



HAL
open science

Manifold embedding for curve registration

Jean-Michel Loubes, Elie Maza, Chloé Dimeglio

► **To cite this version:**

Jean-Michel Loubes, Elie Maza, Chloé Dimeglio. Manifold embedding for curve registration. 2011. hal-00593180v1

HAL Id: hal-00593180

<https://hal.science/hal-00593180v1>

Preprint submitted on 30 Mar 2011 (v1), last revised 13 May 2011 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MANIFOLD EMBEDDING FOR CURVE REGISTRATION

BY CHLOÉ DIMEGLIO

Institut de Mathématiques de Toulouse & Geosys

AND

BY JEAN-MICHEL LOUBES AND ELIE MAZA

Institut de Mathématiques de Toulouse

We focus on the problem of finding a good representative of a sample of random curves warped from a common pattern f . We first prove that such a problem can be moved onto a manifold framework. Then, we propose an estimation of the common pattern f based on an approximated geodesic distance on a suitable manifold. We then compare the proposed method to more classical methods.

1. Introduction. The outcome of a statistical process is often a sample of curves, f_i , $i = 1, \dots, m$, showing an unknown common structural pattern, f , which characterizes the behaviour of the observations. Examples are numerous, among others growth curves analysis in biology and medicine, quantitative analysis of microarrays in molecular biology and genetics, speech signals recognition in engineering, study of expenditure and income curves in economics.... Hence, among the last decades, there has been a growing interest to develop statistical methodologies which enables to recover from the observation functions a single "mean curve" that conveys all the information of the data.

A major difficulty comes from the fact that there are both amplitude (variation in the y -axis) or phase (variation in the x -axis) which prevent any direction extraction of the mean, median, correlations or any other statistical indices for a the standard multivariate procedure such as principal component analysis, and canonical correlations analysis, see [KG92] or [RS05] and references therein. Indeed the classical cross-sectional mean does not provide a consistent estimate of the function of interest f since it fails to capture the structural characteristics in the sample of curves as quoted in [RL98]. Hence curve registration (also called curve alignment, structural averaging, and time warping) methods have been proposed in the statistical literature.

AMS 2000 subject classifications: Primary 62G05; secondary 62M99

Keywords and phrases: Manifold learning, Intrinsic statistics, Graph-based methods, Curve alignment, Curve registration, Structural statistics

We refer to [KG92], [Sil95], [WG97], [Røn01], [LM04], [GG05], [GLM07], [Jam07], and [DLM11] just to name a few.

This issue is closely related to the problem of finding the mean of observations lying in a space with an unknown, non necessarily euclidean, underlying geometry. The problem is thus twofold.

First, the mere definition of the mean should be carefully studied. Indeed, let $\mathcal{E} = \{X_1, \dots, X_n\}$ be a sample of i.i.d random variables of law $X \in \mathcal{M}$ where \mathcal{M} is a submanifold of \mathbb{R}^p . If we denote by d the Euclidean distance on \mathbb{R}^p , then the classical sample mean, or Fréchet sample mean, defined by

$$(1) \quad \hat{\mu} = \arg \min_{\mu \in \mathbb{R}^p} \sum_{i=1}^n d^2(X_i, \mu)$$

is not always a good representative of the given sample \mathcal{E} , and, obviously, of the underlying population. Using the geometry of the manifold, it seems natural to replace Criterion (1) by

$$\hat{\mu}_I = \arg \min_{\mu \in \mathcal{M}} \sum_{i=1}^n \delta^2(X_i, \mu)$$

where δ is the geodesic distance on \mathcal{M} , giving rise to the *intrinsic mean*, whose existence and properties are studied for instance in [BP03]. When dealing with functional data, we assume that the functions f_j can be modeled as variables with values on a manifold, and curve registration amounts to considering an intrinsic statistics that reflects the behaviour of the data. In the following we will consider for $\alpha > 0$,

$$(2) \quad \hat{\mu}_I^\alpha = \arg \min_{\mu \in \mathcal{M}} \sum_{i=1}^n \delta^\alpha(X_i, \mu).$$

In particular for $\alpha = 1$, we will deal with $\hat{\mu}_I^1$, the intrinsic sample median.

