Landmark-Based Registration of Curves via the Continuous Wavelet Transform

Jérémie BIGOT

This article is concerned with the problem of the alignment of multiple sets of curves. We analyze two real examples arising from the biomedical area for which we need to test whether there are any statistically significant differences between two subsets of subjects. To synchronize a set of curves, we propose a new nonparametric landmark-based registration method based on the alignment of the structural intensity of the zero-crossings of a wavelet transform. The structural intensity is a recently proposed multiscale technique that highlights the main features of a signal observed with noise. We conduct a simulation study to compare our landmark-based registration approach with some existing methods for curve alignment. For the two real examples, we compare the registered curves with FANOVA techniques, and a detailed analysis of the warping functions is provided.

Key Words: Curve alignment; FANOVA; Feature detection; Nonparametric estimation; Scale-space representation; Structural intensity; Zero-crossings lines.

1. INTRODUCTION

When studying some biological or physical process in different subjects, we usually see that the observed curves have a common structural pattern. Then, an important issue arises in deciding whether there are any significant differences between two subsets of subjects or in determining the typical shape of the observed process. For instance, when studying growth curves of boys and girls, one might be interested in estimating the curve that best represents the typical shape of the growing process and in testing if its shape shows variations between boys and girls (see Gasser and Kneip 1995; Ramsay and Silverman 2002).

To compare similar objects, it is generally necessary to find a common reference to represent them. Curve alignment consists of finding, for each observed curve, a warping function in order to synchronize all the curves before performing the average or applying any other statistical inferential procedure. A survey on recent developments in the analysis

Jérémie Bigot is Assistant Professor, Laboratoire de Statistique et Probabilités, University Paul Sabatier, Toulouse, France (E-mail: *Jeremie.Bigot@math.ups-tlse.fr*).

DOI: 10.1198/106186006X133023

of deformations and warping can be found in a tutorial by Younes (2000) and extensive references on curve alignment for functional data analysis can be found in Ramsay and Silverman (2005). In what follows, the terms *alignment*, *warping*, *registration*, or *matching* will also be used to refer to the *synchronization* of set of signals. Curve alignment is thus a preliminary task that is often necessary before the statistical analysis of a dataset. Matching two functions can be done by aligning individual locations of corresponding structural points (or landmarks) from one curve to another. Previous approaches to landmark-based registration in a statistical setting include Kneip and Gasser (1992), Gasser and Kneip (1995), Ramsay and Li (1998), Munoz Maldonado, Staniswallis, Irwin, and Byers (2002), and Bigot (2003). For landmark-based matching one needs to detect the landmarks of a set of signals from discrete (noisy) observations. The estimation of the landmarks is usually complicated by the presence of noise whose fluctuations might give rise to spurious estimates which do not correspond to structural points of the unknown signals. Then, it is necessary to determine the landmarks that should be associated. This step is further complicated by the presence of outliers and by the fact that some landmarks of a given curve might have no counterpart in the other curves. Generally, these steps are performed manually (see, e.g., Munoz Maldonado et al. 2002) which can be tedious if the number of signals is large. This article uses the scale-space approach proposed by Bigot (2003, 2005) to estimate the landmarks of a noisy function. This method is based on the estimation of the significant zero-crossings of the continuous wavelet transform of a noisy signal, and on a new tool, the structural intensity, proposed by Bigot (2003, 2005) to represent the landmarks of a signal via a probability density function. The main modes of the structural intensity correspond to the significant landmarks of the unknown signal. In a sense, the structural intensity can be viewed as a smoothing method that highlights the significant features of a signal observed with noise.

Continuous matching approaches are based on the optimization of an appropriate functional. They yield a warping function which minimizes a global misalignment criterion between two signals [see Trouve and Younes (2000) for further details on 1D matching by variational methods; Ramsay and Li (1998), Gervini and Gasser (2004), Liu and Müller (2004)]. Dynamic time warping (DTW), was introduced by Sakoe and Chiba (1978), is probably the most popular technique to automatically register a set of curves [see also Wang and Gasser (1997, 1999) who studied DTW from a statistical point of view]. Bigot (2003) proposed to register by DTW the structural intensities of two noisy functions that have to be aligned. After registration the modes of the structural intensities that are at the same locations correspond to the landmarks that have to be aligned. However, DTW can be time consuming if the number of data points is large. This article proposes a fast and automatic method to align the significant landmarks of a set of noisy signals. This approach is also based on the alignment of the structural intensities, but it avoids the use of DTW. Its computational cost depends only on the number of landmarks and is therefore very low. To justify our wavelet-based approach and to evaluate its performances, we run some simulations to compare it with the continuous matching method of Ramsay and Li (1998) and a DTW algorithm using the standard L^2 cost. Two real examples are also used to illustrate our landmark-based registration approach: a comparison of the distribution of ganglioside

in brain tissues between old and young rats, and an example arising from physiology. The first example was analyzed by Munoz Maldonado et al. (2002) to understand the physical process of aging in brain tissues that could explain changes in performance. In what follows, this dataset will be referred to as the ganglioside dataset. For the second example, data on human movement were recorded by Dr. Amarantini David and Dr. Martin Luc (Laboratoire Sport et Performance Motrice, EA 597, UFRAPS, Grenoble University, France). The motivation of this study was to better understand the processes underlying movement generation under various levels of an external force applied to the knee of various young male subjects. In what follows, this dataset will be referred to as the orthosis dataset.

To illustrate the usefulness of curve alignment, we provide a detailed analysis of the warping functions and their relationship to the registered curves for these two datasets. We also use a functional analysis of variance (FANOVA) model to compare these various sets of curves, and we briefly summarize the results in this article. We chose a fixed-effects FANOVA model, proposed by Abramovich, Antoniadis, Sapatinas, and Vidakovic (2004), to quantify the variations within a set of curves. Abramovich et al. (2004) used the orthosis dataset to illustrate their hypothesis testing procedures. However, their analysis was carried out *without preregistration* of the observed data. We have therefore applied the same hypothesis tests *with preregistration* of the orthosis dataset to see if curve alignment changes significantly the results. This FANOVA model is also used for comparing two samples of curves from the ganglioside dataset. The results are compared with the conclusions of the study carried out by Munoz Maldonado et al. (2002).

The article is organized as follows. Section 2 briefly describes the problem of landmark-based registration and recalls the scale-space approach developed by Bigot (2003, 2005). A new nonparametric curve alignment methodis also described. A real example involving a curve registration step is used to illustrate the methodology. To justify our approach, we use some simulated data to compare our landmark-based registration technique with some existing methods for curve alignment. Then, in Section 3, a detailed analysis of the ganglioside and orthosis datasets is proposed. We carefully analyze the warping functions and the synchronized curves, and we briefly present the results obtained with the FANOVA model of Abramovich et al. (2004). Finally, Section 4 provides concluding remarks and proposes some hints for possible extensions of our methodology.

2. LANDMARK-BASED REGISTRATION

Suppose we are given m ($m \geq 2$) unknown signals $f_j, j = 1, \ldots, m$ (with $f_j : [0,1] \to \mathbb{R}$, $j = 1, \ldots, m$) observed with noise at the same discrete time positions $t_i = i/n$, $i = 1, \ldots, n$:

$$y_{j,i} = f_j(t_i) + \sigma_j \epsilon_{j,i}, \tag{2.1}$$

where $\epsilon_{j,i}$ are iid normal variables with zero mean and variance 1, and σ_j are unknown noise level parameters. The problem considered in this article is the estimation of m smooth warping functions $u_1, \ldots, u_m$ such that the registered curves $\hat{f}_1 \circ u_1, \ldots, \hat{f}_m \circ u_m$ minimize a suitable measure of discrepancy for a set of signals, where $\hat{f}_j$ denotes some estimator of the

curve f_j. Note that, once the warping functions have been estimated, we choose to register the discrete observations by linear interpolation of the raw data. For notation convenience, we assume that all the curves are observed on the same interval $[0, 1]$. If it is not the case, we can define the functions $\tilde{f}_j = f_j(a_j + (b_j - a_j)t), t \in [0, 1]$ when f_j is observed on the interval $[a_j, b_j]$ to retrieve our setting.

Suppose we are given two sets of labeled landmarks $(\tau_{1,1}, \ldots, \tau_{1,N})$ and $(\tau_{2,1}, \ldots, \tau_{2,N})$ extracted from two curves f_1 and f_2. The purpose of landmark-based registration is to find an increasing function u such that $u(\tau_{1,i}) \approx \tau_{2,i}$ for all $i = 1, \ldots, N$. In this article, it will be also required that $u(0) = 0$ and $u(1) = 1$. Further details on landmark-based registration in a statistical setting can be found in Kneip and Gasser (1992), and landmark-based registration for a set of m curves ($m > 2$) will be considered in Section 2.5.

To compute the transformation u, we choose the monotone smoothing technique of Ramsay (1998) which consists in finding a smooth velocity function $\omega(t)$ which minimizes the following penalized least-squares criterion:

$$\sum_{i=1}^{N} (u(\tau_{1,i}) - \tau_{2,i})^2 + \lambda \int_0^1 (\omega^{(p)}(t))^2 dt, \tag{2.2}$$

where $\omega^{(p)}(t)$ denotes the derivative of order p of $\omega(t)$, λ is a regularization parameter, and the warping function u is given by

$$u(t) = C_0 + C_1 \int_0^t \exp\left(\int_0^z \omega(v)dv\right) dz \quad \text{with} \quad C_0, C_1 \quad \text{abitrary constants.} \tag{2.3}$$

The formulation (2.2) transforms the problem of estimating a monotone function u to one of estimating the velocity $\omega(t)$ which is an unconstrained function. Ramsay (1998) proposed to express the function $\omega(t)$ as a linear combination of B-splines $B_k(t)$, $\omega(t) = \sum_{k=1}^{K} c_k B_k(t)$. The criterion (2.2) is then numerically optimized with respect to the coefficients c_k. For all the computations in this article, we chose to take B-splines of order $q = 6$ with $K = N + q - 2$ and $p = 0$.

