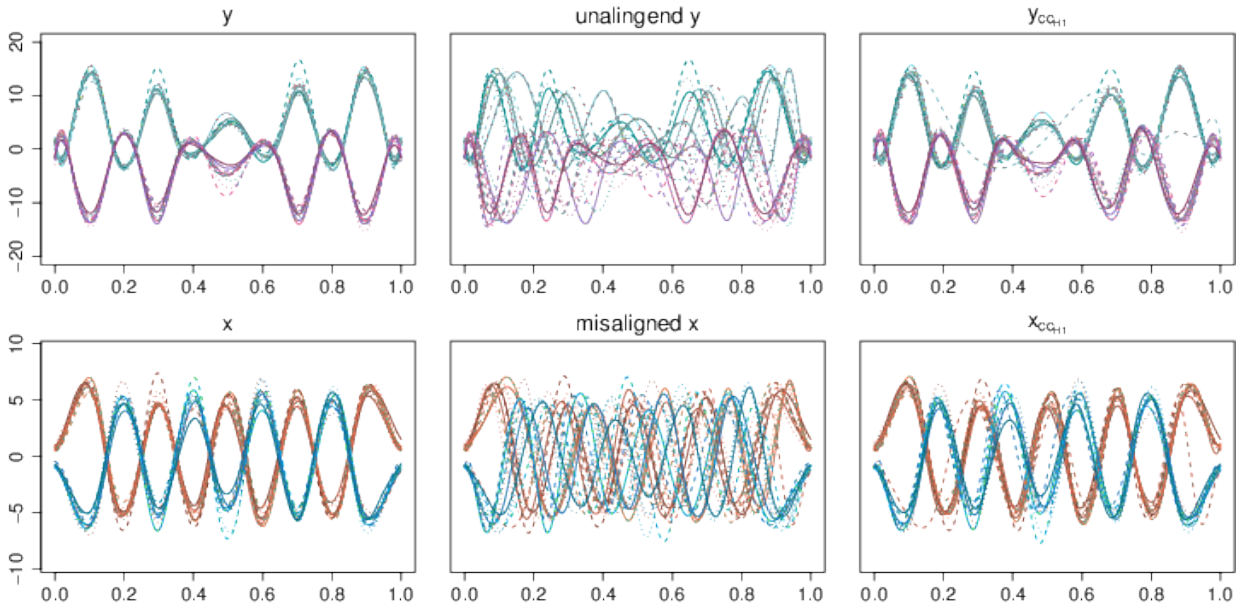

Figure A.2: Simulation 2 original curves (left), misaligned curves (center), registered curves by cc_{H1R} *(right) for* y *(top) and* x *(bottom).*

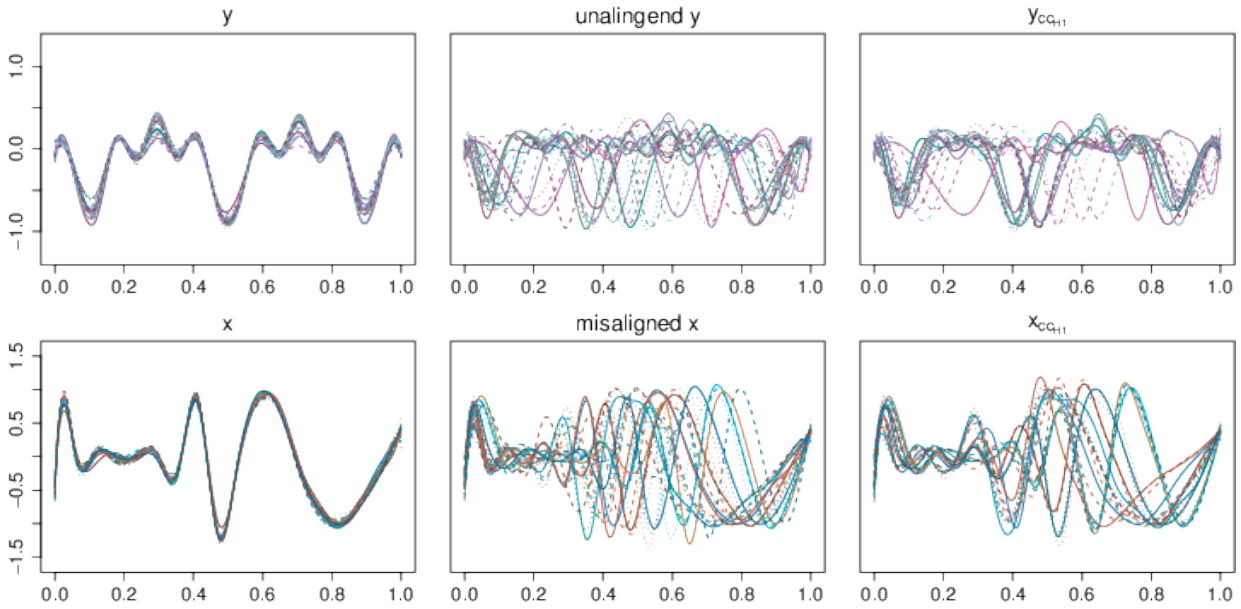

Figure A.3: Simulation 3 original curves (left), misaligned curves (center), registered curves by cc_{H1R} (right) for y (top) and x (bottom).