Second, previous construction relies on the choice of the embedding which may not be unique, then the manifold itself and its underlying geodesic distance. Actually we only have at hand a sample of random variables which are sought to be a discretization of an unobserved manifold. Over the last decade, some new technics have been developed to find and compute the natural embedding of data onto a manifold and to estimate the corresponding geodesic distance, see for instance [dST03] for a review of global (ISOMAP type) and local (LLE type) procedures, while applications have been widely developed, see for instance [Pen06].

In the following, we will consider an approximation, achieved with a graph

theory approach inspired by works on manifold learning and dimension reduction [TdSL00]. We will first show that curve registration for parametric transformations can be solved using a manifold geodesic approximation procedure. Then, we will highlight that this enables to recover a mean pattern which conveys the information of a group of curves. This pattern is used for curve classification for simulated data and real data which consist in predicting a particular landscape using the reflectance of the vegetation. We will compare this procedure to the one built with the pattern coming from the moments alignment procedure described in [Jam07].

This article falls into the following parts. Section 2 is devoted to the construction of the approximated geodesic distance. In Section 3, we describe the manifold framework point of view for curve registration. We then explain how to estimate a representative of a sample of warped curves. The performance of this estimator is then studied in Section 4 using simulated data, and in Section 5 with a real data set. Proofs are gathered in Section 7.

2. A graph construction for topology estimation. Let X be a random variable with values in an unknown connected and geodesically complete Riemannian manifold $\mathcal{M} \subset \mathbb{R}^p$. We observe an i.i.d sample $\mathcal{E} = \{X_i \in \mathcal{M}, i = 1, \dots, n\}$ with distribution X . Set d the Euclidean distance on \mathbb{R}^p and δ the induced geodesic distance on \mathcal{M} . Our aim is to estimate intrinsic statistics defined by Equation (2). Since the manifold \mathcal{M} is unknown, the main issue is to estimate the geodesic distance between two points on the manifold, that is $\delta(X_i, X_j)$.

Let γ_{ij} be the geodesic path connecting two points X_i and X_j , that is the minimum length path on \mathcal{M} between points X_i and X_j . Denoting by $L(\gamma)$ the length of a given path γ on \mathcal{M} , we have that $\delta(X_i, X_j) = L(\gamma_{ij})$.

In the Isomap algorithm, [TdSL00] propose to learn manifold topology from a graph connecting k -nearest neighbors for a given integer k . In the same way, our purpose is to approximate the geodesic distance δ with a suitable graph connecting nearest neighbors. Our approximation is carried out in three steps. Thereafter, we denote g_{ij} a path connecting two points X_i and X_j on a given graph, and $L(g_{ij})$ the length of such a graph.

Step 1. Consider $K = (\mathcal{E}, E)$ the complete Euclidean graph associated to sample \mathcal{E} , that is the graph made with all the points of the sample \mathcal{E} as vertices, and with edges $E = \{\{X_i, X_j\}, i = 1, \dots, n-1, j = i+1, \dots, n\}$. For an Euclidean graph, the edge weights are the edge lengths, that is, the Euclidean distances between each pair of points.

Step 2. Let $T = (\mathcal{E}, E_T)$ be the Euclidean Minimum Spanning Tree (EMST) associated to K , that is, the spanning tree that minimizes

$$\sum_{\{X_i, X_j\} \in E_T} d(X_i, X_j).$$

We point out that, the underlying idea in this construction is that, if two points X_i and X_j are relatively close, then we have that $\delta(X_i, X_j) \approx d(X_i, X_j)$. This may not be true if the manifold is very twisted and may induce bad approximations, hence the algorithm will produce a good approximation for relatively regular manifolds. It also requires a large number of sampling points on the manifold in order to guarantee the quality of this approximation. This drawback is well known when dealing with graph based approximation of the geodesic distance. Then, the graph T is a connected graph spanning K which mimics the manifold \mathcal{M} . Furthermore, an approximation of the geodesic distance $\delta(X_i, X_j)$ is provided by the sum of all the euclidean distance of the edges of the shortest path on T connecting X_i to X_j , namely

$$\hat{\delta}(X_i, X_j) = \min_{g_{ij} \in T} L(g_{ij}).$$

However, this approximation is too sensitive to perturbations of the data, and hence, very unstable. To cope with this problem, we propose to add more edges between the data to add extra paths in the data sample and thus to increase stability of the estimator. The idea is that paths which are close to the ones selected in the construction of the EMST could provide alternate ways of connecting the edges. Close should be here understood as lying in balls around the observed points. Hence, these new paths between the data are admissible and should be added to the edges of the graph. This provides redundant information but also stabilizes the constructed distance, and may also provide an answer to the the main defect of the algorithm that considers that two points very close with respect to the Euclidean distance are also close with respect to the geodesic distance.