When aligning data, a regularization parameter (λ in Equation (2.2)) controls the trade-off between exact matching and the amount of warping (i.e., the distance between u and the identity, which is controlled by the second term in Equation (2.2)). For landmark-based registration, exact matching corresponds to $u(\tau_{1,i}) = \tau_{2,i}$ for all $i = 1, \ldots, N$. The calibration of this regularization parameter is crucial because there is always a choice to be made between how much warping to allow, and retaining a "reasonable" between-curve variability. However, it is quite difficult to design a data-based method to automatically choose such a regularization parameter. For the datasets studied in this article, the value of this parameter is usually chosen by visual inspection of the registered curves and the smoothness of the warping functions. Note that we shall discuss the choice of this parameter throughout the paper (especially in Section 3 for the analysis of the orthosis and ganglioside datasets).

2.1 SCALE-SPACE ESTIMATION OF THE SIGNIFICANT LANDMARKS

This section briefly recalls the scale-space approach proposed by Bigot (2003, 2005) to automatically estimate the landmarks of an unknown signal. This approach is based on the detection of the significant zero-crossings of the continuous wavelet transform of a signal observed with noise.

Let $f \in L^2(\mathbb{R})$ and $\psi = (-1)^r \theta^{(r)}$ where $r \geq 1$ is the number of vanishing moments of the wavelet ψ, and θ is a smooth function with a fast decay such that $\int_{-\infty}^{\infty} (\psi(u))^2 du = 1$. Then, by definition, the continuous wavelet transform of f at a given scale $s > 0$ is

$$W_s(f)(x) = \int_{-\infty}^{+\infty} f(u)\psi_s(u-x)du \quad \text{for} \quad x \in \mathbb{R},$$

where $\psi_s(u) = \frac{1}{\sqrt{s}}\psi(\frac{u}{s})$. The term *wavelet maxima* is used to describe any point (m_0, s_0) in the time-scale plane such that $z \mapsto |W_{s_0}(f)(z)|$ is locally maximum at $z = m_0$. Mallat and Hwang (1992) showed that the local regularity of a function is related to the decay of the wavelet transform amplitude in the time-scale plane, and have introduced the so-called wavelet transform modulus maxima (WTMM) method for detecting singularities in a signal (see Mallat 1999, chap. 6 for further details). The term *zero-crossings* will be used to describe any point (z_0, s_0) in the time-scale space such that $z \mapsto W_{s_0}(f)(z)$ has exactly one zero at $z = z_0$ in a neighborhood of z_0. We will say that z_0 is a positive zero-crossing if $W_{s_0}(f)(z) > 0$ for all $z \in [z_0 - \epsilon, z_0[$ with ϵ sufficiently small, otherwise z_0 is said to be a negative zero-crossing. We will call *zero-crossings line* (respectively wavelet maxima line) any connected curve $z(s)$ in the time-scale plane (x, s) along which all points are zero-crossings (respectively wavelet maxima). Now, note that if f is C^r in an interval $[a, b]$, then for all $x \in]a, b[$

$$\lim_{s \to 0} \frac{W_s(f)(x)}{s^{r+1/2}} = K f^{(r)}(x), \quad \text{where} \quad K = \int_{-\infty}^{+\infty} \theta(u)du \neq 0. \tag{2.4}$$

Hence, Equation (2.4) shows that at fine scales the zero-crossings of $W_s(f)(x)$ converge to the zeros of $f^{(r)}$ in $]a, b[$ (if any) . Thus, one can find the locations of the extrema (respectively the points of inflexion) of a function by following the propagation at small scales of the zero-crossings when ψ has $r = 1$ (respectively $r = 2$) vanishing moment(s).

The WTMM method suggests that the singularities of a function can be detected by following the propagation of the wavelet maxima at fine scales. However, we are not guaranteed that for a wavelet ψ, any sequence of modulus maxima converges when the scale decreases. For instance, if $W_s(f)(x)$ has a modulus maxima located at (m_1, s_1), then $|W_s(f)(x)|$ may have no more maxima in the neighborhood of m_1 when s goes to zero. Hummel and Moniot (1989) and Yuille and Poggio (1986) showed that this is never the case if θ is a Gaussian. Hence, using wavelets that are derivatives of Gaussian guarantees that all the wavelet maxima lines $m(s)$ are never interrupted when s goes to zero. The same property also holds for the zero-crossings: if θ is a Gaussian then all the zero-crossings lines propagate up to fine scales. Given that we want to detect the landmarks of a signal by following the propagation of the zero-crossings at fine scales, we will suppose that θ is a Gaussian in the rest of the

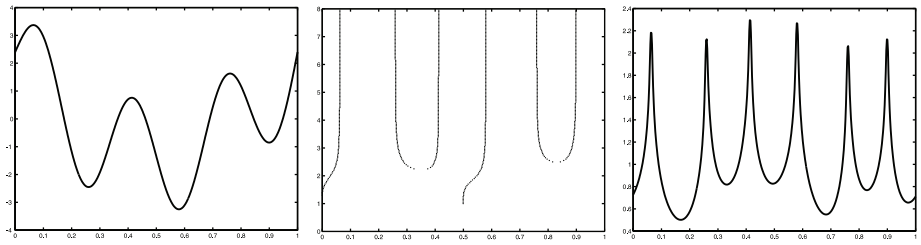

Figure 1. (a) Signal with various extrema, (b) zero-crossings computed for a Gaussian wavelet with $r = 1$ vanishing moment: the vertical axis gives $-log_2(s)$, (c) structural intensity of the zero-crossings.

article. In Figure 1, the zero-crossings of a simulated signal are computed for a Gaussian wavelet with $r = 1$ vanishing moment. One can see that the zero-crossings lines are never interrupted and converge to the extrema of the signal when the scale s goes to zero.

The WTMM method has been developed for the characterization of the singularities of a *deterministic* signal. Bigot (2005) considered the problem of estimating the singularities of a function when it is observed with noise. This approach is based on the adaptation of the WTMM method to the case of noisy observations. A nonparametric approach was proposed by Bigot (2005) to estimate the wavelet maxima of an unknown signal at various scales. The procedure is based on the estimation of some appropriate scales at which the wavelet maxima due to the presence of a significant singularity in the signal dominate the wavelet maxima created by the fluctuations of the noise. An estimation of the true wavelet maxima is then obtained by an appropriate thresholding of the modulus of the continuous wavelet transform at these scales.

When a smooth signal is observed with noise, its local extrema can be detected by estimating the significant zero-crossings of its continuous wavelet transform at various scales. In Bigot (2003, 2005) the estimation of the wavelet maxima and the zero-crossings lines was considered in the white noise model, that is, when f is observed according to

$$Y(dx) = f(x)dx + \tau B(dx), \quad x \in [0,1], \tag{2.5}$$

where τ is a noise level parameter, and B is a standard Brownian motion. When $\tau = \frac{\sigma}{\sqrt{n}}$, the white noise model (2.5) is closely related to the standard nonparametric regression problem (2.1) with n equi-spaced observations of the function f on $[0,1]$ (see Brown and Low 1996; Donoho and Johnstone 1999). The wavelet transform of $f \in L^2([0,1])$ at a scale $s > 0$ is defined to be $W_s(f)(x) = \int_0^1 f(u)\psi_s(u-x)du$, while the wavelet transform of the white noise $B(du)$ is: $W_s(B)(x) = \int_{-\infty}^{+\infty} \psi_s(u-x)B(du)$, for $x \in [0,1]$. Then, the wavelet transform of Y is

$$W_{s,n}(Y)(x) = \int_{-\infty}^{+\infty} \psi_s(u-x)Y(du) = W_s(f)(x) + \frac{\sigma}{\sqrt{n}} W_s(B)(x). \tag{2.6}$$

Let $q^n_{1-\alpha}$ be the quantile such that $P(\max_{x\in[0,1]} |W_{n^{-1}}(B)(x)| \leq q^n_{1-\alpha}) \to 1-\alpha$ as $n \to \infty$ for some $0 < \alpha < 1$. Because $\{W_s(B)(x); x \in [0,1]\}$ is a stationary Gaussian process, the expression of $q^n_{1-\alpha}$ in a closed form can be easily derived by using Lemma 5.2

of Bigot (2005) which gives the asymptotic distribution of $\max_{x\in[0,1]} |W_s(B)(x)|$ when the scale s tends to zero:

$$q^n_{1-\alpha} = \frac{1}{\sqrt{2\log(n)}} \left(-\log\left(-\frac{1}{2}\log(1-\alpha) \right) + 2\log(n) + \log\left(\frac{\sqrt{\int_{-\infty}^{+\infty} \{\psi'(u)\}^2}}{2\pi} \right) \right).$$