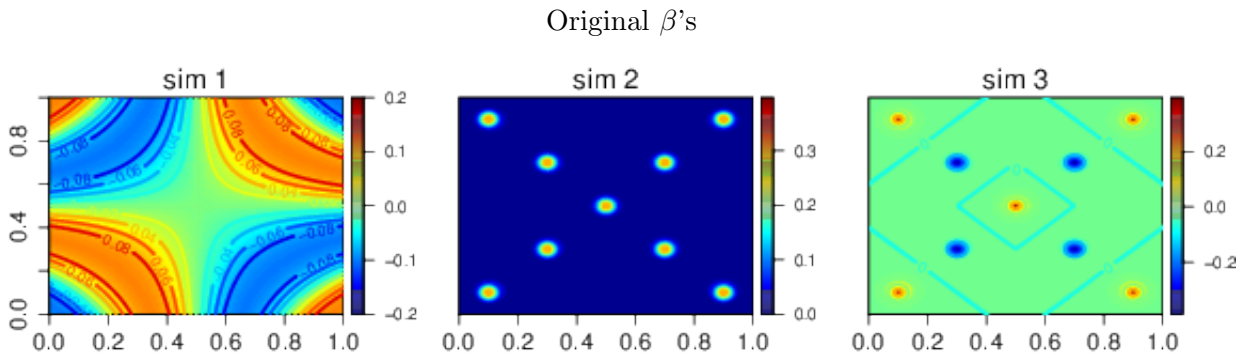

Figure A.4: Original β for simulation 1 (left), simulation 2 (center) and simulation 3 (right).

Projections - simulation 1

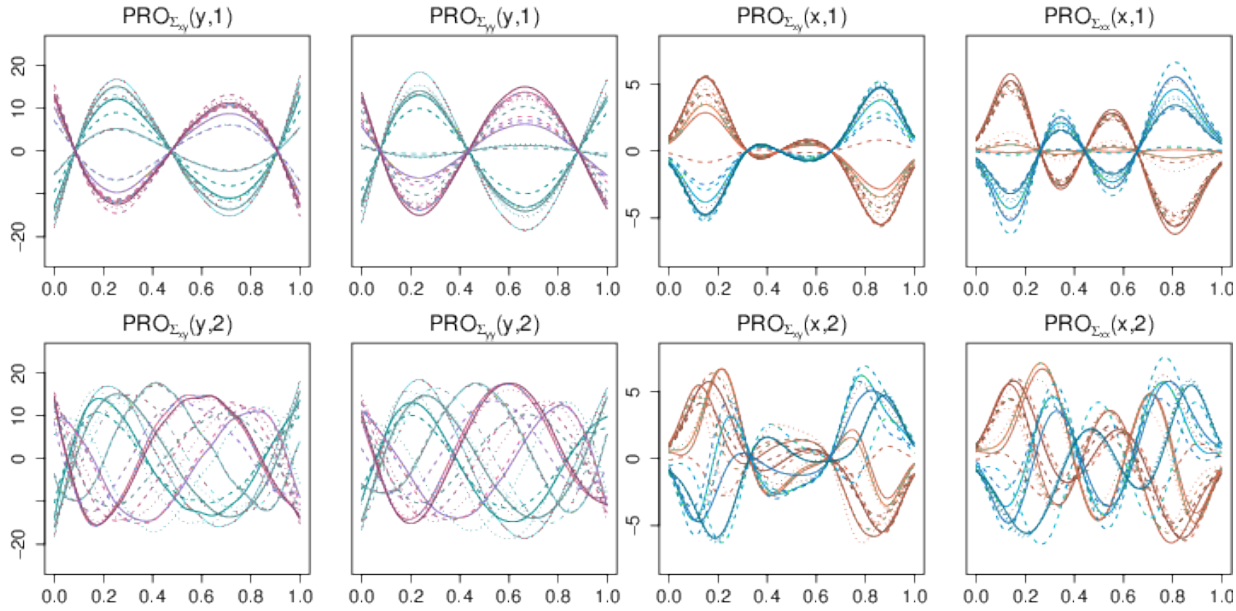

Figure A.5: Simulation 1 projections of the misaligned curves on the first (top) and second (bottom) principal components (second and fourth column) and canonical directions (first and third columns) of y (first and second columns) and x (third and fourth columns).