Step 3. Let $B(X_i, \epsilon_i) \subset \mathbb{R}^p$ the open ball of center X_i and radius ϵ_i defined by $\epsilon_i = \max_{\{X_i, X_j\} \in E_T} d(X_i, X_j)$. Let graph $K' = (\mathcal{E}, E')$ defined by

$$\{X_i, X_j\} \in E' \iff \overline{X_i X_j} \subset \bigcup_{i=1}^n B(X_i, \epsilon_i)$$

where $\overline{X_i X_j} = \{X \in \mathbb{R}^p, \exists \lambda \in [0, 1], X = \lambda X_j + (1 - \lambda)X_i\}$. Then, K' is the graph which gives rise to our estimator of the distance δ :

$$(3) \quad \hat{\delta}(X_i, X_j) = \min_{g_{ij} \in K'} L(g_{ij}).$$

Hence, $\hat{\delta}$ is the distance associated with K' , that is, for each pair of points X_i and X_j , we have $\hat{\delta}(X_i, X_j) = L(\hat{\gamma}_{ij})$ where $\hat{\gamma}_{ij}$ is the minimum length path between X_i and X_j associated to K' .

We note that, the 3-steps procedure described above contains widespread graph-based methods to achieve our purpose. In this article, our graph-based calculations, such as MST estimation or shortest path calculus, were carried out with R Language [R D10] with the *igraph* library for network analysis [CN06].

An example of this 3-steps procedure and its behaviour when the number of observations increases are displayed respectively in Figure 1 and Figure 2. In Figure 1, points $(X_i^1, X_i^2)_i$ are simulated as follows :

$$(4) \quad X_i^1 = \frac{2i - n - 1}{n - 1} + \epsilon_i^1 \text{ and } X_i^2 = 2 \left(\frac{2i - n - 1}{n - 1} \right)^2 + \epsilon_i^2$$

where ϵ_i^1 and ϵ_i^2 are normally distributed with mean 0 and variance 0.01. In Figure 2, we give some results of graph K' for $n \in \{10, 30, 100, 300\}$. We can see in such a figure that graph K' tends to be close to the manifold $\{(t, t^2) \in \mathbb{R}^2, t \in \mathbb{R}\}$.

The main difference between our algorithm and the Isomap algorithm lies in the treatment of points which are far from the others. Indeed, the first step of the original Isomap algorithm consists in constructing the k -nearest neighbor graph or the ϵ -nearest neighbor graph for a given integer k or a real $\epsilon > 0$. Hence, points which are not connected to the biggest graph, since they are too distant, are not used for the construction of the estimated distance. Such a step is not present in our algorithm since in the applications we consider a distant point is not always an outlier. Hence, we do not exclude any points, and rather, for the construction of the EMST, all points of the data set are connected. Moreover, the Isomap algorithm requires the choice of parameters which are closely related to the local curvature of the manifold (see, for instance, [BS02]). This involves a heavy computing phase which is crucial for the quality of the construction, while, in our version we tend to give an automatic selection of parameters.

In the following section, we present a new application of manifold learning to the curve alignment problem.

3. Application to curve alignment. Consider a function $f : \mathbb{R} \rightarrow \mathbb{R}$, which will be the pattern to be recovered, observed in a translation effect framework. Let A be a real valued random variable with unknown distribution on an interval $(b, c) \subset \mathbb{R}$. The observation model is defined by

$$(5) \quad X_i^j = f(t_j - A_i), \quad i \in \{1, \dots, n\}, \quad j \in \{1, \dots, m\},$$