In Bigot (2003), the following statistical test is considered to estimate the zero-crossings of $W_s(f)(x)$ at a given scale s: fail to reject the null hypothesis $H_0^{s,x} : W_s(f)(x) = 0$ if $|W_{s,n}(Y)(x)| \leq \frac{\sigma}{\sqrt{n}} q^n_{1-\alpha}$, versus: conclude significant evidence for $W_s(f)(x)$ being positive or negative if $W_{s,n}(Y)(x) > \frac{\sigma}{\sqrt{n}} q^n_{1-\alpha}$ or $W_{s,n}(Y)(x) < -\frac{\sigma}{\sqrt{n}} q^n_{1-\alpha}$, respectively. Each point of significant zero-crossings of $W_s(f)(x)$ is then located between a pair of points $x_1, x_2 \in [0,1]$ such that H_0^{s,x_1} and H_0^{s,x_2} are rejected, and $W_{s,n}(Y)(x_1)$ and $W_{s,n}(Y)(x_2)$ have opposite signs. Let $s_{\max}$ be the largest scale at which one wants to estimate the zero-crossings of $W_s(f)(x)$. The following empirical estimation of the zero-crossings is then proposed in Bigot (2003): for $s \leq s_{\max}$, let $\{(\hat{x}_{1,s,i}, \hat{x}_{2,s,i}); 1 \leq i \leq \hat{p}_s\}$ be the sequence of points of $]0,1[\times]0,1[$ such that: $H_0^{s,\hat{x}_{1,s,i}}$ and $H_0^{s,\hat{x}_{2,s,i}}$ are rejected, $W_{s,n}(Y)(\hat{x}_{1,s,i})$ and $W_{s,n}(Y)(\hat{x}_{2,s,i})$ have opposite signs, and all the hypotheses $H_0^{s,x}$ for $\hat{x}_{1,s,i} < x < \hat{x}_{2,s,i}$ fail to be rejected. For $1 \leq i \leq \hat{p}_s$, if $W_{s,n}(Y)$ has only one zero-crossing $z_{1,s,i}$ in $[\hat{x}_{1,s,i}; \hat{x}_{2,s,i}]$ define $\hat{z}_i(s) = z_{1,s,i}$ otherwise: discard the zero-crossings located in $[\hat{x}_{1,s,i}; \hat{x}_{2,s,i}]$ since most of them are likely to be due to the fluctuations of the noise. Note that we choose to discard co-located zero-crossings because the significant extrema of a signal are the most useful landmarks, and these appear as isolated zero-crossings in the time-scale plane.

The algorithm described above yields an estimation of the zero-crossings lines at various scales. However, this only gives a visual representation that indicates "where" the landmarks are located, but there is generally no analytical expression of these lines in a closed form. The structural intensity is a new tool introduced by Bigot (2003, 2005) to identify the limits of the wavelet maxima or the zero-crossings lines when they propagate to fine scales. The definition of the structural intensity is based on the following remarks: if $x_0 \in \mathbb{R}$ is a landmark of some signal, then all the lines $z_{x_0}(s)$ that may converge to it are included in a "small" neighborhood of x_0 at fine scales (see, e.g., Figure 1). Hence, if we could compute the "density" of the points $z_{x_0}(s)$ along various scales, it would be expected that the resulting intensity would possess exactly one mode located at x_0. This idea is similar to the method proposed by Gasser and Kneip (1995) to identify features that occur consistently in a set of curves. For instance, when one searches to identify common local maxima in a set of curves, Gasser and Kneip (1995) proposed to retrieve all local maxima in each individual curves, to sort them in one array and then to submit it to kernel density estimation. Common maxima then give rise to peaks in the resulting density. In our setting, the functions $x \mapsto W_s(f)(x)$ can be viewed as a set of curves indexed by the scale parameter s. Hence, the structural intensity method consists in using the zero-crossings of $W_s(f)(x)$ at various scales to compute a "density" whose local maxima will be located at the landmarks of f. More precisely, the structural intensity of the estimated zero-crossings is defined to be:

$$G_z(x) = \sum_{i=1}^{\hat{p}} \int_{\hat{s}_i^1}^{\hat{s}_i^2} \frac{1}{s} \theta\left(\frac{x - \hat{z}_i(s)}{s}\right) ds, \tag{2.7}$$

where $[\hat{s}_i^1, \hat{s}_i^2]$ is the support of the zero-crossings line $\hat{z}_i(.)$ in the time-scale plane, and $\hat{p}$ is the number of estimated zero-crossings lines. The *landmarks of the unknown signal* f are then defined as the *local maxima* of $G_z(x)$ on $[0,1]$. The structural intensity is therefore a tool to identify the limits of the lines $\hat{z}_i(.), i = 1, \ldots, \hat{p}$ in the time-scale plane. In Figure 1, one can see that the local maxima of the structural intensity correspond to the extrema of the signal (note that in Figure 1, the structural intensity is computed with the true zero-crossings). In Bigot (2005), the structural intensity is also used to identify the limits of the estimated wavelet maxima lines that correspond to the significant singularities in a signal. If the structural intensity is normalized to be a probability density function, then a mass can be attributed to each local maxima of G_z to determine its strength. In Bigot (2005), a measure of modality (proposed by Fisher and Marron 2001) is used to select the main modes of a structural intensity that reveal significant evidence of a landmark, and to discard other minor modes that contain no information and can therefore be neglected. Thus, all the structural intensities plotted in this article are normalized to be probability density functions.

2.2 Practical Implementation and a Real Example

Note that for the implementation of the continuous wavelet transform, we do not need a dyadic sequence of sampling points since all the computations for discrete data are based on the discrete Fourier transform. However, we need to assume that the design is equally spaced. If it is not the case, then one has to adapt the definition of the discrete continuous wavelet transform to the case of nonequally spaced designs. We believe that it can be done if the continuous wavelet transform is considered as a kernel smoothing method where the scale s corresponds to the usual bandwidth parameter. Such a scale-space view of kernel smoothing methods was proposed by Chaudhuri and Marron (1999, 2000) in the context of nonparametric curve estimation for determining the significant features in a functional dataset.

To perform the detection of the zero-crossings, one needs to estimate the level of noise σ. For all the simulated and real examples proposed in this article, we use the robust estimate suggested by Donoho and Johnstone (1998) based on the *median absolute deviation* of the empirical wavelet coefficients associated with an orthonormal wavelet basis of $L^2([0,1])$ (here, we took the Symmlet 8 wavelet basis). Again, note that the number of data points does not need to be dyadic, since the estimation of σ is based on the discrete wavelet coefficients at the finest scale, which are simply obtained by applying an appropriate high-pass filter to the data (see Mallat 1999, chap. 7 for further details).

To illustrate this approach, we provide a real example of curve registration. The two curves $\tilde{f}_1$ and $\tilde{f}_2$ plotted in Figure 2(a) are haddock sounds which were recorded for the purpose of automatically identifying an individual fish from the sounds it produces (for both curves the number of data points is $n = 200$). These curves are part of a larger dataset presented in detail and analyzed by Wood (2002). As explained by Wood (2002), classifying fish sounds is a problem which can be solved by registration techniques. In Figure 2(b), we plotted the estimated zero-crossings lines of $\tilde{f}_1$ and $\tilde{f}_2$ computed for $r = 1$. We denote by $\tilde{G}^+_{z1}$ and $\tilde{G}^+_{z2}$ (respectively $\tilde{G}^-_{z1}$ and $\tilde{G}^-_{z2}$) the structural intensities computed with the

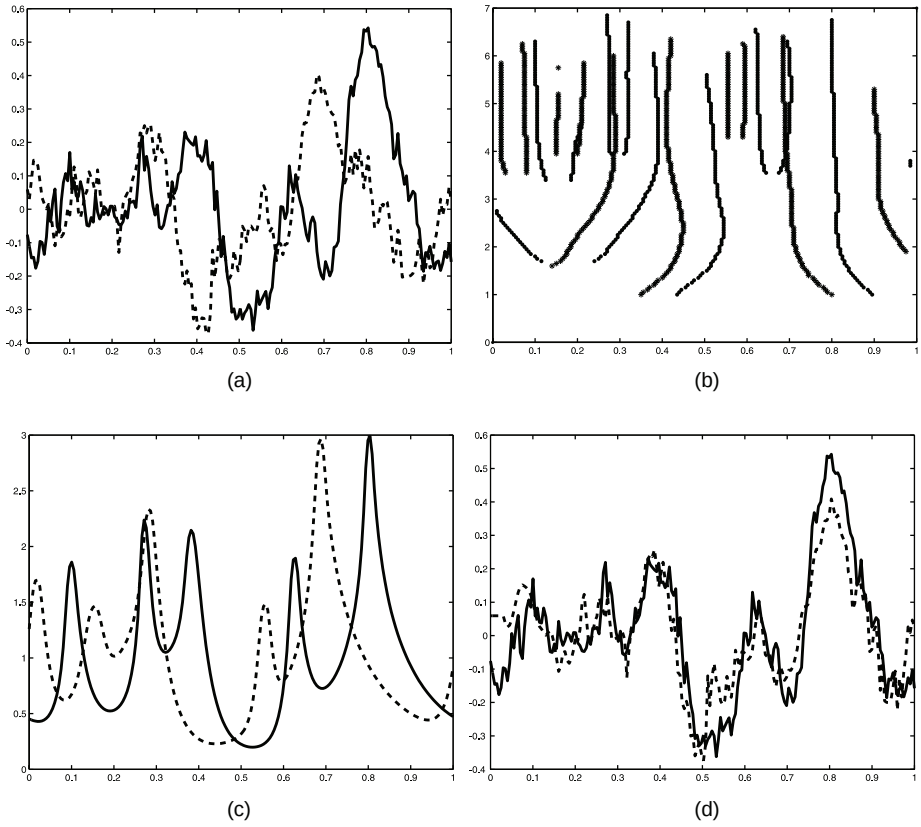

Figure 2. Fish sounds registration: (a) Noisy curves $\tilde{f}_1$ (solid curve) and $\tilde{f}_2$ (dashed curve). (b) Estimation of the zero-crossings for $r = 1$ of $\tilde{f}_1$ (black lines) and $\tilde{f}_2$ (gray lines): the vertical axis gives $-log_2(s)$. Structural intensities: (c) $\tilde{G}^+_{z1}$ and $\tilde{G}^+_{z2}$, (d) registration by dynamic correspondence

zero-crossings which correspond to the local maxima (respectively minima) of $\tilde{f}_1$ and $\tilde{f}_2$, respectively. These structural intensities are computed by eliminating either the positive or the negative zero-crossings in expression (2.7). One can see in Figure 2(c) that the modes of the structural intensities $\tilde{G}^+_{z1}$ and $\tilde{G}^+_{z2}$ correspond to the visually significant maxima of $\tilde{f}_1$ and $\tilde{f}_2$. This tool can thus be viewed as a smoothing method which highlights the main features of noisy observations.