Projections - simulation 2

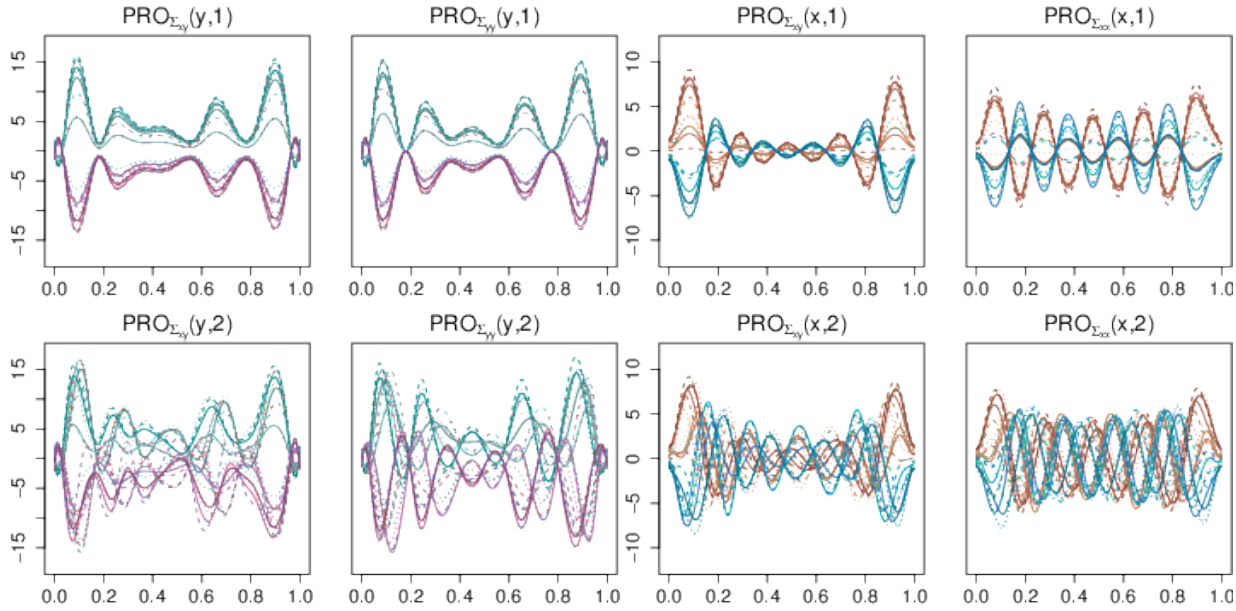

Figure A.6: Simulation 2 projections of the misaligned curves on the first (top) and second (bottom) principal components (second and fourth column) and canonical directions (first and third columns) of y (first and second columns) and x (third and fourth columns).

Projections - simulation 3

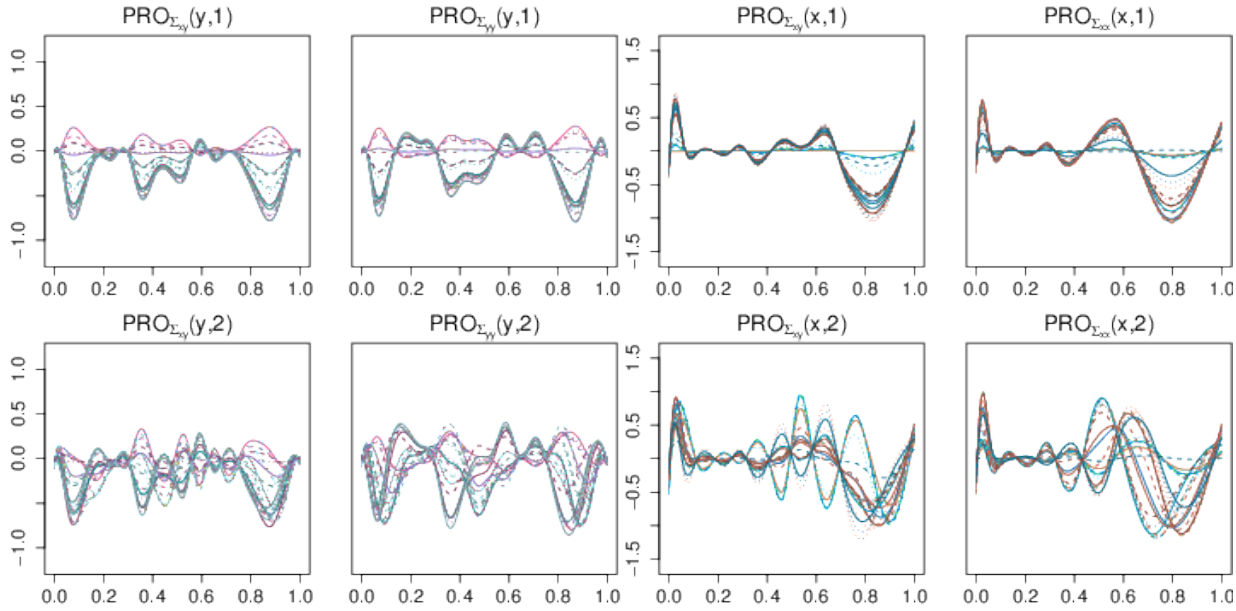

Figure A.7: Simulation 3 projections of the misaligned curves on the first (top) and second (bottom) principal components (second and fourth column) and canonical directions (first and third columns) of y (first and second columns) and x (third and fourth columns).

