INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0325

LANDMARK-BASED REGISTRATION OF 1D CURVES AND FUNCTIONAL ANALYSIS OF VARIANCE WITH WAVELETS

J. BIGOT

http://www.stat.ucl.ac.be

Landmark-based registration of 1D curves and functional analysis of variance with wavelets

Jérémie Bigot
Institut de statistique, UCL
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve, Belgique
email: bigot@stat.ucl.ac.be

Abstract

This paper is concerned with the problem of the alignment of multiple sets of curves and their comparison with FANOVA techniques. A nonparametric approach is proposed to estimate the zero-crossings lines of the continuous wavelet transform of a 1D signal observed with noise. A new tool, the "structural intensity", is introduced to represent the locations of the significant landmarks of an unknown curve via a probability density function. This technique yields an automatic landmark-based registration method to synchronize a set of curves. A fixed-effects FANOVA model is then used to test the significance of main/interaction effects and to show the usefulness of curve alignment. Some real examples arising from the biomedical area are used to illustrate the methodology.

Keywords: Curve alignment, landmark-based registration, feature detection, scale-space representation, wavelets, zero-crossings lines, structural intensity, FANOVA, nonparametric estimation.

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 consists in deciding whether there is 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 which 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)).
Analysis of variance (ANOVA) is certainly one the most used technique for data analysis in applied statistics. This approach aims at determining how the variations of a variable can be explained by the observation of other quantities. This technique, which is very useful for handling low dimensional data, has its limitations when the observed variables are functional responses (e.g. growth curves for different subjects). Although we never observe a whole function, a sufficiently large number of observations is usually available. 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 well suited (especially due to the "curse of dimensionality", see Fan and Li (1997), and Abramovich *et al.* (2002) 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. Given the various applications involving the analysis of functional data, FANOVA models have recently gained popularity, and there exits a wide literature on this subject. Detailed reviews of these techniques can be found in Ramsay and Silverman (1997) and Stone *et al.* (1997).
However, the functions to be analysed may be not observed on the same intervals, and they might also exist some variations in dynamic and time from one curve to another (due for example to the recording process) that must be modified since they could lead to erroneous conclusions. Hence, to compare similar objects, it is generally necessary to find a common referential to represent them. A possible approach consists in 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 of deformations and warping can be found in a tutorial by Younes (2000) while extensive references on curve alignment for functional data analysis can be found in Ramsay and Silverman (1997). In what follows, the terms *alignment*, *warping*, *registration* or *matching* will also be used to refer to the *synchronisation* of set of signals. Curve alignment is then a pre-required task which is often necessary before the statistical analysis of a data set. 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 registration in a statistical setting include Kneip and Gasser (1992), Gasser and Kneip (1995) with the use of kernel estimators to retrieve the structural points of a function, and Munoz Maldonado *et al.* (2002). Usually, landmark-based matching comprises the following stages:

1. Definition of the structural points that will be used as landmarks of a signal.

2. Detection of the landmarks of a signal 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 signal.

3. Determination of the landmarks that are common to a set of signals and of 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.

4. Computation of warping functions that synchronize the common landmarks.

5. Deformation of the observed curves with the transformations obtained by landmark-based matching.

The crucial points of this method are steps (2) and (3). Generally, they are performed manually (see e.g. Munoz Maldonado *et al.* (2002)) which can be tedious if the number of signals is large. In this paper, we will consider automatic methods to estimate the landmarks of a noisy function and to determine the landmarks that should be aligned when comparing various signals. We propose to use the scale-space approach developed in Bigot (2002, 2003) to detect the structural points of a signal corrupted by noise. In Bigot (2002, 2003) a nonparametric approach is designed to estimate the zero-crossings lines of the continuous wavelet transform of a noisy signal, and a new tool, called the *"structural intensity"*, is introduced to identify the limits of these lines. The structural intensity can be considered as a probability density function whose main modes are shown to be located at the significant landmarks of the unknown signal. Then, this technique yields a new method to automatically decide which pairs of landmarks should be associated when comparing two functions to be aligned.
Two real examples are used to illustrate our landmark-based registration approach:

- a comparison of the distribution of ganglioside in brain tissues between old and young rats. This set of curves is analysed in 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 data set will be referred to as the ganglioside data set.

- an example arising from physiology. 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 is to better understand the processes underlying movement generation under various levels of an externally force applied to the knee of various young male subjects. In what follows, this data set will be referred to as the orthosis data set.

Abramovich *et al.* (2002) have recently proposed a fixed-effects FANOVA model for which hypothesis testing procedures based on wavelet decompositions have been constructed to quantify the variations within a set of curves. To illustrate their procedures, Abramovich *et al.* (2002) applied these tests to the orthosis data set. However, their analysis has been carried out *without pre-registration* of the observed data. We have therefore applied the same hypothesis tests *with pre-registration* of

the orthosis data set to illustrate the usefulness of curve alignment. The FANOVA model proposed by Abramovich *et al.* (2002) is also used for comparing two samples of curves from the ganglioside data set. The obtained results are compared with the conclusions of the study carried out in Munoz Maldonado *et al.* (2002).
The paper is organized as follows: in section 2 we present the problem of landmark-based matching and review some existing approaches. In section 3, we briefly recall the scale-space approach developed in Bigot (2002, 2003). Then, in section 4, we explain how the notion of structural intensity can be used to design a new technique to automatically align a set of curves from the estimation of their landmarks. As an illustration, we combine in section 5 our landmark-based registration technique with the FANOVA model of Abramovich *et al.* (2002) to analyse the ganglioside and orthosis data sets. Finally, in section 6 we provide concluding remarks and propose some hints for possible extensions of our methodology.

2 A matching problem in one dimension

Suppose we are given two functions $f_1, f_2 : [0,1] \to \mathbb{R}$, and that we want to synchronise them. The issue is then to find for each point $x \in [0,1]$ a point $x' \in [0,1]$ such that $f_1(x) \approx f_2(x')$. If every point of one curve is uniquely matched to some point of the other curve, the problem is then to find a one-to-one function $u : [0,1] \to [0,1]$ such that:

- $f_1 \approx f_2 \circ u$.
- u is not "too far" from the identity function.

The first constraint represents the accuracy of the alignment, while the second constraint aims at finding a smooth transformation that does not "too much" deform f_2. Two approaches can be investigated to solve this problem. One consists in aligning individual locations of corresponding structural characteristics from one curve to another, while the second approach is based on the minimization of an appropriate functional. To have a sufficiently smooth warping function, we will require that u is a strictly monotically increasing and differentiable function on $[0,1]$.

2.1 Landmark-based matching

A first approach to determine a warping function u is to find the landmarks of each curve $f_j, j = 1, 2$. Landmarks are ordered series of locations (in $[0,1]$) that are associated to typical features of a curve. Typical examples of landmarks are the locations of extrema or inflection points of a function. Suppose we are given two sets of labelled 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 diffeomorphism u such that for all $i = 1, \ldots, N$: $\tau_{2,i} = u(\tau_{1,i})$ (*exact matching*) or $\tau_{2,i} \approx u(\tau_{1,i})$ (*inexact matching*). Generally, it will be also required that $u(0) = 0$ and $u(1) = 1$.

Let $f_1, \ldots, f_m$ be a set of m curves ($m > 2$) and $(\tau_{j,1}, \ldots, \tau_{j,N}), j = 1, \ldots, m$ be m sets of labelled landmarks from the curves $f_j, j = 1, \ldots, m$, then the problem of synchronizing these m curves from the alignment of their landmarks consists in finding m transformations $u_1, \ldots, u_m$ such that for all $i = 1, \ldots, N$ and $j = 1, \ldots, m$: $u_j(\tau_{j,i}) = \tau_{*,i}$ or $u_j(\tau_{j,i}) \approx \tau_{*,i}$, where $\tau_{*,i} \in [0, 1]$ is the location of the i-th landmarks $\tau_{j,i}, j = 1, \ldots, m$ after alignment (e.g. their average location).

One of the first issue encountered by landmark-based matching methods is the correspondence problem between two sets of features extracted from two functions to be aligned. For the 1D case, it is generally assumed that one is given two sets of labelled landmarks which consist of pairs of structural points to be matched. When two unknown signals are observed with noise, the problem of labelling two sets of landmarks is further complicated by the presence of outliers and the fact that there could be many features in either set that have no counterpart in the other. The identification of the landmarks that should correspond is then generally done manually (by an expert or from *a priori* informations, see e.g. Munoz Maldonado *et al.* (2002)). Gasser and Kneip (1995) have proposed a heuristic method to determine the typical features that occur consistently in a set of curves. If $f_1, \ldots, f_m$ is a set of curves eventually observed with noise, then Gasser and Kneip (1995) suggest the following methodology to identify common landmarks of the same type (e.g. local maxima):

- First, all local maxima of the curves $f_i, i = 1, \ldots, m$ are estimated. In Gasser and Kneip (1995), this step is performed by retrieving all the appropriate zeros of the first derivative of each curve obtained with kernel estimators.
- All the locations of these local maxima are sorted into one array that is submitted to kernel density estimation.

Given that in practice inter-individual variations of landmark locations are generally small compared to intra-individual variations, the common maxima should give rise to peaks in the density obtained in this way. In particular, this method allows one to define small intervals (on $[0, 1]$) where the local maxima of each curve should be estimated. Kneip and Gasser (1992) also proposed a set of statistical tools to study landmark-based matching. They introduced the notion of structural functional that can be used to define the common landmarks of a set of curves. But in practice, to determine a set of structural functionals it is necessary to define a set of sub-intervals of $[0, 1]$ where the landmarks are likely to be located. The heuristic approach described by Gasser and Kneip (1995) might help the statistician to determine such intervals, but, as remarked by Kneip and Gasser (1992), defining structural functionals in such a way is rather difficult and prone to error: some modes in the density may be due to spurious estimates while important landmarks for some curves may be ignored. Moreover, a sufficiently large number of signals should be observed in order to apply this approach. Hence, it cannot be used for the alignment of only two curves. In section 4, we therefore propose a method to automatically solve this correspondence problem.

Once the estimation of the landmarks and the correspondence problem have been solved, one need to compute a deformation that aligns two sets of features. The following methods can be used to compute monotonic transformations:

• **Constrained smoothing splines**: In many nonparametric regression problems, it is required that the curve estimate obtained from smoothing satisfy certain constraints, such as monotonicity. A general framework for constrained smoothing has been proposed by Mammen *et al.* (2001) in which many references on this subject can be found. Much of the work in constrained nonparametric regression has been done in the context of splines (note that constrained smoothing splines are also discussed in Mammen and Thomas-Agnan (1999)). Smoothing splines are defined as minimizers of a penalized sum of squares, see (2.1). Constrains are easily incorporated by minimizing over a restricted set. Given two sets of landmarks $(\tau_{1,1}, \ldots, \tau_{1,N})$ and $(\tau_{2,1}, \ldots, \tau_{2,N})$, a possible approach to find a warping function u that aligns these two sets, is then to search the minimum among all strictly monotone and differentiable $h : [0, 1] \mapsto [0, 1]$ of:

$$\|h\|^2 = \frac{1}{N}\sum_{i=1}^{N}(\tau_{2,i} - h(\tau_{1,i}))^2 + \lambda \int_0^1 \{h'(t)\}^2 dt, \tag{2.1}$$

where λ is the smoothing parameter (note that the case $\lambda = 0$ corresponds to exact matching).

• **A differential equation for monotone functions**: Ramsay (1998) has proposed to simplify the problem of nonparametric regression under monotonicity constraints, by considering warping functions h of the form $h = \beta_1 + \beta_2 D^{-1}\exp[D^{-1}w]$, where w is a square-integrable function, β_1 and β_2 are arbitrary constants, and $D^{-1}h(t) = \int_0^t h(x)dx$. This parametric form ensures both that the function h is strictly monotone and that its first derivative is smooth and bounded almost everywhere. For h in this particular class of functions, the warping function u can be defined as the minimizer of the following functional F:

$$F(h) = \frac{1}{N}\sum_{i=1}^{N}(\tau_{2,i} - \beta_1 - \beta_2 m(\tau_{1,i}))^2 + \lambda \int_0^1 \{w(t)\}^2 dt,$$

where λ is a smoothing parameter, $m(t) = \left(D^{-1}\exp D^{-1}w\right)(t)$ and $h(t) = \beta_1 + \beta_2 m(t)$, $t \in [0, 1]$. The principal advantage brought by this parametric form for the warping function is the transformation from the constrained regression problem (2.1) to the one of computing the unconstrained function w. Details for the practical implementation of this approach can be found in Ramsay (1998) and Ramsay and Li (1998), while Matlab programs for creating such warping functions are available from the ftp site:
http://www.psych.mcgill.ca/faculty/ramsay/ramsay.html.