2.3 Curve Registration via the Alignment of the Structural Intensities

One of the first issues encountered by landmark-based matching methods is the correspondence problem between two sets of features. This step is usually performed manually which can be tedious and prone to error. Let G_{z1} and G_{z2} denote the structural intensities of the (estimated) zero-crossings of two signals f_1 and f_2. Bigot (2003) observed that the features that one would align manually correspond to the modes of G_{z1} and G_{z2} whose shape and height are similar. To automatically solve this correspondence problem, Bigot (2003)

suggested aligning the structural intensities by DTW to measure this similarity in an automatic way. After registration by DTW, the modes of G_{z1} and G_{z2} that are at the same locations correspond to the landmarks that have to be aligned. However, DTW is time consuming, and one may wonder what are the advantages of using a DTW approach to align the structural intensities instead of directly registering the curves by DTW.

Usually the number of landmarks is small, and the functions G_{z1} and G_{z2} therefore have only a few modes. To solve the correspondence problem, we propose computing a piecewise linear warping function which aligns the common modes of G_{z1} and G_{z2}. Let $0 < x_1 < \cdots < x_{p_1} < 1$ and $0 < y_1 < \cdots < y_{p_2} < 1$ denote the locations of successive maxima of G_{z1} and G_{z2}, respectively. Let $\epsilon > 0$ be fixed. Let $I = \{i_1, \ldots, i_q\}$ (with $i_1 < \cdots < i_q$) and $J = \{j_1, \ldots, j_q\}$ (with $j_1 < \cdots < j_q$) be two subsets of cardinality $q \geq 1$ of $\{1, \ldots, p_1\}$ and $\{1, \ldots, p_2\}$, respectively, such that $|x_{i_k} - y_{j_k}| < \epsilon$ for $k = 1, \ldots, q$. For two such sets I and J, the piecewise linear warping function u_I^J is defined as the function $u_I^J : [0,1] \to [0,1]$ such that: u_I^J is continuous on $[0,1]$, u_I^J is linear on $[x_{i_k}, x_{i_{k+1}}]$ for $k = 0, \ldots, q$, and $u_I^J(x_{i_k}) = y_{j_k}$ for $k = 0, \ldots, q+1$, with $x_{i_0} = 0$, $y_{j_0} = 0$, $x_{i_{q+1}} = 1$ and $y_{j_{q+1}} = 1$. Then, the best piecewise linear warping function that aligns the modes of G_{z1} onto the modes of G_{z2} is defined as the function u_I^J which minimizes:

$$\int_0^1 (G_{z1}(x) - G_{z2} \circ u_I^J(x))^2 dx, \tag{2.8}$$

for all possible subsets I and J of cardinality $1 \leq q \leq \min(p_1, p_2)$. The minimizer of problem (2.8) can be efficiently computed by dynamic programming. Since the functions u_I^J are piecewise linear, the expression (2.8) can be easily evaluated for each pair of subsets I and J. Note that the cost of a dynamic programming algorithm to minimize (2.8) is $\mathcal{O}(p_1 p_2)$, which is very low because the number of landmarks is usually small. Once the minimizer $u_{I^*}^{J^*}$ of (2.8) has been found, a smoother warping function (at least C^1) is computed by the method of Ramsay (1998) to align the landmarks $(x_{i_k}, y_{j_k}), k = 0, \ldots, q^* + 1$. The parameter ϵ controls the distance between $u_{I^*}^{J^*}$ and the identity function. It avoids the computation of unrealistic warping as we shall see in Section 3, where we discuss the choice of ϵ and the calibration of the regularization parameter λ used for the monotone smoothing method of Ramsay (1998). In what follows we shall refer to this method as *dynamic correspondence*.

The registration of the raw data $\tilde{f}_1$ and $\tilde{f}_2$ by dynamic correspondence (with $\epsilon = 0.3$ and $\lambda = 10^{-3}$) between the modes of $\tilde{G}_{z1}^+$ and $\tilde{G}_{z2}^+$ is plotted in Figure 2(d). Note that after aligning the raw data, any smoothing method can be used to estimate these curves. One can see that the significant maxima of $\tilde{f}_1$ and $\tilde{f}_2$ are correctly aligned after registration. To match both the significant maxima and the significant minima of $\tilde{f}_1$ and $\tilde{f}_2$, one can compute a piecewise linear warping function u_I^J which minimizes the L^2 distance between $\tilde{G}_{z1}^+ - \tilde{G}_{z1}^-$ and $\tilde{G}_{z2}^+ \circ u_I^J - \tilde{G}_{z2}^- \circ u_I^J$. The warping function u_I^J is thus computed by aligning some modes of $\tilde{G}_{z1}^+$ onto some modes of $\tilde{G}_{z2}^+$, and some modes of $\tilde{G}_{z1}^-$ onto some modes of $\tilde{G}_{z2}^+$. In what follows, this procedure will be referred to as *dynamic correspondence MinMax*.

2.4 A SIMULATED DATA EXAMPLE

This section reports on a simulation study to evaluate the quality of the landmark-based method. We compare our methodology to the continuous matching approach of Ramsay and Li (1998), and with the performances of a DTW algorithm for the standard L^2 cost. The goal of this simulation study is to check if these methods can recover the "right" amount of warping between the simulated curves. In particular, it is important to check that if the original curves have no significant differences in time but only vary in amplitude, then the registration process does not introduce a bias in the sense that all the estimated warping functions are close to the identity.

We simulate $m = 100$ noisy signals sampled at $n = 150$ equally spaced points on $[0, 1]$, so that each curve in this dataset is a random deformation of one of the following functions (mixture of Gaussian) $h_1(t) = a_{1,1}g_1(t) + a_{1,2}g_2(t) + a_{1,3}g_3(t) + a_{1,4}g_4(t)$, $h_2(t) = a_{2,2}g_2(t) + a_{2,3}g_3(t) + a_{2,4}g_4(t)$, and $h_3(t) = a_{3,1}g_1(t) + a_{3,2}g_2(t) + a_{3,3}g_3(t)$, where $g_k(t) = \exp\left(-\frac{(t-x_k)^2}{2\sigma_k^2}\right)$, $t \in [0, 1]$, with $0.5 < x_1 < x_2 < x_3 < x_4 < 0.95$ and $|x_k - x_{k'}| \geq 0.2$ for $k \neq k'$. The main features of these curves are the locations and the amplitudes of their local extrema. Then, we simulate m signals $f_1, \ldots, f_m$ by perturbing the locations and the amplitudes of the maxima of the curves $h_l, l = 1, 2, 3$ as follows: for $j = 1, \ldots, m$ repeat the procedure: select randomly an integer l between 1 and 3, if $l = 1$, compute $h_j^*(t) = A_{1,1}g_1(t) + A_{1,2}g_2(t) + A_{1,3}g_3(t) + A_{1,4}g_4(t)$, else if $l = 2$, compute $h_j^*(t) = A_{2,2}g_2(t) + A_{2,3}g_3(t) + A_{2,4}g_4(t)$, else if $l = 3$, compute $h_j^*(t) = A_{3,1}g_1(t) + A_{3,2}g_2(t) + A_{3,3}g_3(t)$, where $A_{l,k} \sim U([a_{l,k} - \delta_1, a_{l,k} + \delta_1])$ (uniform random variable on an interval). Then, we randomly generate a smooth warping function and we compute the test function $f_j = h_j^* \circ u_j$.

Noisy observations are then obtained by adding Gaussian noise to the curves $f_j, j = 1, \ldots, m$ according to the model (2.1). A sub-sample of 10 noisy curves is displayed in Figure 3.

To register this set of $m = 100$ noisy curves, we compare our landmark-based matching method via the alignment of the estimated maxima (abbreviated as LAND_Max), our

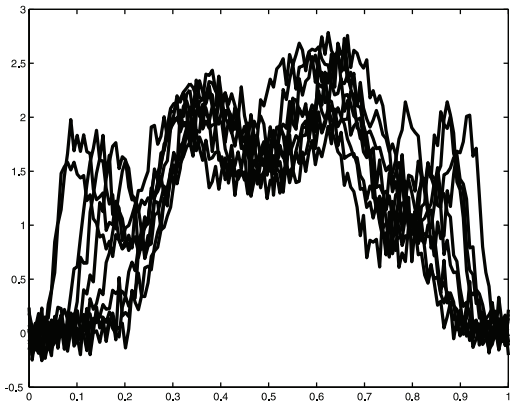

Figure 3. A subsample of 10 noisy curves from the simulated dataset.