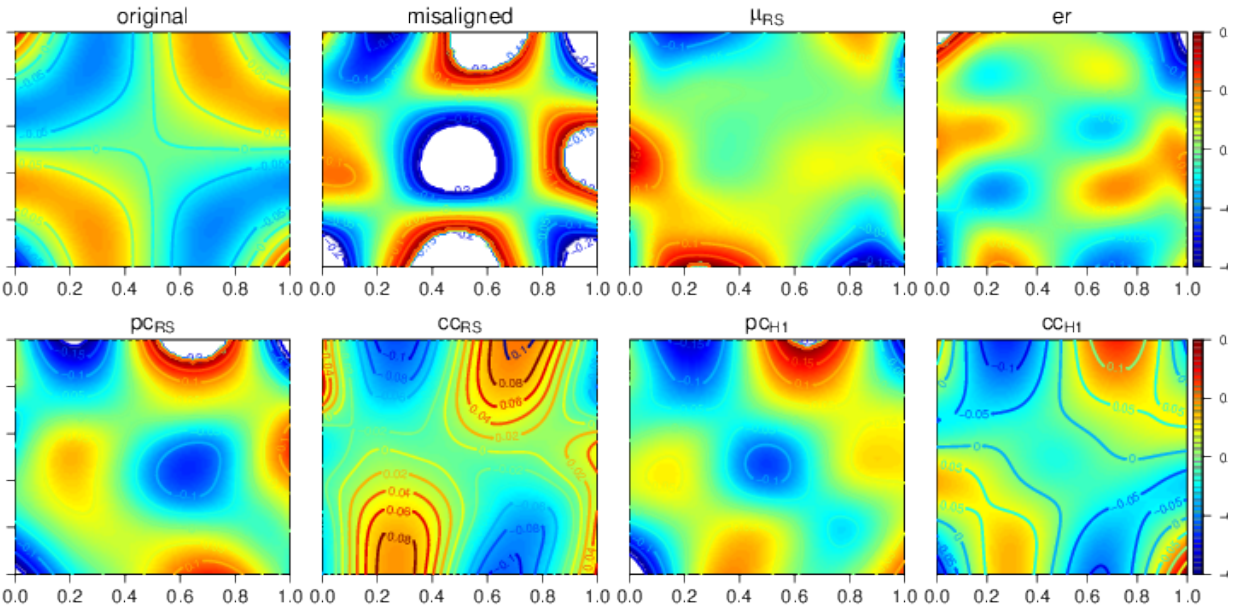

Figure A.8: Simulation 1 estimated β's by different regressions. Original: the regression is performed on the original curves; misaligned: the regression is performed on the misaligned curves. In the other cases the regression is performed on the curves registered by the indicated algorithm. The white zones are values outside the z-range on the right.