2.2 Dense matching

Let F be a function defined on $]0, \infty[\times \mathbb{R} \times \mathbb{R}$. Another approach to find a warping function to align two curves f_1 and f_2, is to search the minimum, among all increasing diffeomorphisms u of $[0,1]$, of the functional:

$$L_{f_1,f_2}(u) = \int_0^1 F(u'(x), f_1(x), f_2 \circ u(x))dx.$$

The function F and therefore the functional L_{f_1,f_2} have to satisfy some specific properties to guarantee that a solution of this minimization problem corresponds to a solution of our matching problem. For further references on matching 1D structures by variational methods see the paper by Trouve and Younes (2000).
Wang and Gasser (1997, 1999) have studied the problem of dense matching from a statistical point of view and have proposed to use the following functional:

$$L_{f_1,f_2,\alpha}(u) = \int_0^1 \alpha^2 \left(\frac{f_1(x)}{\|f_1\|_\infty} - \frac{f_2(u(x))}{\|f_2\|_\infty} \right)^2 + (1-\alpha)^2 \left(\frac{f_1'(x)}{\|f_1'\|_\infty} - \frac{f_2'(u(x))}{\|f_2'\|_\infty} \right)^2 + 2\phi(u(x))dx,$$

where $\alpha \in [0,1]$ and $\phi(u(x))$ is a regularization term which penalizes too irregular warping functions. The solution suggested by Wang and Gasser (1997, 1999) is a compromise between alignment of the functions and alignment of their derivatives. The incorporation of the derivatives of the functions is intended to align common extrema, a feature which is required in most applications, while the normalization with the sup-norm $\|.\|_\infty$ is intended to reduce the differences in amplitude between the curves. In Wang and Gasser (1997, 1999) the functional $L_{f_1,f_2}(u)$ is then minimized by dynamic time warping (DTW). We refer to Sakoe and Chiba (1978) for an introduction to this concept and to Wang and Gasser (1997, 1999) for computational details on the calculation of the warping function which minimizes the cost $L_{f_1,f_2,\alpha}(.)$.

2.3 A nonparametric regression setting

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 $t_i = i/n$:

$$y_{j,i} = f_j(t_i) + \sigma_j \epsilon_{j,i},\ j = 1, \ldots, m,\ i = 1, \ldots, n,$$

where $\epsilon_{j,i}$ $(i = 1, \ldots, n; j = 1, \ldots, m)$ are independent identically distributed (i.i.d.) normal variables with zero mean and variance 1, and σ_j are unknown noise level parameters. The problem that we consider in this paper is the alignment of the functions $f_1, \ldots, f_m$ from the alignment of their landmarks, i.e. the estimation of warping functions to synchronize the curves $f_1, \ldots, f_m$ in the sense of section 2.1. Suppose that it is possible to find m warping functions $u_1, \ldots, u_m$ via landmark-based matching when the curves $f_1, \ldots, f_m$ are noise free. A natural idea to estimate the functions $u_1, \ldots, u_m$ is to first estimate the landmarks of $f_1, \ldots, f_m$ when they are observed with noise, and then to use the resulting estimators to perform one of the landmark-based matching procedures described in section 2.1. Note that this issue has been studied from a statistical point of view by Kneip and Gasser (1992) by the use of kernel estimators to retrieve the locations of structural points of a function and via the concept of structural functionals. Let H_N denote the set of all $\tau = (\tau_1, \ldots, \tau_N)' \in [0,1]^N$ with $\tau_i < \tau_j$ for all $i < j$, and $C^1([0,1])$ be the set of all continuously differentiable functions on $[0,1]$. Following the notations in Kneip and Gasser (1992), we will say that an operator $G : H_N^2 \rightarrow C^1([0,1])$ is a *warping operator* if the following conditions are satisfied:

- (i) For all $(\tau_1, \tau_2) \in H_N^2$, $G_{(\tau_1,\tau_2)}(.)$ is a strictly monotonically increasing real function. Furthermore, for all $t \in [0,1]$, $G_{(.,.)}(t)$ is continuous.

- (ii) For all $\tau_1 = (\tau_{1,1}, \ldots, \tau_{1,N})' \in H_N$ and all $\tau_2 = (\tau_{2,1}, \ldots, \tau_{2,N})' \in H_N$, either:
$$G_{(\tau_1,\tau_2)}(\tau_{1,i}) = \tau_{2,i}, \text{ for all } i = 1, \ldots, N,$$
and in this case G is called an *exact warping operator*, or:
$$G_{(\tau_1,\tau_2)}(\tau_{1,i}) \approx \tau_{2,i}, \text{ for all } i = 1, \ldots, N,$$
while in this case G is called an *inexact warping operator*.

Let $\tau_j = (\tau_{j,1}, \ldots, \tau_{j,N}), j = 1, \ldots, m$ be m sets of N labelled landmarks for the curves $f_j, j = 1, \ldots, m$ respectively, and let $\tau_* = (\tau_{*,1}, \ldots, \tau_{*,N})$ be the vector of average locations of the landmarks. For some warping operator G, $u_j(.) = G_{(\tau_*,\tau_j)}(.)$ is called a warping function to synchronize the curve $f_j, j = 1, \ldots, m$. In practice, landmarks and warping functions have to be estimated. Assume that one has a method to estimate the landmarks of an unknown function (see section 3) and a method to automatically solve the correspondence problem (see section 4), then, based on a warping operator G, an estimate $\hat{u}_j$ of the warping function u_j is obtained by $\hat{u}_j(.) = G_{(\hat{\tau}_*,\hat{\tau}_j)}(.)$, where the $\hat{\tau}_j = (\hat{\tau}_{j,1}, \ldots, \hat{\tau}_{j,\hat{N}})$ denote estimated and labelled landmarks of the curve $f_j, j = 1, \ldots, m$ and $(\hat{\tau}_* = \hat{\tau}_{*,1}, \ldots, \hat{\tau}_{*,\hat{N}})$ is the vector of average locations of the estimated landmarks.

We finish this section with some remarks:

- Before registration, the curves $f_1, \ldots, f_m$ must be estimated. In this paper, wavelet estimators will be used, and we refer to Antoniadis *et al.* (2001) for a review of these methods.

- In practice, we only have discrete observations, so the functions $\hat{f}_j \circ \hat{u}_j, j = 1, \ldots, m$ will be calculated by linear interpolation of $\hat{f}_j$

- For notation convenience, we have supposed that all the curves are observed on the same interval $[0,1]$. If it is not the case, it is always possible to 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 $[a_j, b_j], j = 1, \ldots, m$ to retrieve our setting, since aligning the set $f_1, \ldots, f_m$ and aligning the set $\tilde{f}_1, \ldots, \tilde{f}_m$ are equivalent procedures.

3 Estimating the landmarks of a function

3.1 Wavelets and structure of a signal

In many practical experiments, typical examples of landmarks are extrema and inflections points of a smooth function (see e.g. Ramsay (1998), Kneip and Gasser (1992)). More generally, the landmarks considered in this paper are the points y_0 where a function f is r-times ($r \geq 1$) continuously differentiable in a neighbourhood of y_0 and such that the r-th derivative of f has a zero at y_0. To characterize such landmarks, we propose to use the framework considered in Bigot (2002, 2003) which consists in following the propagation at fine scales of the zero-crossings and the wavelet maxima lines of the continuous wavelet transform of a signal. In Bigot (2002, 2003), it is shown how to detect the singularities of a function via its wavelet maxima lines, but in this paper, we restrict our study to the estimation of the zeros of the r-th derivative of a smooth signal.

3.1.1 The continuous wavelet transform

We assume that we are working with an admissible real-valued wavelet ψ with r vanishing moments ($r \in \mathbb{N}$). We will suppose that the wavelet ψ has a fast decay and has no more than r vanishing moments which implies (see Theorem 6.2 of Mallat (1998)) that there exists θ with a fast decay such that: $\psi(u) = (-1)^r \frac{d^r \theta(u)}{dt^r}$, and $\int_{-\infty}^{+\infty} \theta(u) du \neq 0$. Moreover, we will assume that the wavelet ψ is normalized to one i.e.: $\int_{-\infty}^{+\infty} (\psi(u))^2 du = 1$. Then, by definition, the continuous wavelet transform of a function $f \in L^2(\mathbb{R})$ at a given scale s is: $W_s(f)(x) = \int_{-\infty}^{+\infty} f(u)\psi_s(u-x)du$, where $\psi_s(u) = \frac{1}{\sqrt{s}}\psi(\frac{u}{s})$.

3.1.2 Zero-crossings of the wavelet transform

Suppose that a function $f \in L^2(\mathbb{R})$ is r-times continuously differentiable in an interval $[a, b]$. By commuting the convolution and differentiation operators, and given that θ has a fast decay, we have that for all $x \in]a, b[$,

$$\lim_{s \to 0} \frac{W_s(f)(x)}{s^{r+1/2}} = \lim_{s \to 0} f^{(r)} \star \frac{1}{s}\theta\left(\frac{-x}{s}\right) = K f^{(r)}(x), \text{ where } K = \int_{-\infty}^{+\infty} \theta(u) du \neq 0 \tag{3.1}$$

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 neighbourhood of z_0. Equation (3.1) proves that the zero-crossings of the wavelet transform converge to the zeros of the rth derivative of f that are located in $[a, b]$ when the scale $s \to 0$. Hence, one can find the location of the extrema (resp. the points of inflexion) of a function by following the propagation of the zero-crossings of its wavelet transform when the scale s decreases and when the mother wavelet has $r = 1$ (resp. $r = 2$) vanishing moment(s).

3.1.3 Some scale-space properties of the zero-crossings lines

We will call *zero-crossings line* any connected curve $z(s)$ in the time-scale plane (x, s) along which all points are zero-crossings. From equation (3.1), one might think that it is sufficient to follow any zero-crossings line to detect the landmarks of a function. However, we are not guaranteed that, for any wavelet $\psi = (-1)^r \theta^{(r)}, r \geq 0$, a zero-crossings located at (z_1, s_1) belongs to a zero-crossings line that propagates towards finer scales. A zero-crossings line might be interrupted when the scale decreases. However, if θ is a Gaussian (i.e $\theta(x) = K\frac{1}{\sqrt{2\pi}\beta}e^{-\frac{x^2}{2\beta^2}}$, for some non-negative reals K and β), then for any $f \in L^2(\mathbb{R})$, the zero-crossings of $W_s(f)(x)$ belong to connected curves that are never interrupted when the scale decreases (see e.g. Proposition 6.1 of Mallat(1998)). In Figure 1, we plotted the zero-crossings of the Time Shifted Sine $f_T = 0.3\sin 3\pi[g(g(g(g(x)))) + x] + 0.5,\ x \in [0, 1]$, where $g(x) = (1 - \cos(\pi x))/2$, for a Gaussian wavelet with $r = 1$ and $\beta = 0.5$. It can be seen in Figure 1c that all the zero-crossings lines propagate to fine scales (note that in this paper, for all the Figures representing the time-scale plane, the vertical axis gives $-log_2(s)$ and the horizontal axis is the interval $[0, 1]$). In Bigot (2002, 2003), the properties of the zero-crossings lines are investigated in detail by using the notion of causality arising from the scale-space literature (see Lindeberg (1994)). However, in this paper, we will not investigate the properties of these lines for any wavelet but will restrict our study to wavelets that are derivatives of a Gaussian. Then, in this case, one obtains that the limits of the zero-crossings lines correspond to the landmarks of a signal (see Proposition 3.3 in Bigot (2002)).

3.1.4 Structural intensity of the zero-crossings lines

As pointed out in Bigot (2002, 2003), we only have a visual representation of the zero-crossings lines in the time-scale plane that indicates "where" the landmarks are located, but the analytical expression of these lines is generally unknown. To overcome this drawback, a new tool called "structural intensity" has been introduced in Bigot (2002) (see Proposition 3.4) to identify the limits of the zero-crossings lines computed for a scale-space representation derived from B-Spline and Gaussian wavelets. A possible definition for the structural intensity is the following:

Definition 3.1 *Let $f \in L^2(\mathbb{R})$ and $\psi = (-1)^r \theta^{(r)}, r \geq 1$ where θ is a Gaussian. Suppose that there exists p zero-crossings lines $z_i(s)$ that respectively converge to $y_i \in \mathbb{R}, i = 1, \ldots, p$ as s tends to zero. For $x \in \mathbb{R}$, define the structural intensity of the zero-crossings $G_z(x)$ as:*

$$G_z(x) = \sum_{i=1}^{p} \int_0^{s_{z_i}} \frac{1}{s}\theta(\frac{x - z_i(s)}{s})ds,$$

where $[0, s_{z_i}]$ is the support of the zero-crossings line $z_i(.)$.