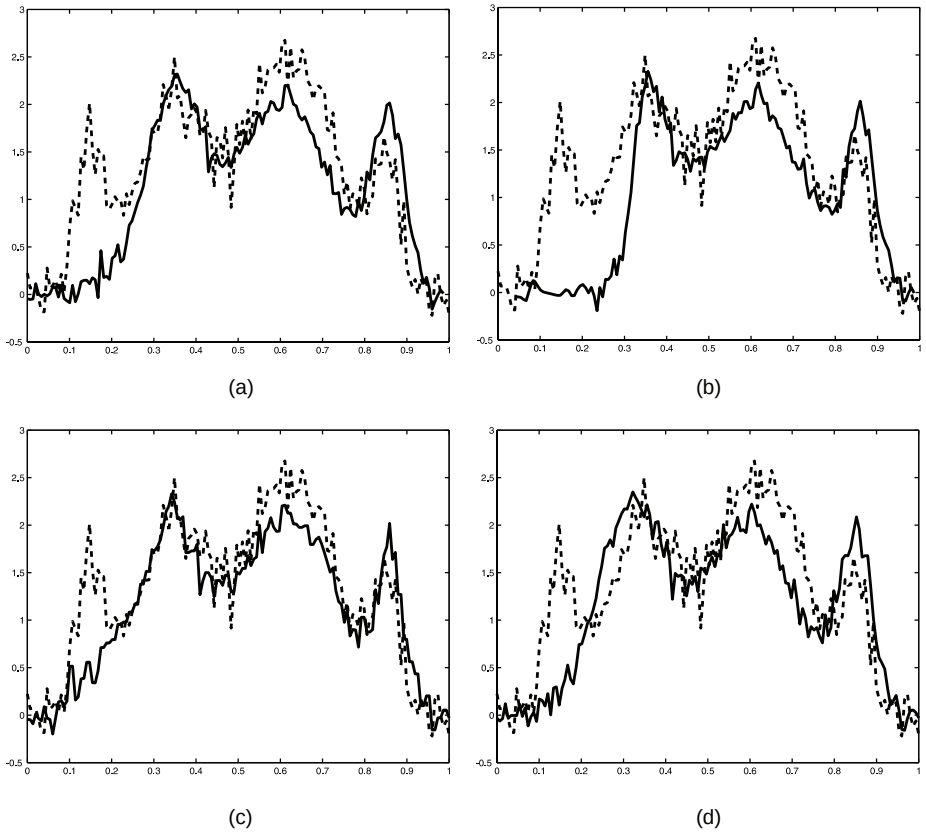

Figure 4. Reference curve h_e is plotted as a dashed line and the solid lines correspond to some registered curve $\tilde{f}$: (a) by LAND_Max, (b) by LAND_MinMax, (c) by DTW, (d) by MINEIG.

landmark-based matching method via the alignment of the estimated maxima and minima (abbreviated as LAND_MinMax), the continuous matching approach of Ramsay and Li (1998), abbreviated as MINEIG, and a DTW algorithm using the empirical L^2 cost between two regression functions. For LAND_Max and LAND_MinMax, we took $\epsilon = 0.2$ (if the distance between two landmarks is larger than 0.2, then they cannot be aligned). We also chose to perform an *exact* landmark registration by taking a very small value for the smoothing parameter in the algorithm proposed by Ramsay (1998). The criterion MINEIG is a powerful alternative to the standard L^2 cost for matching two functions, which produces much more satisfactory results (see Ramsay and Li 1998; Ramsay and Silverman 2002), and which is implemented in the routine `registerfd` available at *http://www.psych.mcgill.ca/faculty/ramsay/ramsay.html*. To compare these methods, we have chosen to register our simulated data to the *same* noisy curve h_e which is the sample of the curve h_1 (at the design points) corrupted by a Gaussian noise. The sampled curve h_e is displayed in Figure 4 as a dashed curve. For all the methods, we present results for the registered raw data onto h_e.

Figure 4 gives an example of the alignment of a curve $\tilde{f}$ with three significant maxima

Table 1. $\boldsymbol{L}^2$ Distance Between the Structural Mean $\boldsymbol{h}^*$ and the Pointwise Mean of the Registered Curves for Each Registration Method

	LAND_Max	***LAND_MinMax***	***DTW***	***MINEIG***
$\boldsymbol{L}^2$ distance	0.0064	0.0269	0.0140	0.0134

onto the reference curve h_e for each method. For LAND_Max, the three significant maxima of $\tilde{f}$ are correctly aligned onto the corresponding maxima of h_e. The left part of $\tilde{f}$ (for $t < 0.3$) is left unchanged since no landmark is found for $\tilde{f}$ in the interval $[0, 0.3]$. For LAND_MinMax, the three significant maxima of $\tilde{f}$ are also correctly aligned onto the corresponding maxima of h_e. However, this method also aligns the local minima of $\tilde{f}$ around $t = 0.1$ onto the local minima of h_e around $t = 0.23$, which changes the shape of the registered curve $\tilde{f}$ in the interval $[0, 0.3]$. This is one drawback of our method: if ϵ is too large, then LAND_MinMax may align two landmarks that would have not been associated manually (for this example, this problem can be avoided by choosing $\epsilon \leq 0.1$). DTW also yields a satisfactory alignment of the significant maxima of both curves. Because DTW tends to reduce the L^2 distance between the two curves, the warping modifies the amplitude of $\tilde{f}$ in the interval $[0, 0.3]$, but the global shape of the curve in this interval is preserved by the registration. For MINEIG, one can see that the common maxima of $\tilde{f}$ and h_e are also correctly aligned after registration. The criterion MINEIG also modifies locally the shape of the registered curve $\tilde{f}$ in the interval $[0, 0.3]$. However in the interval $[0.2, 0.3]$, MINEIG preserves the amplitude difference between the two curves, and does not warp the curve $\tilde{f}$ to make it look exactly like h_e as it is the case for DTW.

Let $h^* = \frac{1}{m}\sum_{j=1}^{m} h_j^*$ be the pointwise mean of the curves h_j^* (called the structural mean in what follows). Table 1 gives the empirical L^2 distance between the structural mean h^* and the pointwise mean of the registered curves for each registration method. LAND_Max gives the best result as it yields an estimate which is very close to the structural mean. For LAND_MinMax, the quality of the estimation is not so satisfactory since this method introduces a bias in the interval $[0.1, 0.35]$ which is due to the misalignment of some local minima as it has been observed in Figure 4(b). For DTW, the amplitude of the estimated mean is highly oscillating and is biased on some parts of $[0, 1]$ which explains its poor performances. LAND_Max also clearly outperforms MINEIG.

To compare these methods, we have also computed the L^2 distance between the amplitude variability of the curves h_j^* defined as: $\frac{1}{m}\sum_{j=1}^{m}(h_j^*(t_i) - h^*(t_i))^2$, and the estimated pointwise variability of the registered data defined as: $\frac{1}{m}\sum_{j=1}^{m}(\hat{h}_j^*(t_i) - h^*(t_i))^2$, for each registration method, where $\hat{h}_j^*(t_i), i = 1, \ldots, n$ denotes the registered raw data for a given method. The results given in Table 2 confirm the superiority of LAND_Max since the estimated variability is close to the true pointwise variability. This shows that LAND_Max retains the "right" amount of amplitude variability between the curves.

Finally, it is important to check if these methods do not introduce a bias when the data have no significant differences in time. In this case, all the estimated warping functions should be close to the identity function. Thus, we have simulated another dataset by adding

Table 2. L^2 Distance Between the Amplitude Variability of the Curves $h_j^*, j = 1, \ldots, M$ and the Estimated Pointwise Variability of the Registered Data for Each Method

	LAND_Max	LAND_MinMax	DTW	MINEIG
L^2 distance	0.0030	0.0319	0.0103	0.0084

Gaussian noise to the sample curves h_j^*. All these curves do not necessarily have the same number of peaks, but their common landmarks are at the same locations. Therefore, a registration method applied to this dataset should leave the curves unchanged. If $\hat{u}_j^{-1}$ denotes the estimated warping function used to align the jth noisy curve h_j^* onto the noisy reference curve h_e (for a given registration method), then the average bias eventually introduced can be measured by $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m}|\hat{u}_j^{-1}(t_i) - t_i|$. The average bias of each registration method is given in Table 3. For LAND_Max and MINEIG, we see that the estimated warping functions are very close to the identity, since the average bias is less than 0.0058 for MINEIG and less than 0.005 for LAND_Max. These results tend to show that LAND_Max and MINEIG yield a good estimation of the "right" amount of time warping within a set of curves. Again, the results are not so satisfactory for DTW and LAND_MinMax.

This simulation study tends to confirm the superiority of our registration method over DTW and MINEIG if we use local maxima as landmarks for the curves. Indeed, our approach gives a very good alignment of the common peaks of the curves which yields a very satisfactory estimation of the structural mean. It can recover the right amount of time variability within a set of curves, and it preserves the differences in amplitude between the registered curves. Finally, it does not introduce a bias if the data have no differences in time.

Note that we should mention that DTW and MINEIG have been used without preliminary smoothing the data. This step could certainly improve the estimation of the warping functions as long as the smoothing preserves the main features of the curves, but this possibility has not been investigated here. A more detailed study of this simulated data example is available at: *http://www.lsp.ups-tlse.fr/Fp/Bigot/pub_en.html*

This document also contains the values of the various parameters that have been used to create this simulated data example.

2.5 Alignment of a Set of Curves

Let $f_1, \ldots, f_m$ be a set of curves eventually observed with noise ($m > 2$). For the simulated datasets of the previous section, all the curves have been registered onto the same reference curve. However, it is generally difficult to choose such a reference function. Indeed,

Table 3. Average Bias of the Warping Functions for Each Method

	LAND_Max	LAND_MinMax	DTW	MINEIG
Average bias	0.0050	0.0166	0.0158	0.0058

this curve cannot be randomly chosen, since we are not guaranteed that this yields a good estimate of the typical curves in the dataset. To synchronize a set of curves by landmark-based matching, we have therefore investigated two approaches (inspired by Wang and Gasser 1997):

1. Take each curve f_i as a reference curve. For each curve $f_j, j \neq i$, determine the vector $\tau_{i,j}$ that contains the landmarks of f_i that must be aligned onto the landmarks of f_j. Let $\tau_i = \cap_{j=1,j\neq i}^{m} \tau_{i,j}$ be the landmarks of f_i that are "common" to all curves. Let $\tau_{j,i}^*$ be the vector containing the landmarks of f_j that correspond to the landmarks in τ_i. Let f_{i_0} be the curve that maximizes the length of $\tau_i, i = 1, \ldots, m$, and define τ^* as the average by component of the vectors $\tau_{j,i_0}^*, j = 1, \ldots, m$ (with $\tau_{i_0,i_0}^* = \tau_{i_0}$). If various vectors τ_{i_0} of maximum length M_0 can be chosen, take the one that maximizes the "variance": $\sum_{j=1}^{m} \sum_{k=1}^{M_0} (\tau_{j,i_0}^*[k] - \tau^*[k])^2$. The warping functions $u_1, \ldots, u_m$ are then computed such that $u_j(\tau^*[k]) \approx \tau_{j,i_0}[k]$ for $k = 1, \ldots, M_0$.