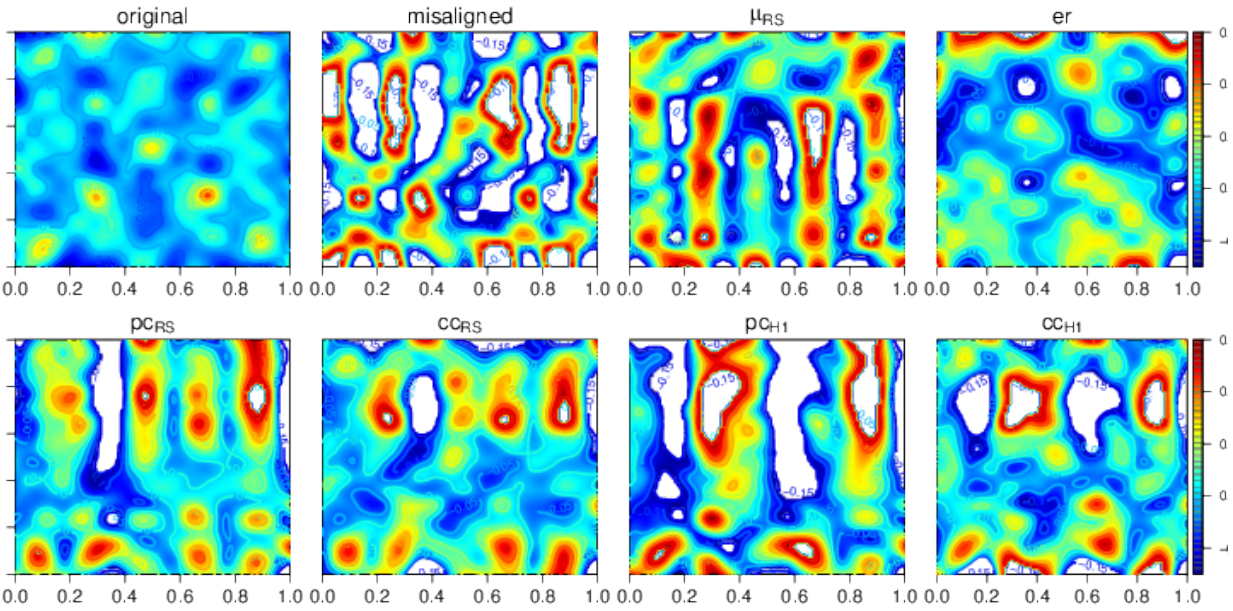

Figure A.9: Simulation 2 estimated β's by different regressions. Original: the regression is performed on the original curves; misaligned: the regression is performed on the misaligned curves. In the other cases the regression is performed on the curves registered by the indicated algorithm. The white zones are values outside the z-range on the right.

$\hat{\beta}$'s - simulation 3

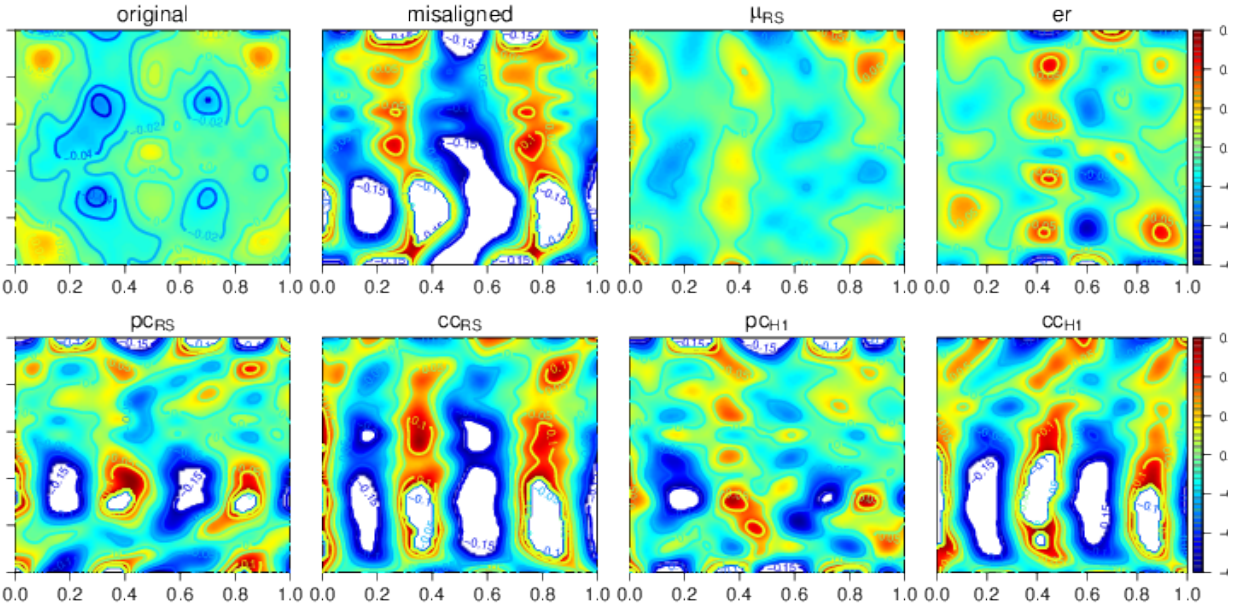

Figure A.10: Simulation 3 estimated β's by different regressions. Original: the regression is performed on the original curves; misaligned: the regression is performed on the misaligned curves. In the other cases the regression is performed on the curves registered by the indicated algorithm. The white zones are values outside the z-range on the right.