Roughly speaking, calculating the structural intensity consists in computing the "density" of the zero-crossings along various scales. With this definition for the

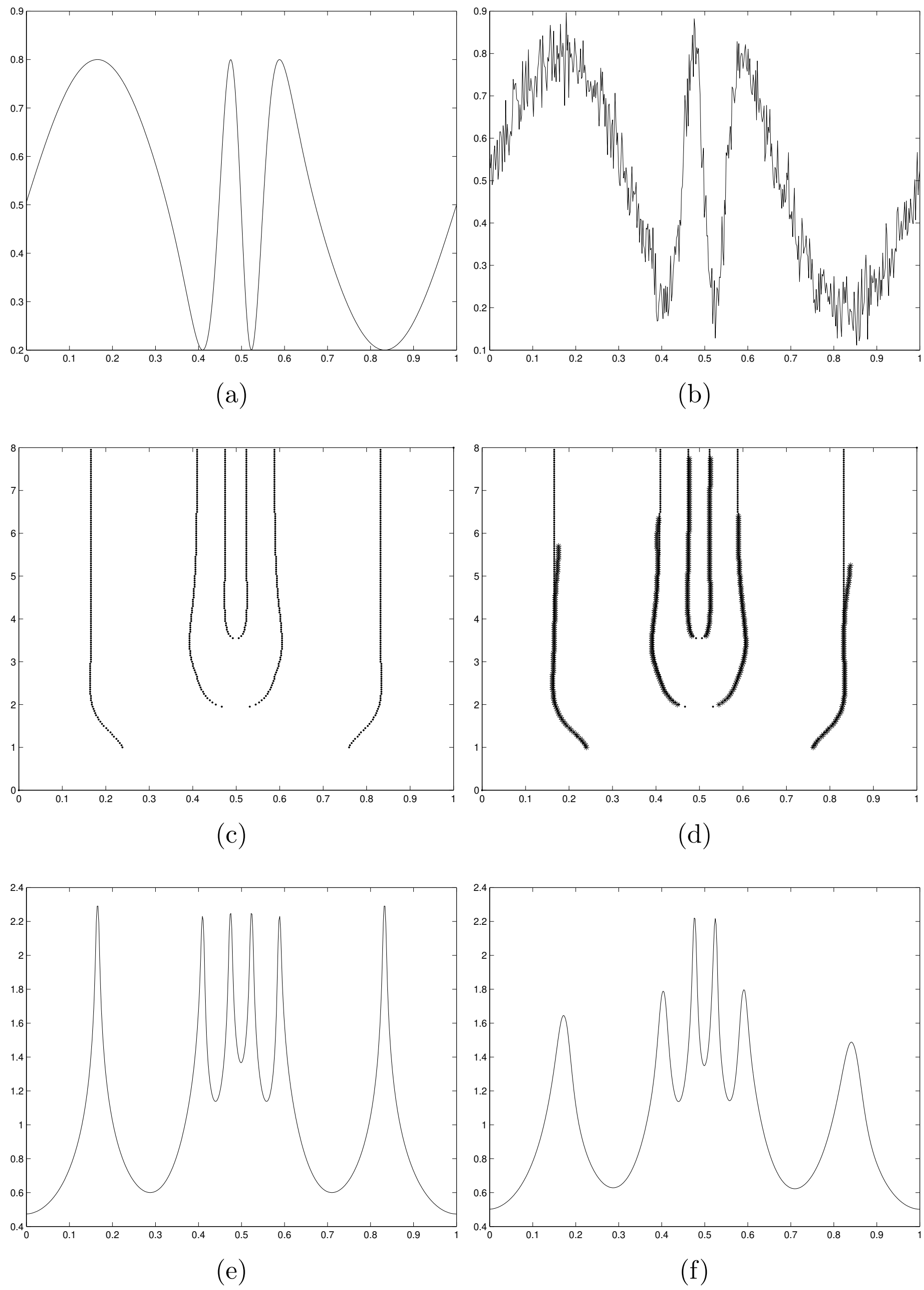


Figure 1: (a) Time function f_T. (c) Zero-crossings lines of f_T computed with a Gaussian wavelet ($\beta = 0.5$) with $r = 1$ vanishing moments. (e) Structural intensity of the zero-crossings. (b) Noisy signal f_T with $n = 512$. (d) Estimated zero-crossings (thick lines) and true zero-crossings (thin lines) of f_T for $r = 1$. (f) Structural intensity of the estimated zero-crossing of f_T.

structural intensity, it is showed in Bigot (2002, 2003), that the function G_z is continuously differentiable on $\mathbb{R} \setminus \{y_1, \ldots, y_p\}$ and such that $G_z(x) \to +\infty$ as $x \to y_i, i = 1, \ldots, p$. Hence, in practice, the modes of the resulting "density" are located at the landmarks of the signal f as it can be seen in Figure 1e where the local maxima of $G_z(.)$ correspond to the limits of the zero-crossings lines of the signal f_T. Note that in practice, the structural intensities are normalized to be probability density functions.

3.2 The nonparametric regression and white noise models

In Bigot (2002, 2003), the estimation of the zero-crossings lines has been considered in the white noise model, i.e. when f is observed according to:

$$Y(dx) = f(x)dx + \tau B(dx),\ x \in [0, 1], \tag{3.2}$$

where τ is a noise level parameter, f an unknown function which may have various landmarks, and B is a standard Brownian motion. When $\tau = \frac{\sigma}{\sqrt{n}}$, the white noise model (3.2) is closely related to the following nonparametric regression problem (see Brown and Low (1996), Donoho and Johnstone (1999, 1994) for further details):

$$y_i = f(x_i) + \sigma\epsilon_i,\ i = 1, \ldots, n, \tag{3.3}$$

where $x_i = \frac{i}{n}$, f is an unknown function, σ is the level of noise and ϵ_i are i.i.d. normal variables with zero mean and variance 1. Assume that $f \in L^2([0, 1])$ is observed according to the white noise model (3.2). If we modify a function that belongs to $L^2(\mathbb{R})$ by multiplying it by the indicator function of $[0, 1]$, we do not modify its regularity and its landmarks on $]0, 1[$. The wavelet transform of f at a scale $s > 0$ is then equal to: $W_s(f)(x) = \int_0^1 f(u)\psi_s(u - x)du$, and the wavelet transform of the white noise $B(du)$ is defined to be: $W_s(B)(x) = \int_{-\infty}^{+\infty} \psi_s(u - x)B(du)$, for $x \in [0, 1]$. Then, the wavelet transform of Y is:

$$W_{s,n}(Y)(x) = \int_{-\infty}^{+\infty} \psi_s(u - x)Y(du) = W_s(f)(x) + \frac{\sigma}{\sqrt{n}}W_s(B)(x) \tag{3.4}$$

In the following section, we will assume that the wavelet ψ is the r-th derivative of a Gaussian and will show how to estimate the landmarks of f from the zero-crossings of $W_{s,n}(Y)(x)$.

3.3 Estimation of the zero-crossings

In Bigot (2002) (see section 6.1), the following statistical test is considered to estimate the zero-crossings of $W_s(f)(x)$ at a given scale s: let $q^n_{1-\alpha}$ be 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 $0 < \alpha < 1$, then define the statistical test that accepts the null hypothesis $H_0^{s,x} : W_s(f)(x) = 0$ if $|W_{s,n}(Y)(x)| \leq \frac{\sigma}{\sqrt{n}} q^n_{1-\alpha}$ or concludes significant evidence for $W_s(f)(x)$ being positive or negative if $W_{s,n}(Y)(x) > \frac{\sigma}{\sqrt{n}} q_{1-\alpha}^n$ or $W_{s,n}(Y)(x) < -\frac{\sigma}{\sqrt{n}} q_{1-\alpha}^n$ respectively. Then, if $H_0^{s,x}$ is rejected, we have statistically significant evidence for $W_s(f)(x)$ being positive or negative depending of the sign of $W_{s,n}(Y)(x)$. Each point of significant zero-crossings of $W_s(f)(x)$ at a given scale s 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 $W_s(f)(x_1)$ and $W_s(f)(x_2)$ have opposite signs. Let s_{min} 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 is Bigot(2002): for $s \leq s_{min}$, 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.
- all the hypotheses $H_0^{s,x}$ for $\hat{x}_{1,s,i} < x < \hat{x}_{2,s,i}$ are accepted.

For $1 \leq i \leq \hat{p}_s$, if $W_{s,n}(Y)$ has **only one zero-crossings** $z_{1,s,i}$ in $[\hat{x}_{1,s,i}; \hat{x}_{2,s,i}]$ define $\hat{z}_i(s) = z_{1,s,i}$ else 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. Once various zero-crossings have been estimated at different scales, **the landmarks of the unknown signal f are defined as the local maxima of the structural intensity** of the estimated zero-crossings lines.

To complete this brief review of the methodology developed in Bigot (2002,2003), we give an example of the estimation of the local extrema of the function f_T when it is observed from the model (3.3) with $n = 512$ (see Figure 1b). Throughout this paper, the continuous wavelet transform of the discrete signal $y_i, i = 1, \ldots, n$ is computed at dyadic scales $s = 2^{-j}$ with 20 voices per octave ($1 \leq j \leq (\log_2(n) - 1)$) for a Gaussian wavelet with $r = 1$ and $\beta = 0.5$. To estimate the noise level σ we used the robust estimate suggested by Donoho and Johnstone (1994) based on the *median absolute deviation* (MAD) of the empirical wavelet coefficients associated with an orthonormal wavelet basis of $L^2([0,1])$ (we took the Symmlet 8 wavelet basis (as described on page 198 of Daubechies (1992))).
The result is very satisfactory, since it can been in Figure 1d that all the lines are correctly estimated. The local maxima of the structural intensity correspond to the local extrema of f_T, and the shape of the peaks in Figure 1f is related to the length and the location along the scale axis of the zero-crossings lines. A sharp peak corresponds to a line that propagates to fine scales, while a flat peak is related to a line that has been estimated only at some coarse scales.

4 Automatic landmark registration

We will now explain how the structural intensities of the zero-crossings lines of two curves f_1 and f_2 can be used to construct a method to automatically decide if a landmark of f_1 (resp. f_2) corresponds to a landmark of f_2 (resp. f_1). To present this method, we propose to illustrate, on a real example, the problem of identifying the pairs of landmarks that should correspond. The two curves $\tilde{g}_1$ and $\tilde{g}_2$ plotted in Figure 2ab are Medulla density profiles for two old rats extracted from the ganglioside data set (further details on how these data were recorded are provided in section 5; for both curves the number of data points is $n = 128$). Suppose that for aligning $\tilde{g}_1$ and $\tilde{g}_2$, one wants to match the significant local maxima of each curve. To retrieve the location of these landmarks, we used the procedure described in section 3 to estimate the zero-crossings that correspond to the local maxima of $\tilde{g}_1$ and $\tilde{g}_2$ (computed for a Gaussian wavelet with $r = 1$ vanishing moment and $\beta = 0.5$). The estimated zero-crossings lines are given in Figure 2 with their corresponding structural intensities. In this example, the structural intensity $\hat{G}^+_{z1}$ in Figure 2e has seven modes while the structural intensity $\hat{G}^+_{z2}$ in Figure 2f has six modes. Obviously, one cannot easily decide if the ith mode of $\hat{G}^+_{z1}$ should correspond to the jth mode of $\hat{G}^+_{z2}$ for $i = 1, \ldots, 7$ and $j = 1, \ldots, 6$. In Munoz Maldonado *et al.* (2002), five main peaks are identified in each density profile to register the fitted curves. Clearly, one can say that $\tilde{g}_1$ and $\tilde{g}_2$ have five main maxima located at $x = 0.04, 0.28, 0.42, 0.63, 0.86$ and $x = 0.04, 0.35, 0.51, 0.72, 0.89$, and the shape of these peaks suggests that these five landmarks should be matched. When looking at the zero-crossings lines converging to these five main peaks, it can be seen that the shape of the lines corresponding to the ith peaks in $\tilde{g}_1$ and $\tilde{g}_2$ are similar for $i = 1, \ldots, 5$. In some sense, to identify the pairs of landmarks that should correspond, it would be natural to try to associate the zero-crossings lines whose shape and length in the time-scale plane are similar. Since the width and the height of the modes of the structural intensities are related to the shape of the zero-crossings lines, we propose *to align $\hat{G}^+_{z1}$ and $\hat{G}^+_{z2}$ by dynamic time warping (DTW) to measure similarity between these lines in an automatic way.*