2. If m is large, select randomly a subset of curves of size 2^j. Then, partition this subset into 2^{j-1} pairs. For each pair of functions f_1, f_2, find the two sets of landmarks τ_1 and τ_2 that should correspond and calculate the average by component τ^* of these two vectors. Compute the warping function u_i that maps τ^* onto τ_i, $i = 1, 2$, and calculate the mean $(f_1 \circ u_1 + f_2 \circ u_2)/2$. In this fashion, we obtain 2^{j-1} mean functions. Partition this group into 2^{j-2} pairs and compute again a mean function for each pair. Repeat the procedure until obtaining one mean function that is taken as a reference curve f_e on which we align the curves $f_i, i = 1, \ldots, m$.

The main drawback of Method 1 is its computational cost when m is large. Hence, if m is large it is better to use Method 2, otherwise Method 1 must be chosen.

3. SYNCHRONIZING SAMPLES OF CURVES AND FANOVA

3.1 ALIGNMENT OF THE GANGLIOSIDE DATASET

The aim of the study carried out by Munoz Maldonado et al. (2002) was to better understand the physical process of aging that could explain changes in performance via the comparison of the ganglioside distribution in brain tissues between old and young rats. The ganglioside distribution, represented by a density profile, has been extracted by thin layer chromatography (TLC) in three regions of the brain, the medulla (MD), the locus (LC), and the hippocampus (HY), for five old rats (12 months old) and five young rats (2 months old). Further details on how these data were recorded can be found in Munoz Maldonado et al. (2002). After removing the background noise, the following model was proposed by Munoz Maldonado et al. (2002) for the curves that correspond to a region of the brain: each density profile is a function $Y_{ij}(t)$ observed for $t \in [a_{ij}, b_{ij}]$, where $i = 1, 2$ is the population index and $j = 1, \ldots, 5$ is the subject index within a group, such that

$$Y_{ij}(t) = f_{ij} \circ h_{ij}(t) + \sigma_{ij}\epsilon_{ij}(t), \tag{3.1}$$

where the f_{ij} are the unknown density profiles with population mean f_i, the h_{ij} $(i = 1, 2; j = 1, \ldots, 5)$ are iid random, strictly increasing and C^1 functions, and $u_{ij} = h_{ij}^{-1}$:

$[0, T] \rightarrow [0, T_{ij}]$ are called the warping functions. As explained by Munoz Maldonado et al. (2002), the warping functions model the capillary action across the silica gel plate used for TLC and contain therefore very little information about differences between populations. Hence, the density profiles must be registered before performing any statistical comparison. It was also indicated by Munoz Maldonado et al. (2002) that the amount of the component ganglioside is related to the *area* under a peak of the density profile rather than to the height of a peak. Therefore, to preserve the area under a peak, the registered density profiles in Munoz Maldonado et al. (2002) are given by $\hat{u}'_{ij}(t)Y_{ij} \circ \hat{u}_{ij}(t)$ and not by $Y_{ij} \circ \hat{u}_{ij}(t)$. However, this area-preserving transformation can significantly change the shape of the curves if the warping functions are far from the identity. Thus, we prefer to analyze the "standard" registered raw data $Y_{ij} \circ \hat{u}_{ij}(t)$, because this transformation is shape preserving and eases the interpretation of the registration step. After synchronizing the curves, the aim of this study is to compare the densities f_1 and f_2 (the population means of the density profiles) for each region of the brain (MD, LC, and HY).

Each density profile is a sequence of observations whose size varies between 74 and 155 points. For each region of the brain, we chose to interpolate the samples curves Y_{ij} with cubic splines to form sequences of observations of $n = 128$ points denoted by $\bar{Y}_{ij}$. The number of points is chosen to avoid smoothing, so that the fitted curves agree with the shape of the original data. Then, the fitted curves $\bar{Y}_{ij}$ are linearly rescaled on the unit interval $[0, 1]$ by defining $\tilde{Y}_{ij}(t) = \bar{Y}_{ij}(a_{ij} + (b_{ij} - a_{ij})t), t \in [0, 1]$. To compare the density profiles, we register the curves $\tilde{Y}_{ij}(t)$ $(t \in [0, 1],\ j = 1, \ldots, 5)$ for each subgroup $i = 1, 2$. These rescaled density profiles for old and young rats in the region MD are plotted in Figure 5 (for reasons of space we do not plot the density profiles in the regions LC and HY).

The main characteristics of the density profiles are the locations of the peaks which represent the various components of the ganglioside. We have therefore chosen to register the curves by aligning their significant maxima. For each region of the brain and for each group, the number of curves is relatively small (only 5), and the registration Method 1 with dynamic correspondence can thus be used (see Section 2.5). We took $\epsilon = 0.3$ to allow a relatively large amount of warping for the peaks. The significant maxima of the density profiles are detected via the structural intensity of the significant zero-crossings computed for a Gaussian wavelet with $r = 1$ vanishing moment. For this dataset the landmarks are the local maxima of the density profiles, so the negative zero-crossings are not used to compute the structural intensities. For all curves, we found that the modes of the structural intensities are exactly located at the significant peaks of the density profiles. Moreover, for all curves, only four or five of these peaks are used as common landmarks for the registration. Because these landmarks are the main features of the curves, and given that Munoz Maldonado et al. (2002) argued that the warping functions contain very little information about differences between populations, we chose to perform an exact landmark registration by choosing a very small value for the regularization parameter ($\lambda = 10^{-3}$) used in the algorithm of Ramsay (1998). The registered density profiles for the region MD are displayed in Figure 5. One can see that the maxima that are common to a set of curves are correctly aligned after registration. Analysis of the results shows that the quality of the alignment is particularly satisfactory for the density profiles of the regions MD and HY, and for the old rats in

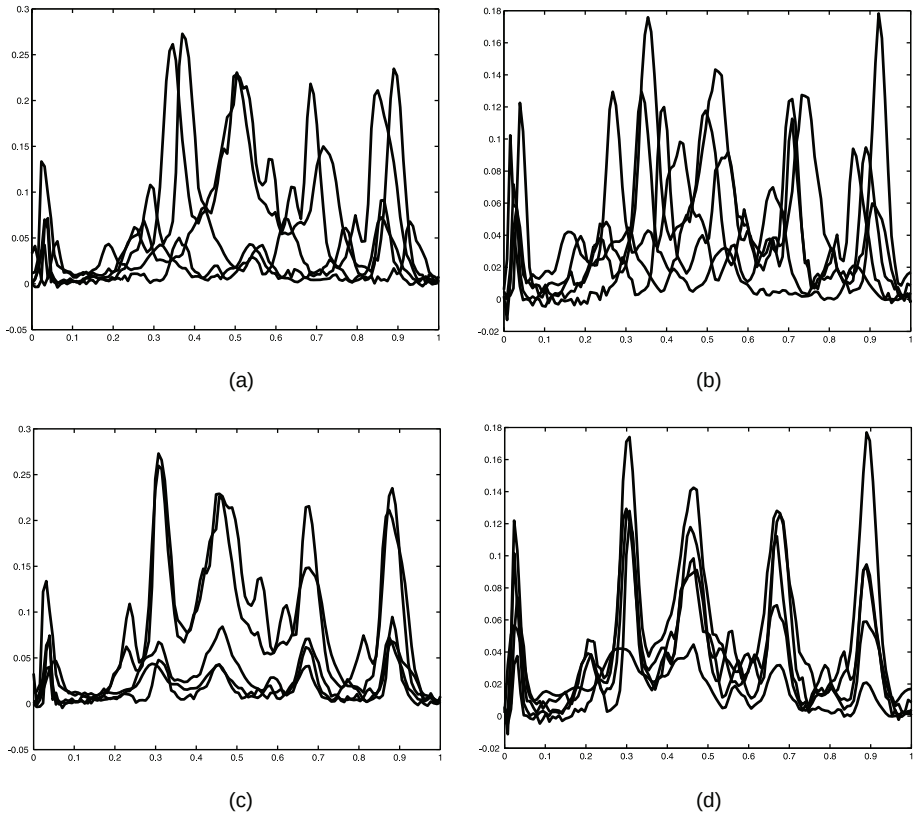

Figure 5. Density profiles of the ganglioside in the region MD. Raw data: (a) five old rats, (b) five young rats. Registered raw data: (c) five old rats, (d) five young rats.

the region LC. The alignment for the curves of the young rats in the region LC is not so satisfactory. This is mainly due to the fact that these density profiles do not have the same number of significant maxima. To improve the quality of the registration, some inflection points were also used as landmarks by Munoz Maldonado et al. (2002), but this possibility has not been investigated here.