Projections - AneuRisk

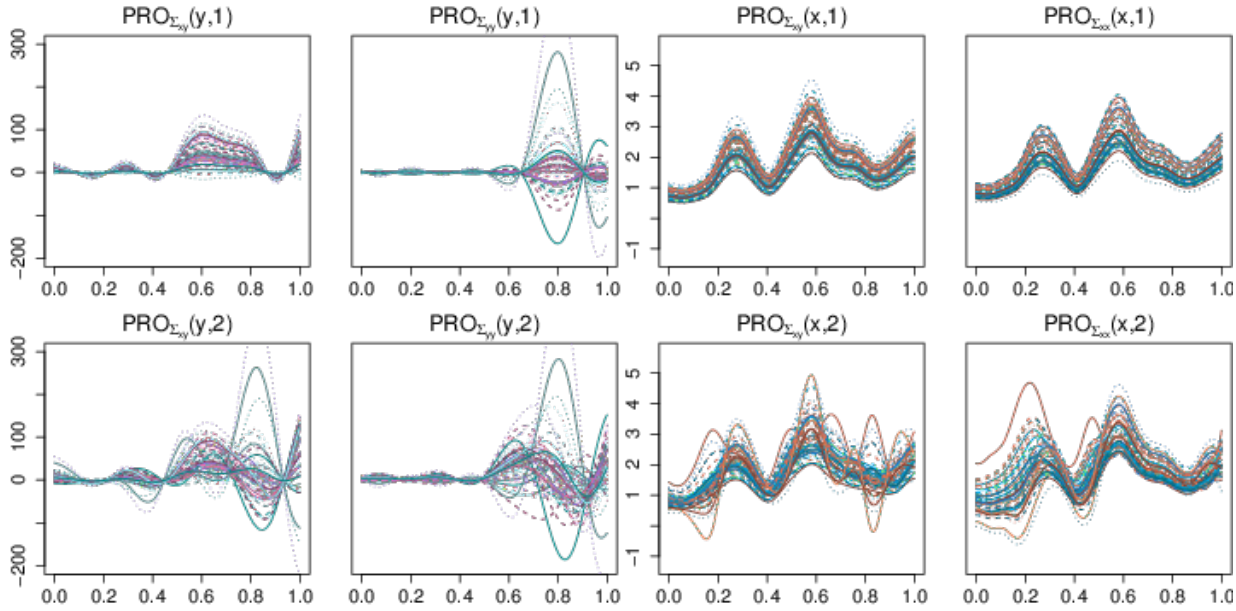

Figure A.11: AneuRisk data, projections of the misaligned curves on the first (top) and second (bottom) principal components (second and fourth column) and canonical directions (first and third columns) of y (first and second columns) and x (third and fourth columns).

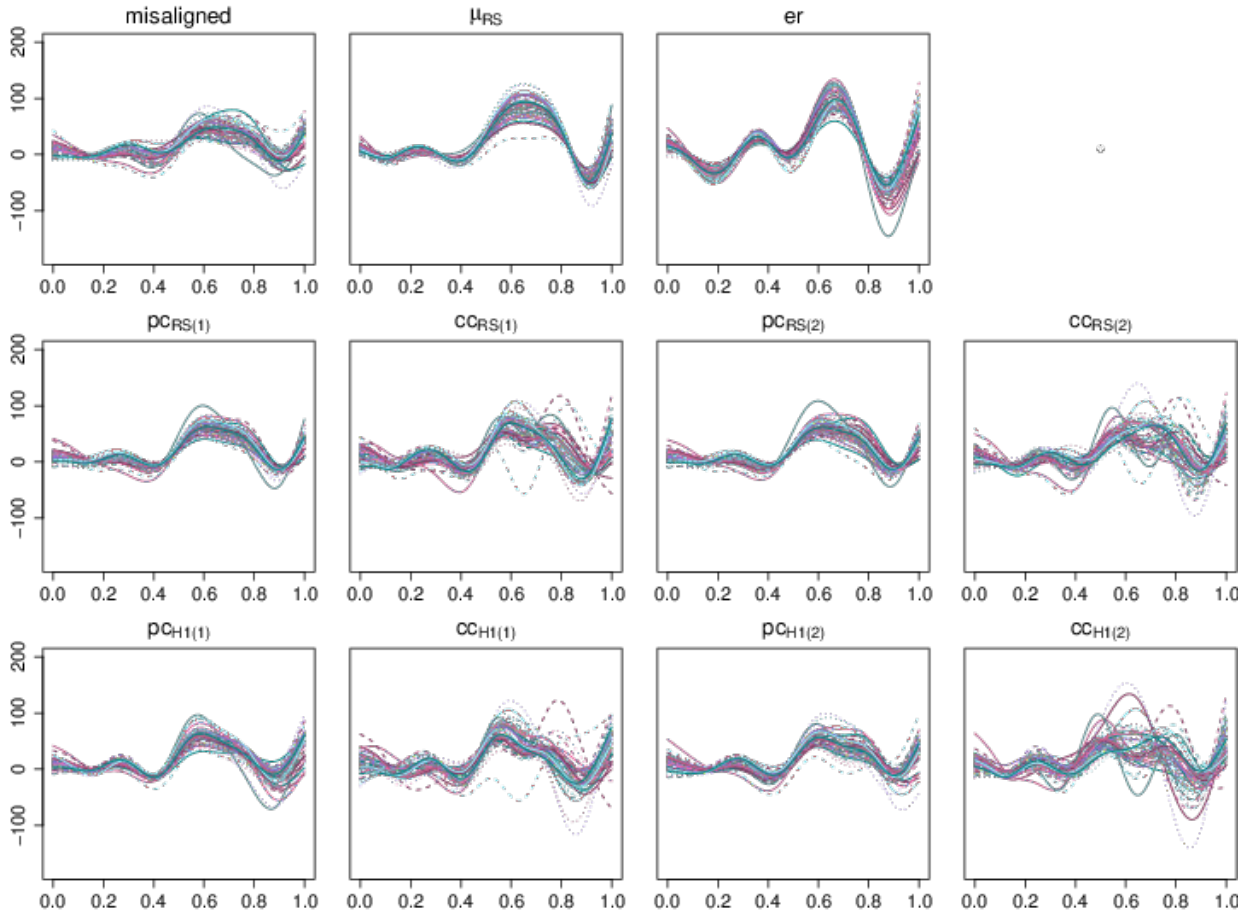

Figure A.12: AneuRisk data, leave-one-out predicted curves by different regressions.