4.1 Alignment of the structural intensities by DTW

Since the functional $L_{f_1,f_2,\alpha}$ proposed by Wang and Gasser (1997, 1999) (see section 2.2) aims at aligning the extrema of two functions, it is expected that, after registration by DTW, the common modes of $\hat{G}^+_{z1}$ and $\hat{G}^+_{z2}$ would be aligned which would therefore indicate which pairs of landmark should be matched. This idea is formulated in the following definition:

Definition 4.1 *Let $G_{z_1}(x)$ and $G_{z_2}(x)$ denote two structural intensities associated to two functions f_1 and f_2. Let $\tau_1 = (\tau_{1,1}, \ldots, \tau_{1,p_1})$ be the set of local maxima of G_{z_1} and $\tau_2 = (\tau_{2,1}, \ldots, \tau_{2,p_2})$ be the set of local maxima of G_{z_1}. Define u as the warping function obtained by the minimization of $L_{G_{z_1},G_{z_2},\alpha}(.)$ (as described in Wang and Gasser (1999)). Let $\epsilon > 0$. Then, an ϵ-***correspondence*** *between the two sets of landmarks τ_1 and τ_2 is defined by the following scheme:*

- *for $i = 1, \ldots, p_1$ let $\tau^*_{2,i} = (\tau_{2,j_1}, \ldots, \tau_{2,j_i})$ be the set of landmarks that are such that $\tau_{2,j_k} \in [u(\tau_{1,i}) - \epsilon, u(\tau_{1,i}) + \epsilon]$, $\tau_{2,j_k} \in \tau^*_{2,i}$ and that have not already been associated to another landmark $\tau_{1,l}, l = 1, \ldots, i-1$.*

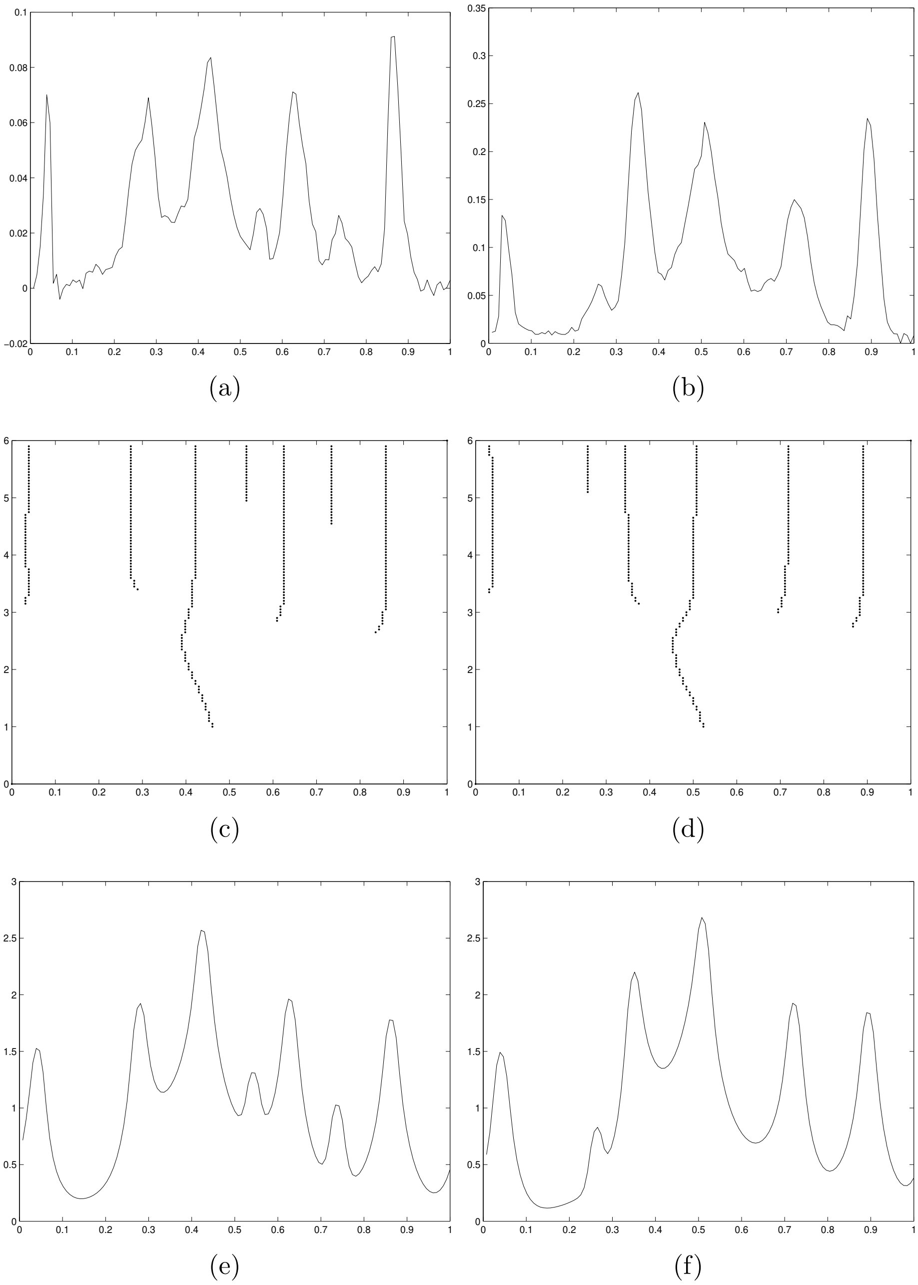


Figure 2: (a) Fitted curve $\tilde{g}_1$. (c) Estimated zero-crossings of $\tilde{g}_1$ that correspond to local maxima with their associated structural intensity $\hat{G}^+_{z1}$ (e). (b) Fitted curve $\tilde{g}_2$. (d) Estimated zero-crossings of $\tilde{g}_2$ that correspond to local maxima with their associated structural intensity $\hat{G}^+_{z2}$ (f).

- *if $\tau^*_{2,i}$ is not empty then the landmark τ_{2,j_1} is associated to the landmark $\tau_{1,i}$.*

The alignment by DTW of the structural intensities $\hat{G}^+_{z1}$ and $\hat{G}^+_{z2}$ whose modes correspond to the significant local maxima of $\tilde{g}_1$ and $\tilde{g}_2$ is given in Figure 3b. In Munoz Maldonado *et al.* (2002), it is explained that, for registering the Medulla density profiles, the five main peaks of $\tilde{g}_1$ located at $x = 0.04, 0.28, 0.42, 0.63, 0.86$ are aligned with the five main peaks of $\tilde{g}_2$ that are located at $x = 0.04, 0.35, 0.51, 0.72, 0.89$. From Figure 3b, it can be seen that if one performs ϵ-correspondence with $\epsilon = 4/128$ between the modes of $\hat{G}^+_{z1}$ and $\hat{G}^+_{z2}$, then the vectors of landmarks that should be matched are $\tau^+_1 = (0.04, 0.28, 0.42, 0.63, 0.86)$ and $\tau^+_2 = (0.04, 0.35, 0.51, 0.72, 0.89)$. Hence, in this example, ϵ-correspondence correctly aligns the five main peaks of $\tilde{g}_1$ and $\tilde{g}_2$ and therefore agrees with the matching that one would have made "by the eyes".

For the purpose of curve alignment, it is often required to match both common local maxima and minima. In Figures 3cd, we plotted the alignment by DTW of the structural intensities $\hat{G}^-_{z1}$ and $\hat{G}^-_{z2}$ of the estimated zero-crossings that correspond to significant minima of $\tilde{g}_1$ and $\tilde{g}_2$. With $\epsilon = 4/128$, the vectors of local minima that should be matched are $\tau^-_1 = (0.10, 0.34, 0.52, 0.79)$ and $\tau^-_2 = (0.15, 0.41, 0.63, 0.82)$. Since a significant maxima should be followed by a significant minima, one could define $\tau_1 = \tau^+_1 \cup \tau^-_1$ and $\tau_2 = \tau^+_2 \cup \tau^-_2$ as the vectors of landmarks to be matched, in such a way that the elements of τ_1 and τ_2 are sorted in increasing order, i.e $\tau_1 = (0.04, 0.10, 0.28, 0.34, 0.42, 0.52, 0.63, 0.79, 0.86)$ and $\tau_2 = (0.04, 0.15, 0.35, 0.41, 0.51, 0.63, 0.72, 0.83, 0.90)$. In this example, this method preserves the correspondence between local maxima and local minima. However, if for some $i > j$ one has $\tau^+_1(i) < \tau^-_1(j)$ and $\tau^+_2(i) > \tau^-_2(j)$ this procedure would associate a maxima with a minima. Hereby, to match both local minima and local minima of two curves that one wishes to align, the following approach is proposed:

- let $\hat{G}^*_{z1} = \hat{G}^+_{z1} - \hat{G}^-_{z1}$ and $\hat{G}^*_{z2} = \hat{G}^+_{z2} - \hat{G}^-_{z2}$.
- define u^* as the warping function which minimizes $L_{\hat{G}^*_{z1}, \hat{G}^*_{z2}, \alpha}(.)$.
- let $\tau^{*,+}_1$ and $\tau^{*,+}_2$ be the vectors obtained by ϵ-correspondence between the *local maxima* of $\hat{G}^*_{z1}$ and $\hat{G}^*_{z2}$, and $\tau^{*,-}_1$ and $\tau^{*,-}_2$ be the vectors obtained by ϵ-correspondence between the *local minima* of $\hat{G}^*_{z1}$ and $\hat{G}^*_{z2}$, with the warping function u^*.
- define $\tau^*_1 = \tau^{*,+}_1 \cup \tau^{*,-}_1$ and $\tau^*_2 = \tau^{*,+}_2 \cup \tau^{*,-}_2$ in such a way that the elements of τ^*_1 and τ^*_2 are sorted in increasing order. Then, take τ^*_1 and τ^*_2 as the vectors of landmarks that should be matched.

Figures 3ef gives an illustration of this procedure for the structural intensities of $\tilde{g}_1$ and $\tilde{g}_2$. With $\epsilon = 4/128$, the correspondence between the local extrema of $\hat{G}^*_{z1}$ and $\hat{G}^*_{z2}$ yields $\tau^*_1 = (0.04, 0.10, 0.27, 0.34, 0.42, 0.52, 0.63, 0.79, 0.87)$ and $\tau^*_2 = (0.04, 0.15, 0.35, 0.41, 0.51, 0.63,$