Recall that for the ganglioside data, one needs to decide if the shape of the density profiles shows variations with aging for each region of the brain. This can be done by quantifying the differences between the empirical population means f_1 and f_2 (see Section 3.1) that are depicted in Figure 6 (obtained by averaging the registered density profiles). The statistical analysis carried out by Munoz Maldonado et al. (2002) provides significant evidence of difference in ganglioside distribution with aging in the regions MD and LC but not in the region HY. As one can see in Figure 6, the population means in the region LC are significantly different. For the region MD, we see that the two population means have a similar shape. These two curves are mainly composed of five significant peaks at the same locations. However, these peaks are more intense for the old rats. This difference in amplitude tends to confirm that the distribution of the ganglioside varies with aging in the

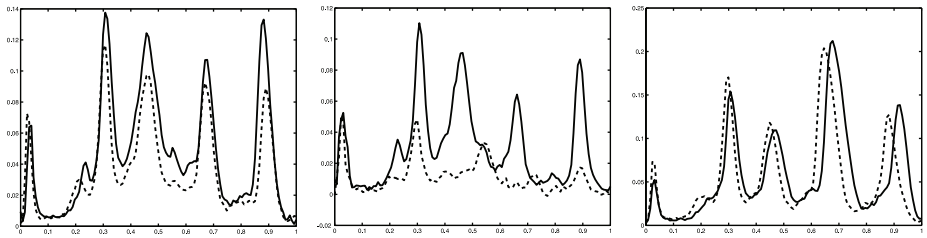

Figure 6. Population means of the registered density profiles for young rats (dashed lines) and old rats (solid lines) for each region of the brain: (a) MD, (b) LC, (c) HY.

region MD. For the region HY, the population mean have the same shape and amplitude. They are also composed of five peaks, but the locations of these maxima clearly vary in time. Hence, up to a time transformation, the population means for the region HY show very little differences which would confirm the conclusions in Munoz Maldonado et al. (2002). This time shift can be explained by the variability of the capillary action across the silica gel plate used for TLC which is not related to differences between the two populations [it mainly depends on the experimental conditions; see Munoz Maldonado et al. (2002) for further details].

3.2 ALIGNMENT OF THE ORTHOSIS DATASET

The purpose of recording the orthosis dataset is to better understand the processes underlying movement generation under various levels of an external force applied to the knee. Seven young male volunteers wore a spring-loaded orthosis of adjustable stiffness under four experimental conditions: a "Control" condition (without orthosis), an "Orthosis" condition (with the orthosis only), and two conditions ("Spring 1," "Spring 2"), in which the movement is perturbated by fitting a spring-loaded orthosis onto the right knee joint. Under each experimental condition, 10 trials of 20 seconds were recorded for each subject. For a stepping-in-place task, the resultant moment at the knee is derived by means of body segment kinematics recorded with a sampling frequency of 200 Hz. Further details on how these data were recorded and computed can be found in Cahouet, Martin, and Amarantini (2002). For each trial, the resultant moment was computed at 256 time points equally spaced and scaled so that a time interval corresponds to an individual gait cycle. Hence, a typical observation is a one-dimensional function of time $t \in [0, 1]$ and the dataset consists of 280 separate curves (10 trials replicated for the seven subjects under the four experimental conditions). For reasons of space we do not display the whole orthosis dataset; see Abramovich et al. (2004) for such a plot.

For each subject and each condition, we have observed that the curves which correspond to the 10 trials have a common structural pattern, and have approximately the same number of significant extrema. To compare these curves it is necessary to eliminate these variations. We have chosen to register the curves for each subject and each experimental condition by aligning their significant extrema. Given that 28 subsets of 10 curves must be synchronized,

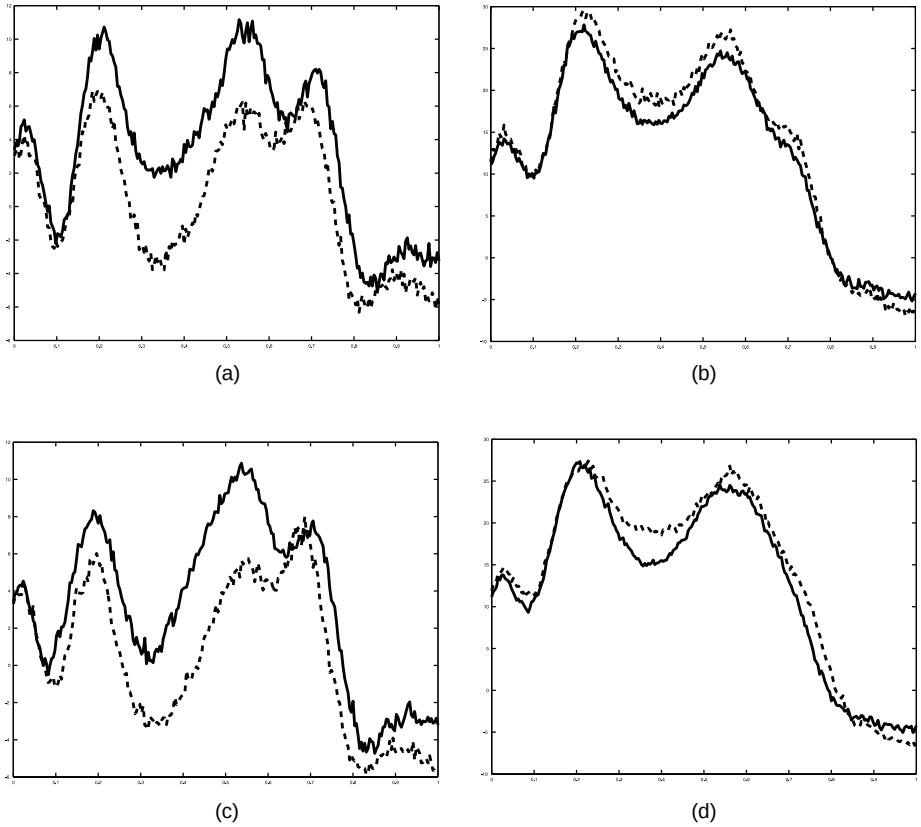

Figure 7. Population mean for the orthosis data: (a) original data under the "Control" (solid line) and "Orthosis" (dashed line) conditions, (b) original data under the "Spring 1" (solid line) and "Spring 2" (dashed line) conditions, (c) registered data under the "Control" (solid line) and "Orthosis" (dashed line) conditions, (d) registered data under the "Spring 1" (solid line) and "Spring 2" (dashed line) conditions.

the registration Method 2 with dynamic correspondence MinMax has been chosen (see Section 2.5). Since a visual inspection of this dataset shows that the amount of time-warping between the curves is a priori small, we took $\epsilon = 0.2$. For all curves, the significant local extrema are correctly detected via the structural intensities of the significant zero-crossing. Since these local extrema are the main features of the curves, we chose to perform an exact landmark registration by choosing again a small value for the regularization parameter ($\lambda = 10^{-3}$) in the algorithm proposed by Ramsay (1998). After registration, we have observed that the common local extrema of the curves are correctly aligned.

The purpose of the analysis of the orthosis dataset is to better understand how a subject behaves under an external perturbation. Obviously, the subjects do not all behave similarly but the researchers who have recorded these data are not interested in quantifying the differences between the subjects. Hence, to determine if the subjects behave similarly under various external perturbations, we propose to quantify the differences between the population means under each experimental condition. For the original and registered data, the

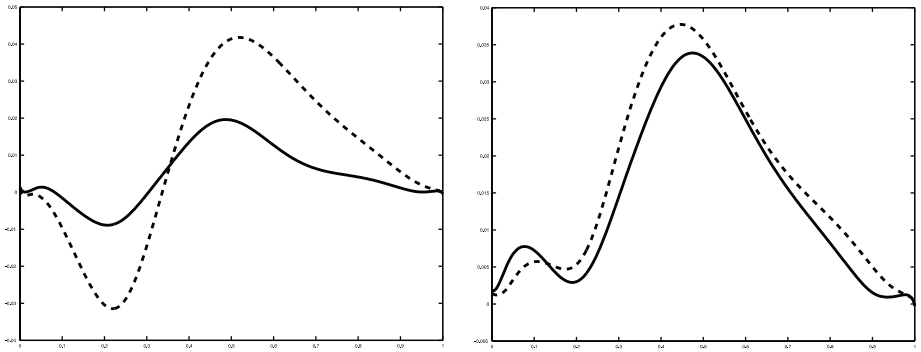

Figure 8. Average time deformation functions under: (a) the "Control" (solid line) and "Orthosis" (dashed line) conditions, (b) the "Spring 1" (solid line) and "Spring 2" (dashed line) conditions.

population means under each experimental condition are displayed in Figure 7. One can see that the synchronization of this dataset modifies locally the amplitude of the population means, but that the shape of these curves is globally preserved by the registration. We also see in Figure 7 that the subjects seem to behave similarly under the conditions "Control" and "Orthosis," or under the conditions "Spring 1" and "Spring 2." The registration step tends to increase the amplitude variation between the population means, but it is difficult to determine from these plots if the curves under the conditions "Control" and "Orthosis," or under the conditions "Spring 1" and "Spring 2" are significantly different. Thus, to quantify the variations within this set of curves in a more automatic way, we propose a fixed-effect FANOVA model in the next section.

To compare the subjects in terms of time variability we have aligned each curve of the original dataset onto the global mean of the registered data (denoted as f_e in what follows). For each experimental condition, we have computed the average time deformation function defined as $\tilde{u}(t) = \frac{1}{\#C}\sum_{j\in C}(\hat{u}_j(t) - t),\ t \in [0,1]$, where the set $\{\hat{u}_j\}_{j\in C}$ denotes the estimated warping functions used to align the original curves onto f_e for a given condition C. Negative values of $\tilde{u}$ indicate an earlier reaction of the muscular action of the subjects with respect to f_e, while positive values correspond to a slower reaction. These average time deformation functions are displayed in Figure 8. Again, these curves tend to show that the subjects behave similarly under similar perturbations. We see that the amplitude of the average time deformation function under the "Spring 2" condition is slightly larger than the one under the condition "Spring 1." This indicates that the muscular reaction of the subjects is faster under the condition "Spring 1." For the "Control" and "Orthosis" conditions, the average time deformation functions have the same shape but differ in amplitude. The dashed curve in Figure 8(a) shows early reaction of the muscular action under the "Orthosis" condition before $t = 0.35$ and then a slower reaction. Thus, it seems that an external perturbation may introduce a phase variation (in time) of the muscular activity.