Leave-one-out predicted curve for one patient - AneuRisk

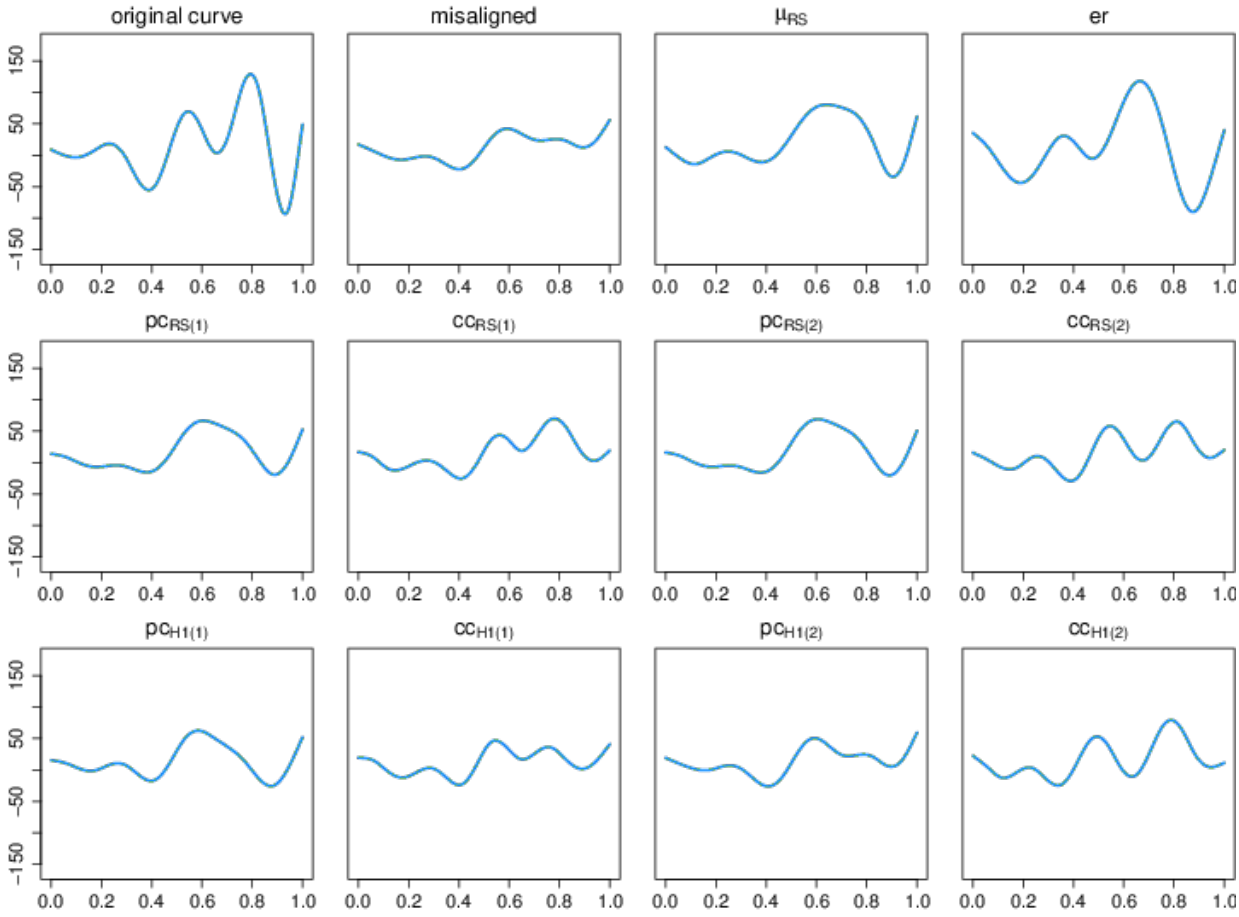

Figure A.13: AneuRisk data, leave one predicted curve by different regressions for just one patient. The top left panel represents the original misaligned curve before any registration.