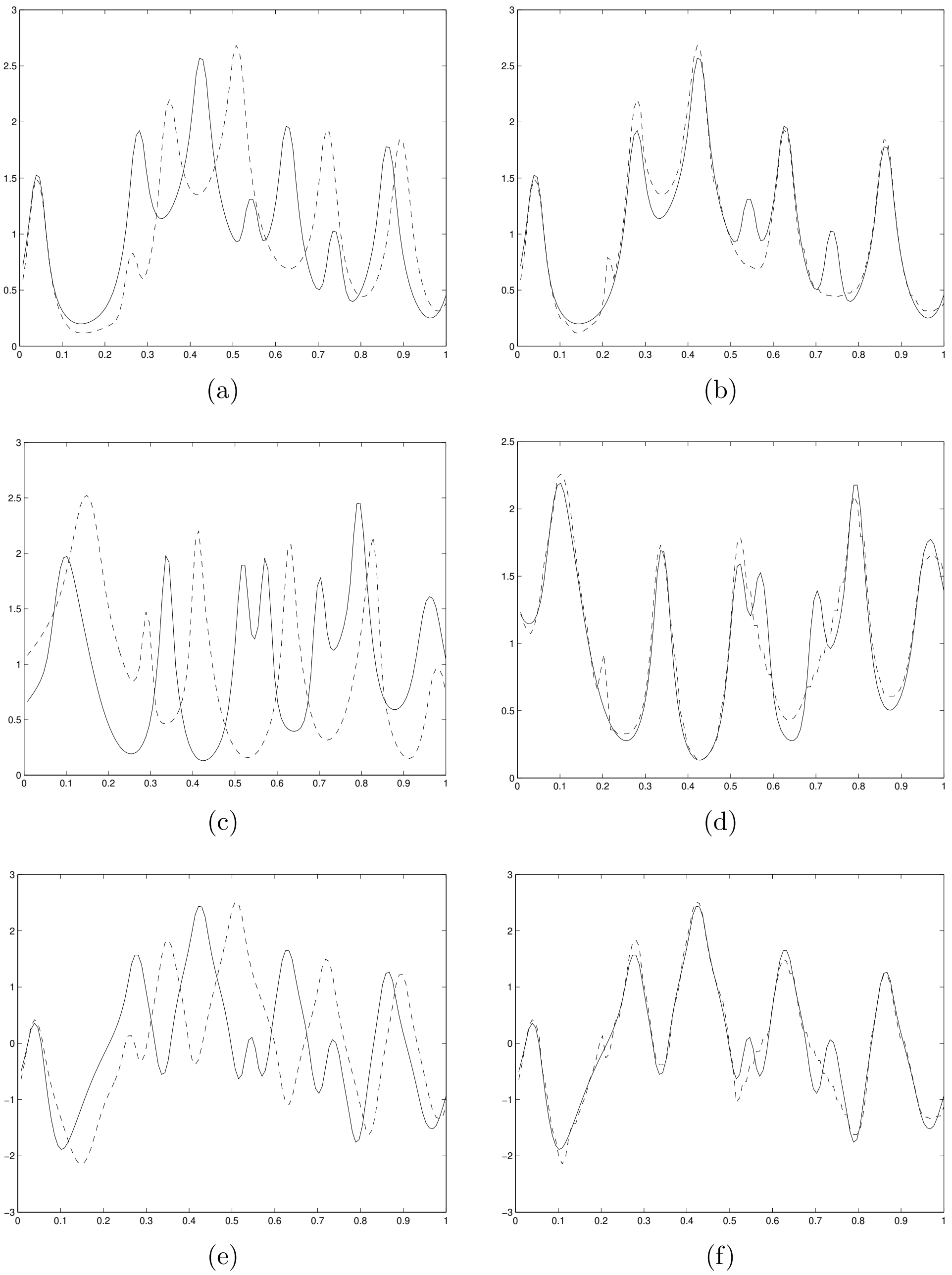


Figure 3: Various type of structural intensities for $\tilde{g}_1$ and $\tilde{g}_2$. Before registration: (a) $\hat{G}^{+}_{z1}$ (solid line) and $\hat{G}^{+}_{z2}$ (dotted line) (c) $\hat{G}^{-}_{z1}$ (solid line) and $\hat{G}^{-}_{z2}$ (dotted line) (e) $\hat{G}^{*}_{z1}$ (solid line) and $\hat{G}^{*}_{z2}$ (dotted line). After registration by DTW: (b) $\hat{G}^{+}_{z1}$ (solid line) and $\hat{G}^{+}_{z2} \circ u$ (dotted line)(d) $\hat{G}^{-}_{z1}$ (solid line) and $\hat{G}^{-}_{z2} \circ u$ (dotted line) (f) $\hat{G}^{*}_{z1}$ (solid line) and $\hat{G}^{*}_{z2} \circ u$ (dotted line).

$0.72, 0.82, 0.90)$. Hence, this method, which will be referred to as **ϵ-correspondence MinMax**, preserves the correspondence between local maxima and local minima, and it can be seen that the vectors τ_j^* and τ_j $(j = 1, 2)$ are very closed to each other.

Once the correspondence problem has been solved, a warping function which aligns the two sets of labelled landmark must be computed, and the observed curves must be estimated before being registered. For all the numerical results presented in this paper, we chose the method of Ramsay (1998) with $\lambda = 10^{-5}$ to calculate the warping functions and the translation invariant wavelet estimator with hard thresholding (TI-H) of Coifman and Donoho (1996) to estimate the curves to be aligned (with the Symmlet 8 wavelet basis). We chose the TI-H procedure because the simulative comparison study between wavelet estimators carried out in Antoniadis *et al.* (2001) shows that TI-H is markedly better than other methods for smooth functions.

4.2 Automatic landmark-based registration versus DTW

Wang and Gasser (1997, 1999) have studied the alignment of curves by DTW from a statistical point of view. They showed that the functional $L_{f_1,f_2,\alpha}$ is an effective tool to automatically align two functions, and they proved that when the data are corrupted by noise, kernel estimators can be used to estimate the curves and their first derivatives which still yields a satisfactory alignment of the estimated curves. Hence, one may wonder why we first estimate the landmarks of the functions, and then use a DTW approach to align the structural intensities instead of directly registering the curves by DTW. In Bigot (2003), a simulation study has been carried out to compare the performances of our automatic landmark-based registration method with those of the algorithm designed in Wang and Gasser (1999) for the basic problem of aligning two curves observed with noise. From these simulations, it seems that our method generally yields a better alignment of the structural points of two functions when a sufficient number of landmarks can be estimated and put into correspondence. To illustrate this, we give a real example of curve registration for which our method yields a better result than the one obtained with the algorithm designed in Wang and Gasser (1999). The two curves $\tilde{f}_1$ and $\tilde{f}_2$ plotted in Figure 4 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 $n = 256$). These curves are part of a larger data set presented in detail and analysed in Wood (2000). As explained in Wood (2002), classifying fish sounds is a problem which can be solved by registration techniques. In Figure 4c, we plotted the difference between the structural intensity of the estimated zero-crossings lines which correspond to local maxima and the structural intensity of the estimated zero-crossings lines which correspond to local minima for each curve $\tilde{f}_1$ and $\tilde{f}_2$ (i.e $\hat{G}_{z1}^* = \hat{G}_{z1}^+ - \hat{G}_{z1}^-$ and $\hat{G}_{z2}^* = \hat{G}_{z2}^+ - \hat{G}_{z2}^-$, see section 4). The landmark-based registration of the fish sounds via ϵ-correspondence MinMax and denoising by TI-H (as explained in section 4) is plotted in Figure 4b, while the alignment by DTW is plotted in Figure 4d (with the algorithm designed in Wang and Gasser (1999) and the kernel method of Gasser and Muller (1984) to estimate $\tilde{f}_1$, $\tilde{f}_2$ and their first derivative). It can be seen from Figure 4 that landmark-based registration yields a very good alignment of the extrema of $\tilde{f}_1$ and $\tilde{f}_2$, and corresponds to the alignment that would have been done manually, while the alignment obtained by DTW is not satisfactory.

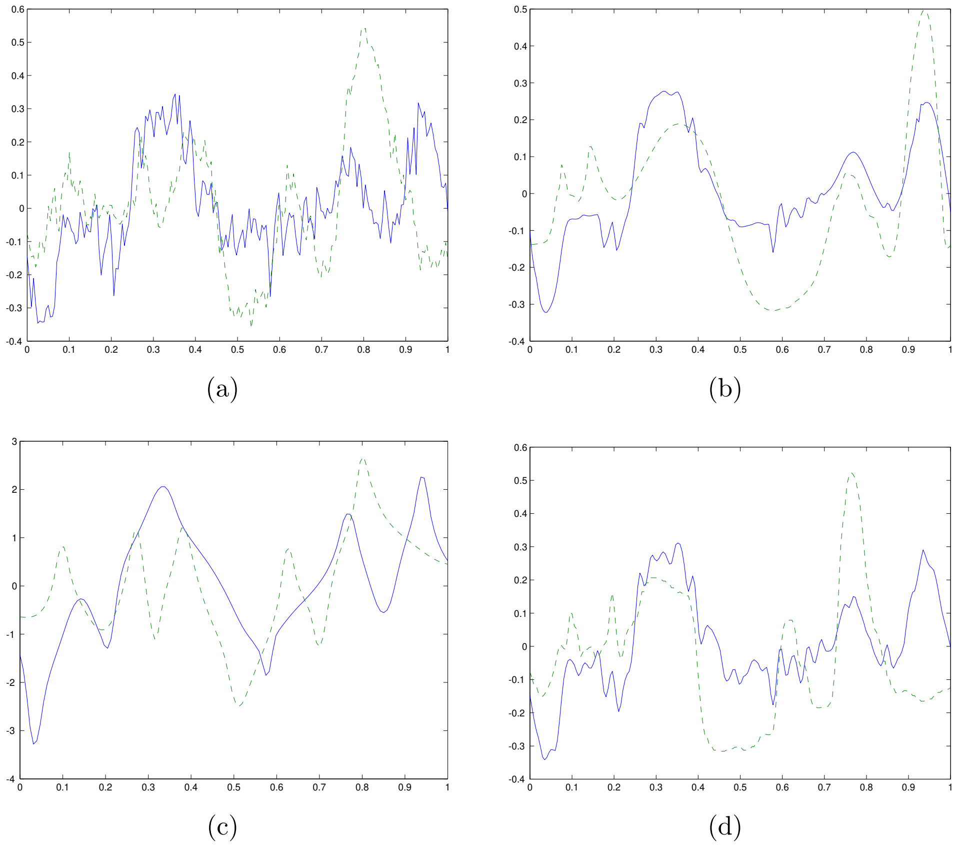


Figure 4: Fish sounds registration: (a) Noisy curves $\tilde{f}_1$ (solid curve) and $\tilde{f}_2$ (dotted curve), (b) Landmark-based registration and denoising by TI-H, (c) $\hat{G}^*_{z1} = \hat{G}^+_{z1} - \hat{G}^-_{z1}$ (solid curve) and $\hat{G}^*_{z2} = \hat{G}^+_{z2} - \hat{G}^-_{z2}$ (dotted curve), (d) Alignment by DTW and denoising by kernel smoothing

4.3 Alignment of a set of curves

Let $f_1, \ldots, f_m$ be a set of curves eventually observed with noise. To synchronise these curves from the estimation of their landmarks, we have investigated two approaches (inspired by Wang and Gasser (1997)):

1. Take each curve $f_i, i = 1, \ldots, m$ as a reference curve. For each curve $f_j, j = 1, \ldots, m, j \neq i$, use ϵ-correspondence to calculate the vector $\tau_{i,j}$ that contains the landmarks of f_i that should be associated to the landmarks of f_j. Let $\tau_i = \cap_{j=1, j\neq i}^{m} \tau_{i,j}$ be the vector that contains the landmarks of f_i that are

"common" to all curves, and denote by $\tau^*_{j,i}$ the landmarks of f_j associated to the landmarks in τ_i. Then, define f_{i_0} as the curve which maximizes the length of the vector $\tau_i, i = 1, \ldots, m$, and define τ^* as the average by component of the vectors $\tau^*_{j,i_0}, j = 1, \ldots, m$ (with $\tau^*_{i_0,i_0} = \tau_{i_0}$). In the case where various vectors τ_{i_0} of maximum length M_0 can be chosen, take the one that maximizes the cost: $\sum_{j=1}^{m} \sum_{k=1}^{M_0} (\tau^*_{j,i_0}[k] - \tau^*[k])^2$ where $\tau[k]$ is the k-th component of a vector τ. Let G be a warping operator. The warping functions $u_1, \ldots, u_m$ to synchronize $f_1, \ldots, f_m$ are then defined by $u_i(t) = G_{(\tau^*, \tau^*_{j,i_0})}(t), t \in [0, 1]$.

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. For some warping operator G, compute $u_i(t) = G_{(\tau^*, \tau^*_i)}(t), t \in [0, 1], i = 1, 2$, 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. Determine the landmarks of f_e, then calculate by ϵ-correspondence the vectors $\tau_{i,e}, i = 1, \ldots, m$ containing landmarks of f_i to be aligned with landmarks of f_e. Then, define $u_i(t) = G_{(\tau_{e,i}, \tau_{i,e})}(t)$ where $\tau_{e,i}$ are the landmarks of f_e associated to the landmarks of f_i. If the curves are observed with noise, a denoising procedure is used before aligning the curves and calculating the mean functions.

The main drawback of method (1) is its computational cost when m is large (since ϵ-correspondence is used $m(m-1)/2$ times). Hence, if m is small method (1) must be chosen, and if m is large it is better to use method (2).

5 Synchronizing samples of curves and FANOVA

5.1 Alignment of the ganglioside data set

The aim of the study carried out in Munoz Maldonado *et al.* (2002) is to 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), locus (LC) and 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 has been proposed by Munoz Maldonado *et al.* (2002) for the curves that correspond to a region of the brain:

- the two populations, old and young rats, are independently sampled.