3.3 A FIXED-EFFECTS FANOVA MODEL

Due to the size of the data that can be collected by modern recording equipments, multivariate ANOVA techniques which treat functional data as multivariate vectors are usually not practical [especially due to the "curse of dimensionality;" see Fan and Li (1998) and Abramovich et al. (2004) for a discussion on the drawbacks of these approaches]. For the comparison of multiple sets of signals, *functional analysis of variance* (FANOVA) methods provide powerful alternatives to classical ANOVA techniques [see Ramsay and Silverman (2005) and Stone, Hansen, Kooperberg, and Truong (1997) for detailed reviews of these techniques]. Abramovich et al. (2004) proposed a fixed-effects FANOVA model, and derived statistical procedures based on wavelet decompositions to test if the main or interaction effects are zero.

Recall that for the ganglioside dataset, we need to measure the similarity between the population means f_1 and f_2 for each region of the brain, and that for the orthosis dataset we would like to know if the subjects behave similarly under various experimental conditions. Now that we have represented these two sets of curves in appropriate referentials for their comparison, we can use a statistical model to quantify the differences between two subsets of subjects. As an illustration, Abramovich et al. (2004) applied their tests to the analysis of the orthosis dataset, but *without pre-registration of the data*. They have concluded that under the conditions "Control" and "Orthosis," or under the conditions "Spring 1" and "Spring 2," the subjects behave similarly. This would therefore indicate that the subjects adjust their posture similarly under similar perturbations. But if the data are preregistered for each subject and each condition, then by applying the same hypothesis tests, we have concluded to the contrary that the individuals behave differently under the conditions "Control" and "Orthosis", or under the conditions "Spring 1" and "Spring 2". This shows that a registration step can change substantially the results of the subsequent testing problem.

To analyze the ganglioside data, we took for each region of the brain the same FANOVA model as the one used for the orthosis data but with only two populations (or conditions). For the three regions—MD, LC, and HY—the null hypothesis (no differences in effects between the two populations) is clearly rejected. Note that this could be directly inferred by looking at the data in Figure 6 which shows that the level of noise is extremely small and that all the curves are therefore different in L^2 norm. Hence, contrary to the conclusions in Munoz Maldonado et al. (2002), this FANOVA model indicates that the distribution of the ganglioside varies with aging in the three regions MD, LC, and HY.

4. CONCLUSION

An important problem in landmark-based registration is the correspondence problem between two sets of features. This article proposed a fast and automatic method to align the significant landmarks of a set of noisy signals. We justified our approach with simulations and showed that our wavelet-based method can improve the results obtained by existing procedures for curve registration. Our procedure was also successfully applied to various real examples and is an effective technique to synchronize a set of curves.

For the ganglioside and orthosis datasets, we showed that the analysis of the warping functions together with their relationship to the registered curves can explain the differences or the similarities between two populations. This landmark-based matching approach has also been combined with the functional hypothesis testing procedures recently developed by Abramovich et al. (2004) to illustrate the usefulness of curve registration for the analysis of functional data. Our results show that curve alignment may influence the conclusions obtained from a FANOVA model. For the orthosis dataset, we derived results opposite to the conclusions in Abramovich et al. (2004) if the curves are preregistered.

For any registration method, some smoothing parameters control the tradeoff between the amount of warping and leaving enough time variability between the curves. For the correspondence problem between two sets of landmarks, a parameter ϵ has been introduced to avoid unrealistic alignment between two landmarks. Once this correspondence is solved, there is still a regularization parameter λ that controls the tradeoff between exact and inexact landmark matching. In this article, the values of ϵ and λ are chosen subjectively which is often the case for curve registration problems (see, e.g., Ramsay and Silverman 2002). This choice can be made by visual inspection of the synchronized curves and depends on the quality of the estimation of the landmarks. However, it would be nice to design a data-based method to automatically calibrate these parameters.

Finally, we would like to indicate that the notion of structural intensity is also an efficient technique to extract the singularities of an unknown signal via the estimation of the wavelet maxima lines of its continuous wavelet transform (see Bigot 2005). In this case, the amplitude of the modes in the structural intensity is related to the order of the singularities of the signal. Hence, we believe that the structural intensity could also be used to register a set of curves by aligning their significant singularities (e.g., jump points).

4.1 SOFTWARE

LandAlignWave is a library of MATLAB procedures allowing the use of the methods developed in this article that is available for download from: *http://www.lsp.ups-tlse.fr/Fp/Bigot/Soft/landalign.html*

ACKNOWLEDGMENTS

This research was supported by funds from the project MADEMO (Project Région Rhône-Alpes) and the project AMOA (Project IMAG). Financial support from the IAP research network nr P5/24 of the Belgian Government (Federal Office for Scientific, Technical and Cultural Affairs) is also gratefully acknowledged. I am very grateful to Y. Munoz Maldonado for providing the ganglioside dataset, to D. Amarantini for providing the orthosis dataset and to G. Horgan for providing the fish sound data. I particularly thank A. Antoniadis for helpful comments and for pointing out some references. I also thank C. Glasbey for helpful discussions and excellent hospitality whilst at BIOSS (Edinburgh, UK). I am also very grateful to the associate editor and to three anonymous referees for helpful comments which have substantially improved the presentation of this article.

REFERENCES

Abramovich, F., Antoniadis, A., Sapatinas, T., and Vidakovic, B. (2004), "Optimal Testing in a Fixed-effects Functional Analysis of Variance Model," *to appear in International Journal of Wavelets, Multiresolution and Information Processing*.

Bigot, J. (2003), "Automatic Landmark Registration of 1D Curves," in *Recent Advances and Trends in Nonparametric Statistics*, eds. M. Akritas and D.N. Politis, Amsterdam: Elsevier, pp. 479–496.

——— (2005), "A Scale Space Approach with Wavelets to Singularity Estimation," *ESAIM: Probability and Statistics*, 9, 143–164.

Brown, L.D., and Low, M. G. (1996), "Asymptotic Equivalence of Nonparametric Regression and White Noise," *The Annals of Statistics*, 3, 2384–2398.

Cahouet, V., Martin, L., and Amarantini, D. (2002), "Static Optimal Estimation of Joint Accelerations for Inverse Dynamic Problem Solution," *Journal of Biomechanics*, 35, 1507–1513.

Chaudhuri, P., and Marron, J.S. (1999), "SiZer for Exploration of Structures in Curves," *Journal of the American Statistical Association*, 94, 807–823.

——— (2000), "Scale Space View of Curve Estimation," *The Annals of Statistics*, 28, 408–428.

Donoho, D.L., and Johnstone, I. M. (1998), "Minimax Estimation via Wavelet Shrinkage," *The Annals Statistics*, 26, 879–921.

——— (1999), "Asymptotic Minimality of Wavelet Estimators with Sampled Data," *Statistica Sinica*, 9, 1–32.

Fan, J., and Li, S K. (1998), "Test of Significance When Data are Curves," *Journal of the American Statistical Association*, 91, 674–688.

Fisher, N.I., and Marron, J. S. (2001), "Mode Testing via the Excess Mass Estimate," *Biometrika*, 88, 499–517.

Gasser, T., and Kneip, A. (1995), "Searching for Structure in Curve Samples," *Journal of the American Statistical Association*, 90, 1179–1188.

Gervini, D., and Gasser, T. (2004), "Self-Modeling Warping Functions," *Journal of the Royal Statistical Society*, Series B, 66, 959–971.

Hummel, B., and Moniot, R. (1989), "Reconstruction from Zero-Crossings in Scale-Space," *IEEE Transaction on Acoustic, Speech and Signal Processing*, 37, 2111–2130.

Kneip, A., and Gasser, T. (1992), "Statistical Tools to Analyze Data Representing a Sample of Curves," *The Annals of Statistics*, 20, 1266–1305.

Liu, X., and Müller, H. G. (2004), "Functional Convex Averaging and Synchronization for Time-Warped Random Curves," *Journal of the American Statistical Association*, 99, 687–699.

Mallat, S. (1999), *A Wavelet Tour of Signal Processing* (2nd ed.), New York: Academic Press.

Mallat, S., and Hwang, W.L. (1992), "Singularity Detection and Processing with Wavelets," *IEEE Transactions on Information Theory*, 38, 617–643.

Munoz Maldonado, Y., Staniswalis, J.G., Irwin, L.N., and Byers, D. (2002), "A Similarity Analysis of Curves," *The Canadian Journal of Statistics*, 30, 373–381.

Ramsay, J.O. (1998), "Estimating Smooth Monotone Functions," *Journal of the Royal Statistical Society*, Series B, 60, 365–375.

Ramsay, J.O., and Li, X. (1998), "Curve Registration," *Journal of the Royal Statistical Society*, Series B, 60, 351–363.

Ramsay, J.O., and Silverman, B. W. (2002), *Applied Functional Data Analysis*, New York: Springer Verlag.

——— (2005), *Functional Data Analysis* (2nd ed.), New York: Springer Verlag.

Sakoe, H., and Chiba, S. (1978), "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, 26, 43–49.

Stone, C. J., Hansen, M., Kooperberg, C., and Truong, Y. (1997), "Polynomial Splines and their Tensor Products in Extended Linear Modeling" (with discussion), *The Annals of Statistics*, 25, 1371–1470.

Trouve, A., and Younes, L. (2000), "Diffeomorphic Matching Problems in One Dimension: Designing and Minimizing Matching Functionals," in *Proceedings ECCV 2000*.

Wang, K., and Gasser, T. (1997), "Alignment of Curves by Dynamic Time Warping," *The Annals of Statistics*, 25, 1251–1276.

——— (1999), "Synchronizing Sample Curves Nonparametrically," *The Annals of Statistics*, 27, 439–460.

Wood, M. (2002), "Discriminant Analysis using Wavelet Derived Features," PhD Thesis, University of Aberdeen.

Younes, L. (2000), "Deformations, Warping and Object Comparison," *Tutorial*. Available online at *http://www.cmla.ens-cachan.fr/~younes*.

Yuille, A.L., and Poggio, T.A. (1986), "Scaling Theorems for Zero Crossings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, 15–25.