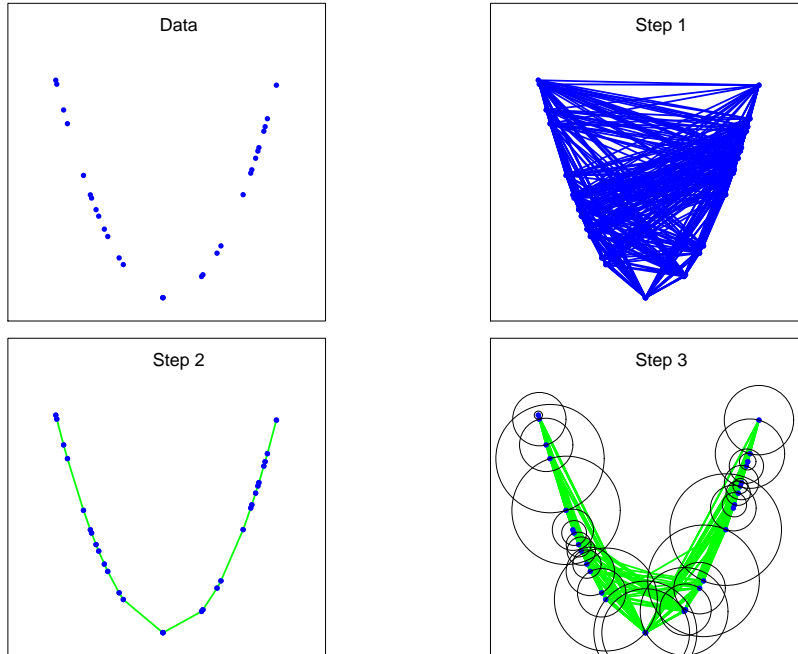


FIG 1. Construction of a subgraph K' from Simulation (4) with $n = 30$ points. On the top left, a simulated data set. On the top right, the associated complete Euclidean graph K (Step 1). On the bottom left, the EMST associated with the complete graph K (Step 2). On the bottom right, the associated open balls and the corresponding subgraph K' (Step 3).

where $(A_i)_i$ are i.i.d random variables drawn with distribution A which model the unknown translation parameters, while $(t_j)_j \in \mathbb{R}^m$ stand for the measurement points.

This situation usually happens when individuals experience similar events, which are explained by a common pattern f , and when the starting times of the events are not synchronized. Such a model has been studied, for instance, in [Sil95] and in [Røn01]. This issue has also received a specific attention in a semi-parametric framework in [GLM07] or [CL09]. In these works, among others, shift parameters are estimated, which enables to align the observations and thus to get rid of the translation issue. Model (5) falls also under the generic warping model proposed in [Maz06] and in [DLM11] which purpose is to estimate the underlying *structure* of the curves. For this, the authors define the *structural median* f_{SM} of the data. In the case of translation effects, it corresponds to

$$(6) \quad f_{SM} = f(\cdot - \text{med}(A))$$

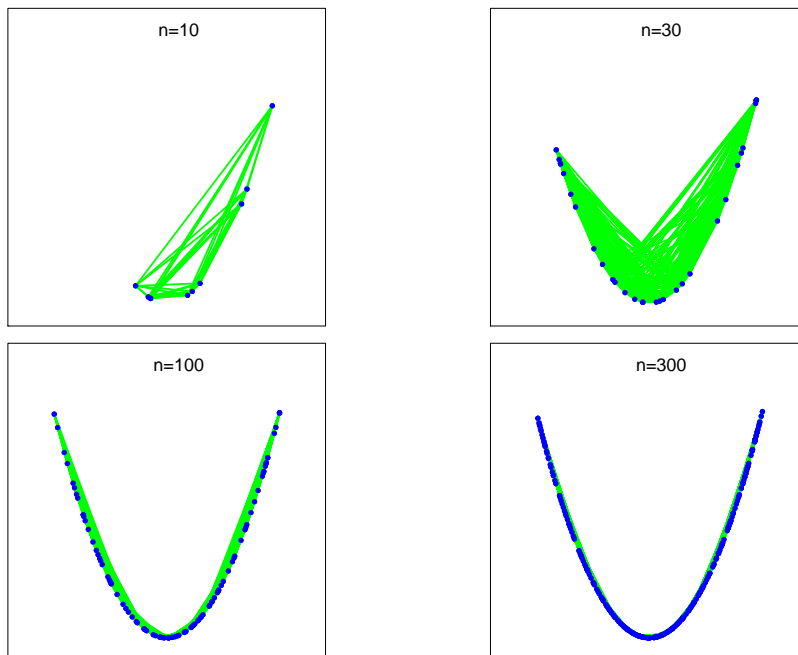


FIG 2. Evolution of graph K' for Simulation (4) and $n \in \{10, 30, 100, 300\}$.