$\hat{\beta}$'s - AneuRisk

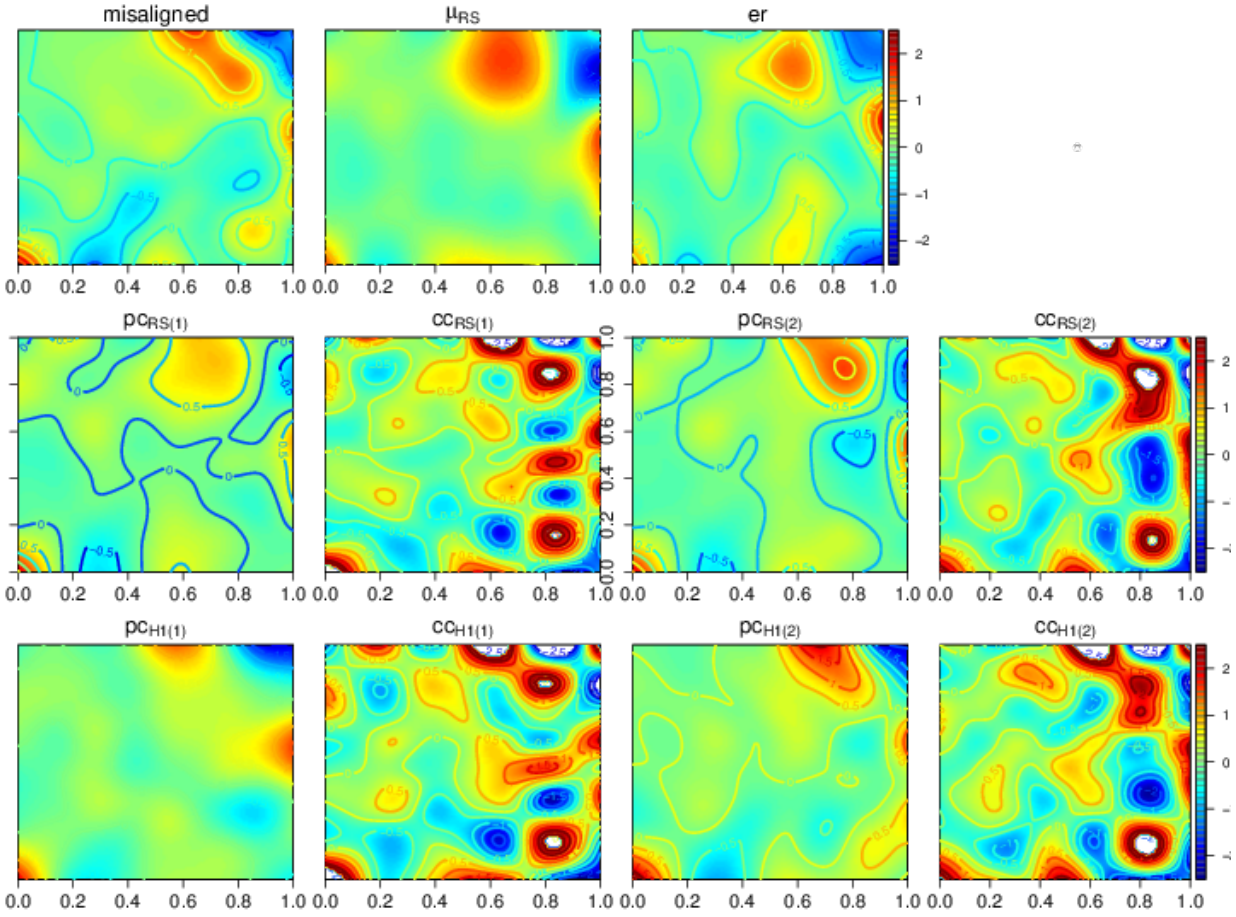

Figure A.14: AneuRisk data, estimated β's by different the different algorithms. The white zones are values outside the z-range on the right.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

64/2018 Menafoglio, A.; Pigoli, D.; Secchi, P.
Kriging Riemannian Data via Random Domain Decompositions

63/2018 Brugiapaglia, S.; Micheletti, S.; Nobile, F.; Perotto, S.
Wavelet-Fourier CORSING techniques for multi-dimensional advection-diffusion-reaction equations

61/2018 Zakerzadeh, R.; Zunino P.
A Computational Framework for Fluid-Porous Structure Interaction with Large Structural Deformation

62/2018 Perotto, S.; Carlino, M.G.; Ballarin, F.
Model reduction by separation of variables: a comparison between Hierarchical Model reduction and Proper Generalized Decomposition

59/2018 Martino, A.; Guatteri, G.; Paganoni, A. M.
Multivariate Hidden Markov Models for disease progression

58/2018 Ferro, N.; Micheletti, S.; Perotto, S.
A sequential coupling of shape and topology optimization for structural design

57/2018 Ferro, N.; Micheletti, S.; Perotto, S.
POD-assisted strategies for structural topology optimization

56/2018 Antonietti, P.F.; Manzini, G.; Verani, M.
The conforming virtual element method for polyharmonic problems

55/2018 Cerroni, D.; Laurino, F.; Zunino, P.
Mathematical analysis, finite element approximation and numerical solvers for the interaction of 3D reservoirs with 1D wells

54/2018 Dal Santo, N.; Deparis, S.; Manzoni, A.; Quarteroni, A.
Multi space reduced basis preconditioners for parametrized Stokes equations