- each density profile is a function $Y_{ij}(t)$ observed for $t \in [0, T_{ij}]$,where $i = 1, 2$ is the population index and $j = 1, \ldots, 5$ is the subject index within a group, such that
$$Y_{ij}(t) = a_{ij} \frac{d}{dt} F_{ij} \circ h_{ij}(t) + \sigma_{ij} \epsilon_{ij}(t), \tag{5.1}$$
where
  - a_{ij} are i.i.d realizations of a random variable that models the variations in the amount of tissue used for TLC.
  - $\epsilon_{ij}(t)$ are i.i.d standard normal variables and σ_{ij} are unknown noise level parameters (note that in Munoz Maldonado *et al.* (2002) it is assumed that the $\sigma_{ij} = \sigma$ are the same for the two groups and all subjects).
  - $F_{i1}, \ldots, F_{i5}$ are i.i.d., random and cumulative distribution functions, with population mean F_i and common support $[0, T]$ for $i = 1, 2$.
  - The h_{ij} ($i = 1, 2; j = 1, \ldots, 5$) are i.i.d. random, strictly increasing and C^1 functions, and $w_{ij} = h_{ij}^{-1} : [0, T] \to [0, T_{ij}]$ are called the warping functions.

The density profiles for old and young rats, and the three regions of the brain are plotted in Figure 5. As explained in 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 is also indicated in 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 the height of a peak. Therefore, to preserve the area under a peak, the registered density profiles are given by $\hat{w}'_{ij}(t)\hat{Y}_{ij} \circ \hat{w}_{ij}(t)$ and not only by $\hat{Y}_{ij} \circ \hat{w}_{ij}(t)$, where $\hat{Y}_{ij}$ is an estimated curve obtained from $\hat{Y}_{ij}$. After having synchronized the curves, the aim of the study carried out in Munoz Maldonado *et al.* (2002) is to compare the densities $f_1(t) = \frac{dF_1(t)}{dt}$ and $f_2(t) = \frac{dF_2(t)}{dt}$ (the population means of the density profiles) for each region of the brain.

5.1.1 Pre-processing of the data

Each density profile of the Figure 5 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 $\tilde{Y}_{ij}(t) = \bar{Y}_{ij}(T_{ij}t), t \in [0, 1]$. Thus, to compare the density profiles, we propose to register the curves $\tilde{Y}_{ij}(t)$ $t \in [0, 1]$ $(i = 1, 2; j = 1, \ldots, 5)$.

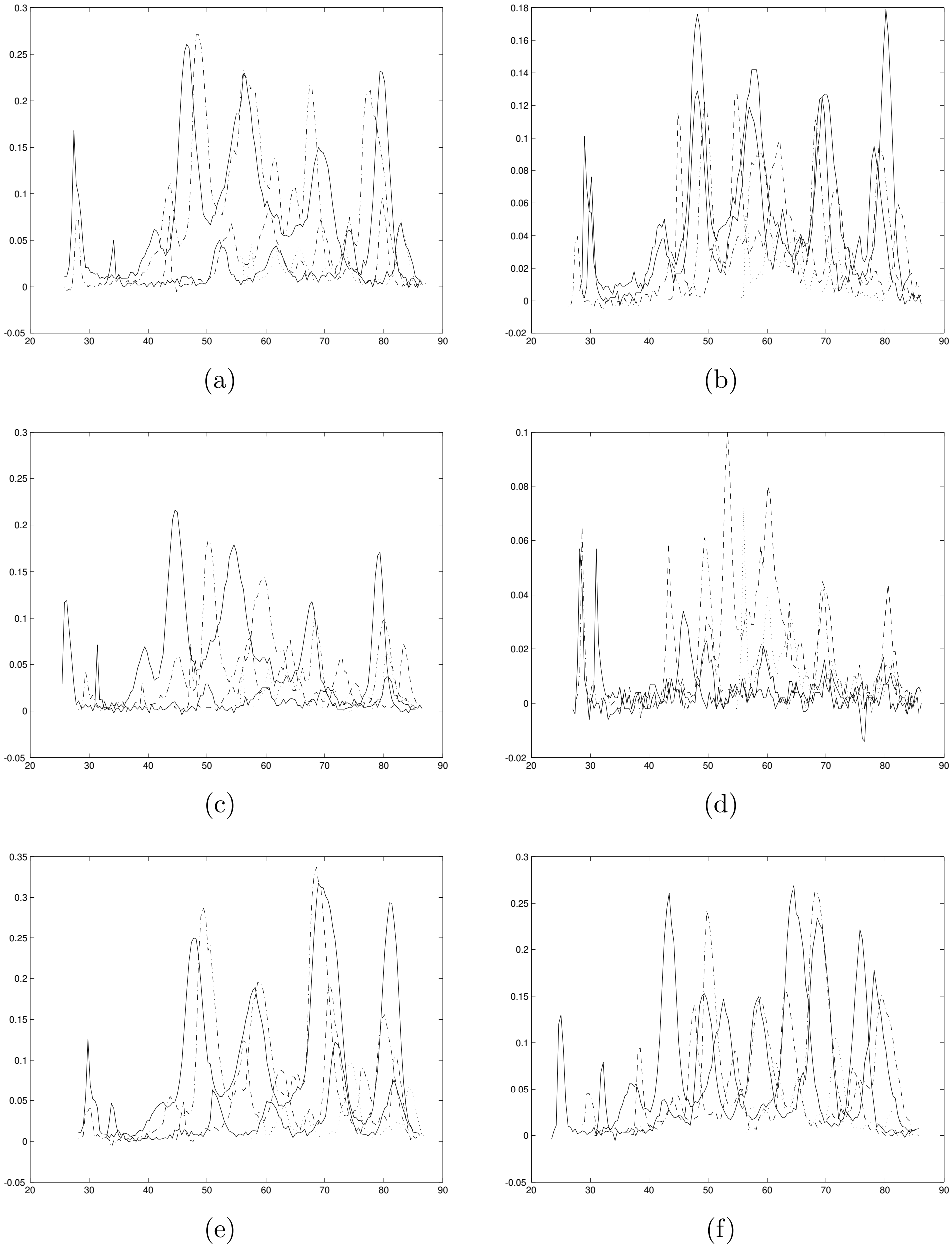


Figure 5: Density profiles of the ganglioside in three regions of the brain: MD (a) 5 old rats (b) 5 young rats, LC (c) 5 old rats (d) 5 young rats, HY (e) 5 old rats (f) 5 young rats

5.1.2 Registration from the maxima

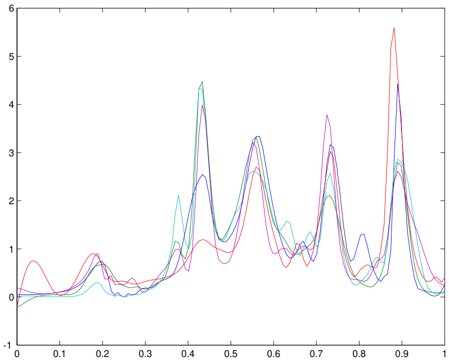
The main characteristics of the density profiles are the locations of the peaks which represent the various components of the ganglioside. We have therefore chosen to register the curves from the alignment of their significant maxima. For each group, the number of curves is relatively small (only 5), and the registration method (1) can then be used (see section 4.3). The significant maxima are estimated from the zero-crossings lines computed for a Gaussian wavelet with $r = 1$ vanishing moment (see section 3.3). Then, ϵ-correspondence is performed with $\epsilon = 4/128$, the warping functions are computed with the method of Ramsay (1998), and before registration, all the curves are estimated with the TI-H procedure (see section 4.1). To suppress the differences in amplitude between the curves, the area under each density profile is normalized to one after registration. For each warping function u, we also imposed $u(0) = 0$ and $u(1) = 1$. The alignment of the ganglioside data set (i.e. the curves $\tilde{Y}_{ij}$) is shown in Figure 6. One can see that the maxima that are common to a set of curves are correctly aligned after registration. The quality of the alignment is particularly satisfactory for the density profiles of the region HY since all the curves have clearly five main peaks which are correctly aligned. The alignment for the curves of the region LC is not so satisfactory. This is mainly due to the fact that the density profiles in the region LC do not have the same number of significant maxima. To improve the quality of the registration, some inflection points were also used as landmarks in Munoz Maldonado *et al.* (2002), but this possibility has not been investigated here.

5.2 Alignment of the orthosis data set

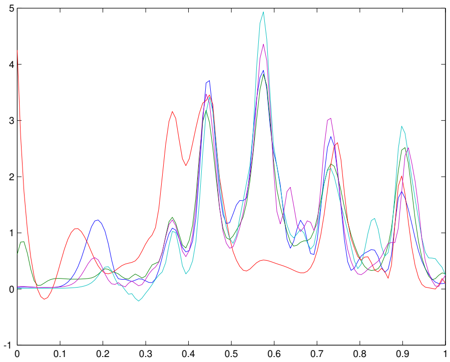
The purpose of recording the orthosis data set is to better understand the processes underlying movement generation under various levels of an externally force applied to the knee. Seven young male volunteers wore a spring-loaded orthosis of adjustable stiffness under four experimental conditions: a control condition (without orthosis), an orthosis condition (with the orthosis only), and two conditions (spring1, spring2), in which the movement is pertubated 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 *et al.* (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 data set consists of 280 separate curves (ten trials replicated for the seven subjects under the four experimental conditions). Figure 7 shows the whole data set: resultant moment as a function of time.

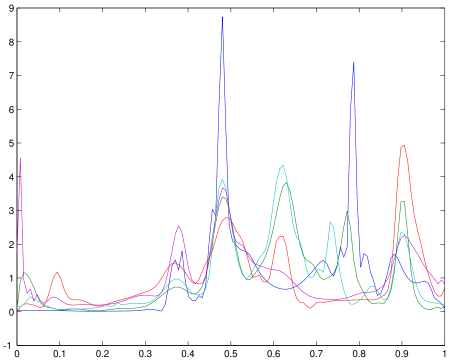
For each subject and each condition, we observe that the curves that correspond to



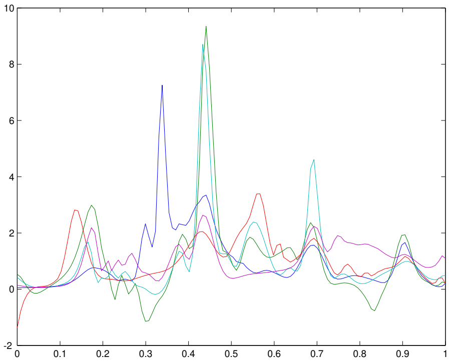
(a)



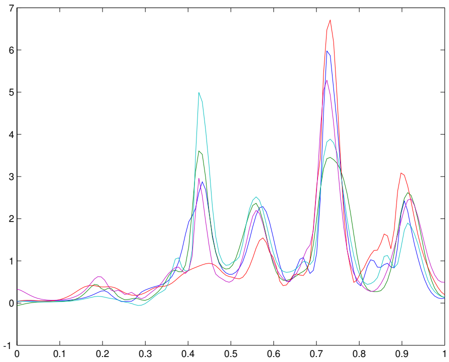
(b)



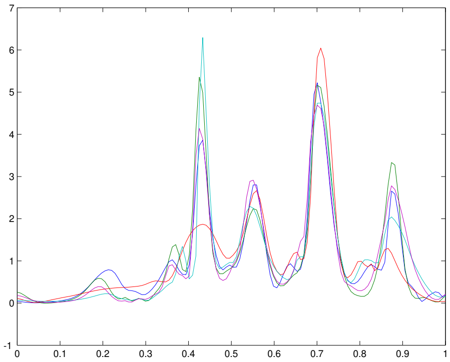
(c)



(d)



(e)



(f)

Figure 6: Alignment of the ganglioside data set for three regions of the brain: MD (a) 5 old rats (b) 5 young rats, LC (c) 5 old rats (d) 5 young rats, HY (e) 5 old rats (f) 5 young rats.