with $\text{med}(A)$ the median of A . Hence, a natural estimator of the structural median f_{SM} , related to Model (5), would be

$$(7) \quad \widehat{f}_{\text{SM}} = \left(f \left(t_1 - \widehat{\text{med}}(A) \right), f \left(t_2 - \widehat{\text{med}}(A) \right), \dots, f \left(t_m - \widehat{\text{med}}(A) \right) \right)$$

with $\widehat{\text{med}}(A)$ the median of sample $(A_i)_i$. However, we first note that the translation parameters $(A_i)_i$ are not observed, and, as a consequence, that the median $\widehat{\text{med}}(A)$ can not directly be calculated. Then, the function f is also unknown, so, estimating $\widehat{\text{med}}(A)$ is not enough to calculate \widehat{f}_{SM} . Our purpose is to show that our manifold point of view provides a direct estimate of f_{SM} without the prior estimation of $\text{med}(A)$.

In order to use the manifold embedding approach, define

$$\begin{aligned} X : \mathbb{R} &\rightarrow \mathbb{R}^m \\ a &\mapsto X(a) = (f(t_1 - a), f(t_2 - a), \dots, f(t_m - a)) \end{aligned}$$

and set

$$\mathcal{C} = \{X(a) \in \mathbb{R}^m, a \in \mathbb{R}\}.$$

As soon as $f' \neq 0$, the map $X : a \mapsto X(a)$ provides a natural parametrization of \mathcal{C} which can thus be seen as a submanifold of \mathbb{R}^m of dimension 1. The corresponding geodesic distance is given by

$$\delta(X(a_1), X(a_2)) = \left| \int_{a_1}^{a_2} \|X'(a)\| da \right|.$$

The observation model (5) can be seen as a discretization of the manifold \mathcal{C} for different values $(a_i)_i$. Finding the median of all the shifted curves can hence be done by understanding the *geometry* of space \mathcal{C} , and thus approximating the geodesic distance between the curves.

The following theorem states that the structural median defined by Equation (7) is equivalent to the median with respect to the geodesic distance on \mathcal{C} , which provides a geometrical interpretation of the structural median.

THEOREM 1. *Recall the definition of the intrinsic median: $\hat{\mu}_I^1 = \arg \min_{\mu \in \mathcal{C}} \sum_{i=1}^n \delta(X_i, \mu)$. Under the assumption that $f' \neq 0$, we get that*

$$\hat{\mu}_I^1 = \hat{f}_{\text{SM}}.$$

Previous theorem can be extended to the more complex case of parametric deformations of the type

$$\begin{aligned} X : \mathbb{R}^3 &\rightarrow \mathbb{R}^m \\ (a, b, c) &\mapsto X(a, b, c) = (af(t_1 - b) + c, \dots, af(t_m - b) + c) \end{aligned}$$

as soon as $a \neq 0$ and $f' \neq 0$. Such a model has been described in [Vim10] and in [BLV10]. In this case, the submanifold is obviously of dimension 3.

In an even more general framework, when the observations can be modeled by a set of curves warped one from another by an unobservable deformation process, this estimate enables to recover the main pattern. It relies on the assumption that all the data belong to a manifold whose geodesic distance can be well approximated by the graph structure of the modified minimal spanning tree described in Section 2.

Finally, we propose the following estimator of the structural median

$$(8) \quad \tilde{\mu}_I^1 = \arg \min_{\mu \in \mathcal{E}} \sum_{i=1}^n \hat{\delta}(X_i, \mu),$$

using the geodesic distance $\hat{\delta}$, estimated by the algorithm described in Section 2. The numerical properties of this estimator is studied using simulations in Section 4, and for real data sets in Section 5.

4. Simulations. We consider the target function $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(t) = t \sin(t)$. We simulate deformations of this function on $j = 1, \dots, m = 100$ equally distributed points t_j of the interval $[-10, 10]$, according to the following model :

$$(9) \quad Y_i(t_j) = A_i f(B_i t_j - C_i) \quad i = 1, \dots, n, \quad j = 1, \dots, m$$

where $(A_i)_i$, and $(C_i)_i$ are i.i.d uniform random variables on $[-10, 10]$ while $(B_i)_i$ are an i.i.d sample of a uniform distribution on $[-1, 1]$. We finally obtain a data set of $n = 100$ curves where each differs from the initial function f by a translation and an amplitude deformation. The data is displayed on the left graph of Figure 4.