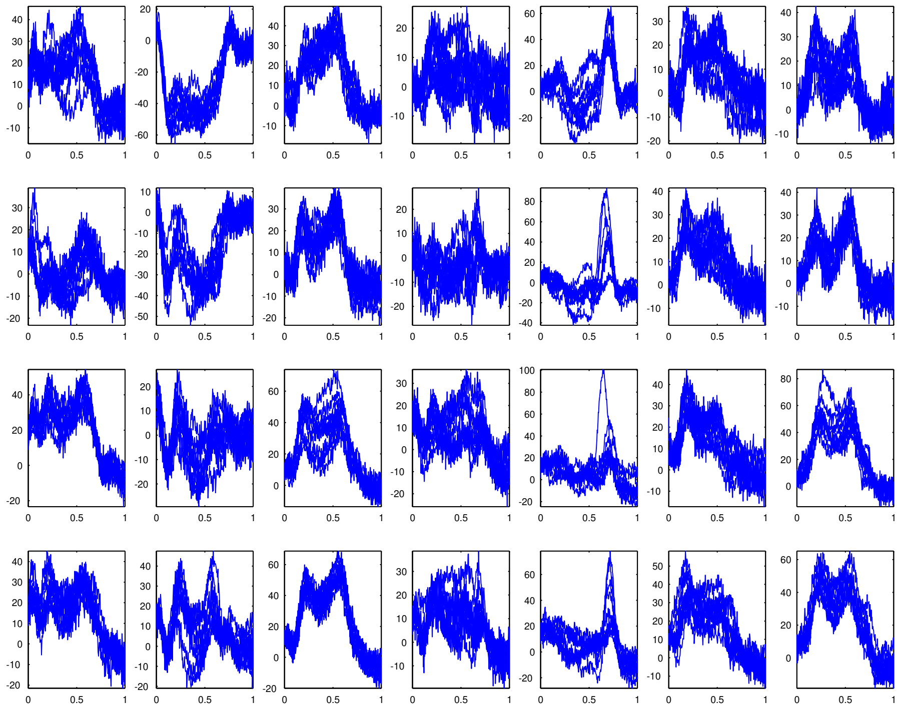
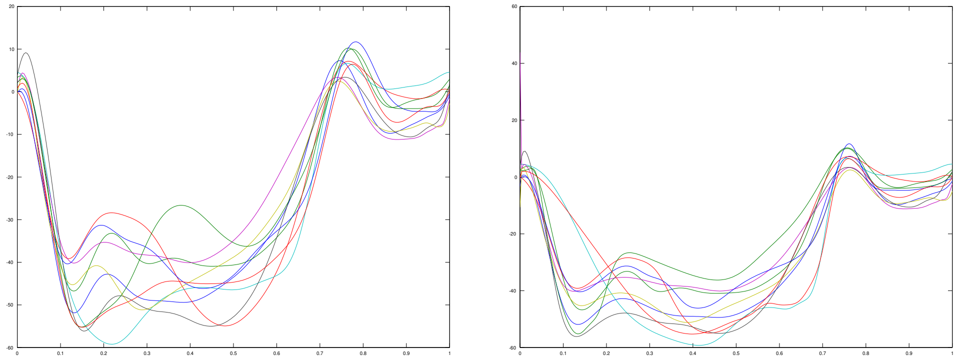


Figure 7: The orthosis data set: rows correspond to the four experimental conditions while columns correspond to the seven subjects

the 10 trials have a common structural pattern, and have approximately the same number of significant extrema. This is more obvious on Figure 8a where we have represented the estimation (with the TI-H procedure) of the resultant moments for the second subject under the control condition (first line in Figure 7). However, the curves show variations in time which can be explained by the fact that the subjects do not repeat exactly the same movement at the same speed for each trial under a given experimental condition. Therefore, to compare these curves it seems reasonable to eliminate these variations. We have chosen to register the curves for each subject and each experimental condition from the alignment of their significant extrema (estimated with zero-crossings lines computed for a Gaussian wavelet with $r = 1$ vanishing moment, see section 3.3). Given that 28 subsets of 10 curves must be synchronized, the registration method (2) has been chosen (see section 4.3). The curves have been registered via ϵ-correspondence MinMax with $\epsilon = 4/256$ (the warping functions are computed with the method of Ramsay (1998) by imposing $u(0) = 0$ and $u(1) = 1$, and before registration, all the curves are estimated with the TI-H procedure (see section 4.1)). For reasons of space we do not plot the whole registered data set. We only show the result of the alignment for the curves in Figure 8. It can bee seen from Figure 8b that the common extrema of the curves are correctly aligned after registration.



(a) (b)

Figure 8: Example of curve alignment for the orthosis data set. (a) estimation with the TI-H procedure of the resultant moments for the second subject under the control condition, (b) registration and estimation of the curves for the second subject under the control condition from the alignment of their significant extrema.

5.3 A FANOVA model with wavelets

Recall that for the ganglioside data set, 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 data set 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 need a statistical model to quantify the differences between two subsets of subjects. Abramovich *et al.* (2002) have recently proposed a fixed-effects FANOVA model, and have derived statistical procedures based on wavelet decompositions to test if the main or interaction effects are zeros against the alternative hypothesis that they are separated away from zero in L^2 norm. In this section, we propose to briefly recall their model and to explain how their procedures can be applied to the ganglioside and orthosis data sets.
Following the notations in Abramovich *et al.* (2002), suppose that we observe a serie of sample paths of a stochastic process which can be written as:

$$dY_i(t) = m_i(t)\ dt + \epsilon\ dW_i(t),\ i = 1, \dots, I;\ t \in [0, 1], \tag{5.2}$$

where ϵ is a noise level parameter, I is an integer, m_i are unknown one-dimensional functions and W_i are independent standard Wiener processes. In Abramovich *et al.* (2002), it is assumed that each of the I functions in the model (5.2) admits the following unique decomposition:

$$m_i(t) = m_0 + \mu(t) + a_i + \gamma_i(t),\ i = 1, \dots, I;\ t \in [0, 1], \tag{5.3}$$

where m_0 is a constant (*the grand mean*), $\mu(t)$ is either zero or a non-constant function of time t (the *main effect* of t), α_i is either zero or a non-constant function of i (*the main effect* of i) and $\gamma_i(t)$ is either zero or a non-zero function which cannot be decomposed as the sum of a function of i and a function of t (the *interaction* component). The components of the decomposition (5.3) also satisfy the following identifiability conditions:

$$\int_{[0,1]} \mu(t)\, dt = 0, \qquad \sum_{i=1}^{r} a_i = 0,$$

$$\sum_{i=1}^{r} \gamma_i(t) = 0, \int_{[0,1]} \gamma_i(t)\, dt = 0, \qquad \forall i = 1, \dots, I;\ t \in [0,1].$$

In this fixed-effects FANOVA model, it is interesting to be able to test the significance of the main and interaction effects. For this purpose, Abramovich *et al.* (2002) proposed to consider the following functional hypothesis tests:

$$H_0: \quad \mu(t) \equiv 0,\ t \in [0,1],\ \text{no global trend} \tag{5.4}$$

$$H_0: \quad \gamma_i(t) \equiv 0,\ \forall i = 1, \dots, r;\ t \in [0,1],\ \text{no interaction component.} \tag{5.5}$$

No parametric form is specified for μ and γ_i for the alternative hypotheses. The set of alternative hypothesis is constructed such that it is sufficiently separated away from the null hypotheses (5.4) and (5.5) for a distance ρ in $L^2([0,1])$. In order to test a large class of alternatives, it is assumed that the functions m_i belong to some Besov ball of radius $C > 0$ on $[0,1]$, $B^s_{p,q}(C)$ where $s > 0$ and $1 \le p, q \le \infty$. For particular choices of s, p and q, Besov spaces include various classes of functions (see Meyer (1992) for a detailed study of Besov spaces). The following alternatives hypotheses are then considered:

$$H_1: \quad \mu \in F(\rho),$$

$$H_1: \quad \gamma_i(t) \in F(\rho),\ \text{at least for one } i = 1, \dots, r.$$

where $F(\rho) = \{f \in B^s_{p,q}(C) : \|f\|_2 \ge \rho\}$ and $\|.\|_2$ is the standard norm on $L^2([0,1])$. Abramovich *et al.* (2002) have then developed functional hypothesis testing procedures for the null hypotheses (5.4) and (5.5) based on wavelet decompositions of the data, and have shown that these tests are asymptotically optimal over the range of Besov balls for the alternatives (5.6) and (5.6) respectively. We will not give the expressions of the test statistics derived from the wavelet decompositions of the data, and refer to Abramovich *et al.* (2002) for further details on their methodology and the asymptotic properties of their procedures. We just mention that these procedures involve the calibration of two integers $j(s)$ and j_η that define the resolution levels that must be used in the wavelet decompositions to calculate the test statistics. Rather than defining an automatic method to calculate $j(s)$ and j_η, Abramovich *et al.* (2002) have performed some numerical simulations to investigate the performances of their tests, and have derived appropriate values for $j(s)$ and j_η depending on the sample size n.

5.3.1 Orthosis data analysis

As an illustration, Abramovich *et al.* (2002) applied their tests to the analysis of the orthosis data set, but **without pre-registration of the data**. Given that the purpose of this analysis is to better understand how a subject behave under an external perturbation, it is necessary to be able to quantify the differences between the means of the individual curves for the various experimental conditions. The following model is proposed in Abramovich *et al.* (2002) for the orthosis data set:

$$dY_{ijk} = m_{ij}(t)\ dt + \epsilon\ dW_{ijk}(t),\ i = 1, \dots, I;\ j = 1, \dots, J;\ k = 1, \dots, K;\ t \in [0, 1],$$

with

$$m_{ij}(t) = m_0 + \mu(t) + \alpha_i + \gamma_i(t) + \beta_j + \delta_j(t),\ i = 1, \dots, I;\ j = 1, \dots, J;\ t \in [0, 1],$$

where i is the condition index, j is the subject index, k is the repetition index and t is the time variable. As explained in Abramovich *et al.* (2002), subjects in the above model are naturally considered as block effects. Indeed, 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 evaluate the effects of each experimental condition, one can use the functional hypothesis tests developed in Abramovich *et al.* (2002) for the following "average" model (on the seven subjects and the ten replications) if we assume standard identifiability conditions:

$$d\bar{Y}_{i\cdot\cdot} = m_i(t)\ dt + \eta\ dW_{i\cdot\cdot}(t),\ i = 1, \dots, I;\ t \in [0, 1],$$

with

$$m_i(t) = m_0 + \mu(t) + \alpha_i + \gamma_i(t),\ i = 1, \dots, I;\ t \in [0, 1],$$

where $\eta = \epsilon/\sqrt{JK}$. Given that the number of sample points is $n = 2^8$, it is recommended in Abramovich *et al.* (2002) to take $j(s) = 4$ et $j_\eta = 6$. We took the compactly supported wavelet *Coiflet 18-tap filter* to analyse these signals, and used the *median absolute deviation* (MAD) of the empirical wavelet coefficients to estimate $\epsilon/\sqrt{JK}$ for each condition. The following null hypotheses have been tested in Abramovich *et al.* (2002) (against alternatives of the kind $H_1 :\ \mu \in F(\rho)$ for a fixed $\rho > 0$):

1. $H_0 :\ \mu \equiv 0$.
2. $H_0 :\ \gamma_i \equiv 0\ i = 1, \dots, 4$.
3. $H_0 :\ \gamma_1 - \gamma_2 \equiv 0$: the effects of conditions "Control" and "Orthosis" are the same.
4. $H_0 :\ \gamma_3 - \gamma_4 \equiv 0$: the effects of conditions "Spring 1" and "Spring 2" are the same.

Ax explained previously, the analysis of the orthosis data in Abramovich *et al.* (2002) has been carried out without pre-registration of the curves. To illustrate the usefulness of curve alignment, we have compared the results obtained in Abramovich *et al.* (2002) with those obtained with pre-registration of the curves as described in section 5.2. However, before registration, the signals have been estimated. To use a FANOVA model for the registered data, it is therefore necessary to **"renoise"** the curves after registration. For each curve, the level of noise has been estimated through MAD and a sequence of Gaussian noise with the same variance has been added to each estimated and registered curve. The empirical estimators of the grand mean and the effects of each condition are given in Figure 9 for the non-registered data and the registered/renoised data. One can see that the alignment of the data modifies locally the shape of the curves for the grand mean and the treatment effects. For the non-registered data and the registered data, the null hypotheses $H_0: \ \mu \equiv 0$ and $H_0: \ \gamma_i \equiv 0 \ i = 1, \dots, 4$ are clearly rejected with a p-value smaller than 10^{-3} in both cases. For the non-registered data, the p-values of the tests $H_0: \ \gamma_1 - \gamma_2 \equiv 0$ and $H_0: \ \gamma_3 - \gamma_4 \equiv 0$ are respectively 0.097 and 0.487, while for the registered/renoised data, the p-values of the tests $H_0: \ \gamma_1 - \gamma_2 \equiv 0$ and $H_0: \ \gamma_3 - \gamma_4 \equiv 0$ are respectively 0.009 and 0.011. Hence at level $\alpha = 5\%$ and for the non-registered data, we find that under the conditions "Control" and "Orthosis", or under the conditions "Spring 1" and "Spring 2", the subjects behave similarly under perturbations of similar nature. This would therefore indicate that the subjects adjust their posture similarly under similar perturbations. But if the data are pre-registered for each subject and each condition, we conclude to the contrary that (at level $\alpha = 5\%$) the individuals behave differently under the conditions "Control" and "Orthosis", or under the conditions "Spring 1" and "Spring 2".

5.3.2 Ganglioside data analysis

For the ganglioside data, one need to decide if the shape of the density profiles shows variations with aging for each region of the brain, which can be done by quantifying the differences between the empirical population means f_1 and f_2 (see section 5.1) that are depicted in Figure 10 (obtained by averaging the registered/renoised density profiles). For each region of the brain, we took the same FANOVA model than the one used for the orthosis data but with $I = 2$ (young or old rats), $J = 5$ (5 young rats and 5 old rats) and $K = 1$ (no replication). Once registered, the data were pertubated with the same level of noise (as explained previously for the orthosis data analysis). To determine if the distribution of the ganglioside shows variations with aging, we have tested the null hypothesis $H_0: \ \gamma_1 - \gamma_2 \equiv 0$ for each region of the brain. With $j(s) = 4$ and $j_\eta = 5$, the null hypothesis is clearly rejected (p-value smaller than 10^{-3}) for the three regions MD, LC and HY. Note that this could be directly inferred by looking at the data in Figure 10 which shows that the level of noise is extremely small. Hence, contrary to the conclusions in Munoz Maldonado *et al.* (2002), our analysis indicates that the distribution of the ganglioside varies with aging in the three regions MD, LC and HY. The statistical analysis carried out in Munoz Maldonado *et al.* (2002) provides significant evidence difference in ganglioside distribution with aging in the regions MD and LC but not in the region HY. Note that the analysis carried out in Munoz Maldonado *et al.* (2002) is not based on functional hypothesis testing procedures, but on three statistics to test the notion of similarity between two curves. The test statistics are

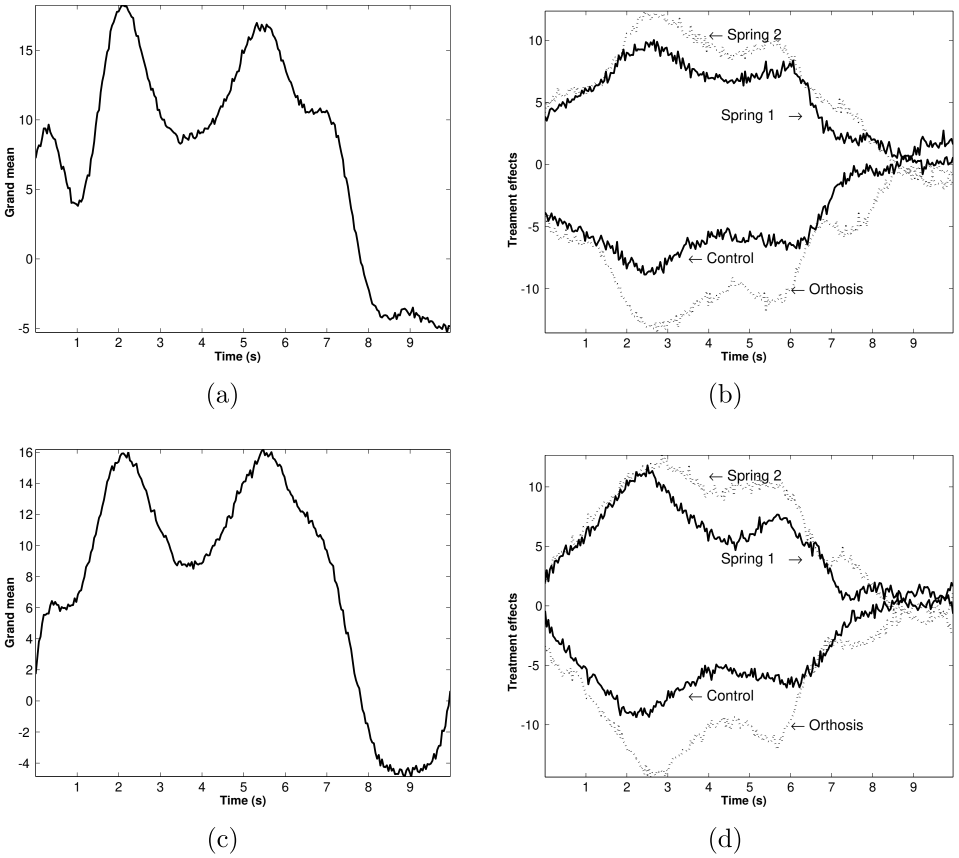
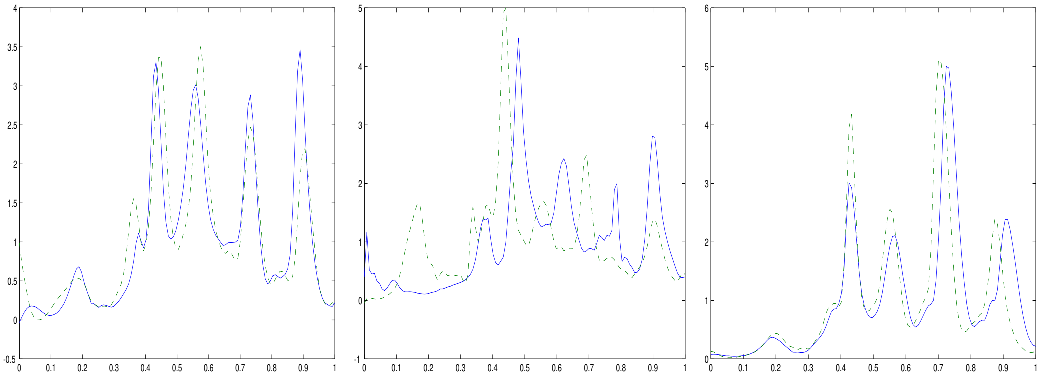


Figure 9: Orthosis data: (a) Empirical estimation of the grand mean $m_0 + \mu(t)$ and (b) empirical estimators of the treatment effects (i.e $\alpha_i + \gamma_i(t)$) *without pre-registration of the data.* (c) Empirical estimation of the grand mean $m_0 + \mu(t)$ and (d) empirical estimators of the treatment effects (i.e $\alpha_i + \gamma_i(t)$) *with pre-registration/renoising of the data.*

based on Pearson's sample correlation coefficient calculated for the vectors obtained by evaluating the registered curves on a common grid (obtained by discretization of a common interval).



(a) (b) (c)

Figure 10: Mean of the density profiles for young rats (dotted lines) and old rats (solid lines) for each region of the brain: (a) MD (b) LC (c) HY.

6 Conclusion

An important problem in landmark-based registration consists in deciding which pairs of structural points should be associated when two functions, corrupted by noise, have to aligned. A scale-space approach with wavelets has been proposed to estimate the landmarks of a function which yields a new method, called ϵ-correspondence, to automatically put into correspondence pairs of landmarks from two functions. This approach has been successfully applied on various real examples and has been shown to be an effective technique to synchronize a set of curves from the alignment of their landmarks. We have combined our landmark-based matching approach with the functional hypothesis testing procedures recently developed by Abramovich *et al.* (2002) to illustrate the usefulness of curve registration for the analysis of functional data.

A possible extension of this work is the study of the consistency of ϵ-correspondence and the analysis of the asymptotic properties of the estimators of the warping functions.

We would like to indicate that the notion of structural intensity has also been used in Bigot (2002, 2003) to extract the singularities of an unknown signal via the estimation of the wavelet maxima lines of its continuous wavelet transform. 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 from the estimation of their singularities.

Software
LandAlignWave is a library of MATLAB procedures allowing the use of the methods developed in this paper. This code can be used to reproduce the Figures displayed in the previous sections and is available for download from:
`http://www.stat.ucl.ac.be/ISpersonnel/bigot/landalignwave.html`

Acknowledgements
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 Dr. Yolanda Munoz Maldonado (Texas A&M University, USA) for providing the ganglioside data set, to Dr. Amarantini David (LABM, Marseille, France) for providing the orthosis data set and to Dr. Graham Horgan (BIOSS, Edinburgh, UK) for providing the fish sound data. I particularly thank Professor Anestis Antoniadis (University Joseph Fourier, Grenoble, France) for helpful comments and for pointing out some references. I am also very grateful to Professor Chris Glasbey (BIOSS, Edinburgh, UK) for helpful discussions and for welcoming me in Edinburgh for a research visit to carry out this work.

References

[1] Abramovich F,. Antoniadis, A. Sapatinas, T. and Vidakovic, B. (2002). Optimal Testing in Functional Analysis of Variance. *To appear in JASA*.

[2] Antoniadis, A. Bigot, J. and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software*, Vol. 6, Issue 6, 1–83.

[3] Bigot, J. (2003). Recalage de signaux et analyse de variance fonctionnelle par ondelettes. Applications au domaine biomédical. Thèse de l'Université Joseph Fourier (Grenoble, France),
`http://www.stat.ucl.ac.be/ISpersonnel/bigot/theseBIGOT.pdf`.

[4] Bigot, J. (2002). A scale-space approach to landmark detection. Technical Report TR2046, PAI (Interuniversity Attraction Pole network),
`www.stat.ucl.ac.be/Iapdp/tr2002/TR0246.pdf`.

[5] Brown, L.D. and Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise, *Ann. Statist.* **3** No.6 , 2384–2398.

[6] 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.

[7] Coifman, R.R. and Donoho, D.L. (1995). Translation-invariant de-noising. In *Wavelets and Statistics*, Antoniadis, A. and Oppenheim, G. (Eds.), Lect. Notes Statist., **103**, pp. 125-150, New York: Springer-Verlag.

[8] Daubechies, I. (1992). Ten Lectures on Wavelets, *Philadelphia: SIAM.*

[9] Donoho, D.L. and Johnstone, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika* **81**, 425–455.

[10] Donoho, D.L. and Johnstone, I.M. (1999). Asymptotic minimality of wavelet estimators with sampled data, *Stat. Sinica* **9**, 1–32.

[11] Fan, J. and Lin, S K. (1998). Test of significance when data are curves, *JASA* **91**, 2, 674–688.

[12] Gasser, T. and Kneip, A. (1995). Searching for Structure in Curve Samples, *J. Am. Statist. Ass.* **90** No.432, 1179–1188.

[13] Gasser, T. and Muller, H. (1984). Estimating regression functions and their derivatives by the kernel method, *Scand. J. Statist.* **11** 171–185.

[14] Kneip, A. and Gasser, T. (1992). Statistical tools to analyze data representing a sample of curves, *Ann. Statist.* **20** No.3 , 1266–1305.

[15] Lindeberg, T. (1994). Scale Space Theory in Computer Vision, *Kluwer, Boston.*

[16] Mallat, S. (1998). A Wavelet Tour of Signal Processing, *Academic Press.*

[17] Mammen, E. Marron, J.S. Turlach, B.Aånd Wand, M.P. (2001). A general projection framework for constrained smoothing, *Statistical Science* 16(3): 232-248.

[18] Mammen, E. Thomas-Agnan, C. (1999) Smoothing splines and shape restrictions. *Scand. J. Statist.* **26**, 239 - 252.

[19] Meyer, Y. (1992). Wavelets and Operators, *Cambridge: Cambridge University Press.*

[20] Munoz Maldonado, Y. Staniswalis, J.G. Irwin, L.N. and Byers, D. (2002). A similarity analysis of curves, *The Canadian Journal of Statistics* **30** No.3, ???–???.

[21] Ramsay, J.O. (1998) Estimating smooth monotone functions. *Journal of the Royal Statistical Society*, Series B, 60, 365-375.

[22] Ramsay, J.O. and Li, X. (1998) Curve registration. *Journal of the Royal Statistical Society*, Series B, 60, 351-363.

[23] Ramsay, J.O. and Silverman, B.W. (1997). Functional data analysis, *New York: Springer Verlag.*

[24] Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition, *IEEE Trans. Acoust. Speech Signal Process.* **26**, 43-49.

[25] Stone, C.J. Hansen, M. Kooperberg, C. and Truong, Y. (1997). Polynomial splines and their tensor products in extended linear modeling (with discussion), *Ann. Statist.* **25**, 1371-1470.

[26] Trouve, A. and Younes, L. (2000). Diffeomorphic matching problems in one dimension: designing and minimizing matching functionals, *Proceedings ECCV 2000*.

[27] Wang, K. and Gasser, T. (1997). Alignment of curves by dynamic time warping, *Ann. Statist.* **25** No.3, 1251–1276.

[28] Wang, K. and Gasser, T. (1999). Synchronizing sample curves nonparametrically, *Ann. Statist.* **27** No.2, 439–460.

[29] Wood, M. (2002). Discriminant Analysis using Wavelet Derived Features, *PhD Thesis, University of Aberdeen*.

[30] Younes, L. (2000). Deformations, Warping and Object Comparison, *Tutorial* `http://www.cmla.ens-cachan.fr/ỹounes`.