We then consider three estimators of the function f . The first one, which minimizes the approximated geodesic distance, defined by Equation (8), will be referred to as the structural median estimator. The second one is obtained by the Curve Alignment by Moments procedure (CAM) developed by [Jam07]. The last one is the mere mean of the data. We recall here that the CAM procedure consists on extracting the mean pattern by synchronization of the moments of the simulated curves. For this, he introduces the *feature function* concept for a given function g , defined as $I_g(t)$:

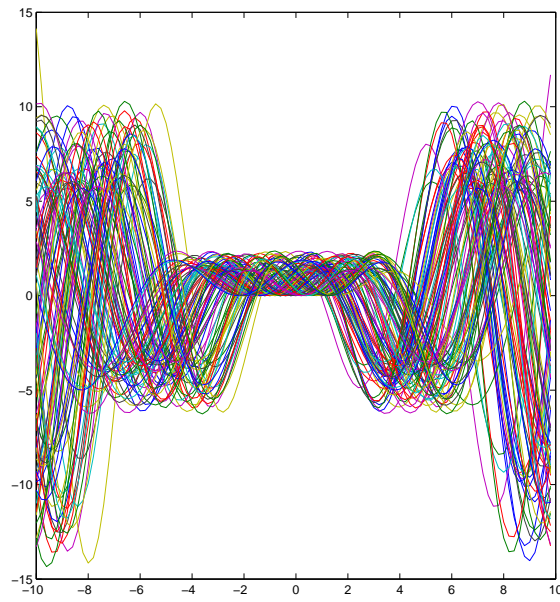
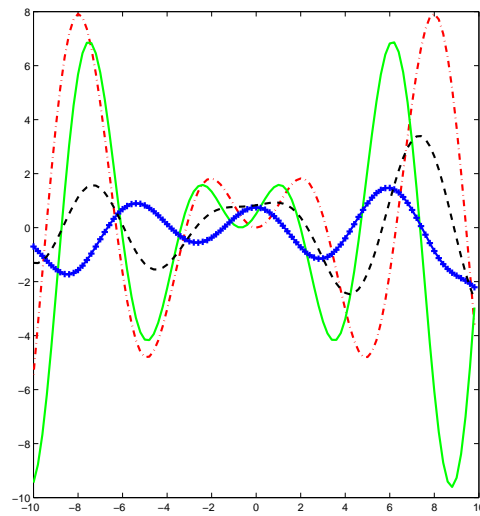
$$(10) \quad I_g(t) \geq 0 \text{ and } \int I_g(t) dt = 1$$

and the moments

$$(11) \quad \mu_g^{(1)} = \int t I_g(t) dt \text{ and } \mu_g^{(k)} = \int (t - \mu_g^{(1)})^k I_g(t) dt, \quad k \geq 2$$

Then, the CAM procedure is align the curves by warping their moments, for instance, the amplitude of the peaks, at the location they occur, the variance around these peaks, and so on. This method relies on the choice of a proper feature function, for instance $I_g^{(l)}(t) = \frac{|g^{(l)}(t)|}{\int |g^{(l)}(s)| ds}$ for a given $l \geq 0$, on an approximation of the functions by splines, and the selection of the number of moments to be synchronized. Hence, it highly depends on the choice of these tuning parameters. We have chosen the optimal value of the parameters over a grid.

These three estimators are shown on Figure 4. With the CAM or the mere mean procedure, the average curve does not reflect the structure of the initial curves, or the amplitude of their variations. On the contrary, the structural median extracted by Manifold Warping has the characteristics of the closest target curve, but is also its best approximation.

FIG 3. *Data simulated from Model (9)*FIG 4. *The target function f (red dotted line), an estimation of the structural median by Manifold Warping (green solid line), an estimation with mean (blue dashed line), and an estimation obtained with CAM procedure (black dashed line)*

5. Real data. Consider the real data case where an observation curve represents the reflectance of a particular landscape and fully characterizes the nature of this landscape. The purpose of this study is to predict the different landscapes while observing the reflectance profiles. In Figures 5 and 6, we present two data sets corresponding to reflectance patterns of two landscapes in the same region with the same period. However, the reflectance depends on the vegetation whose growth depends on the weather condition and the behavior in soil. It is therefore relevant to consider that these profiles are deformations in translation and/or amplitude of a single representative function of the reflectance behaviour of each landscape in this region at this time.

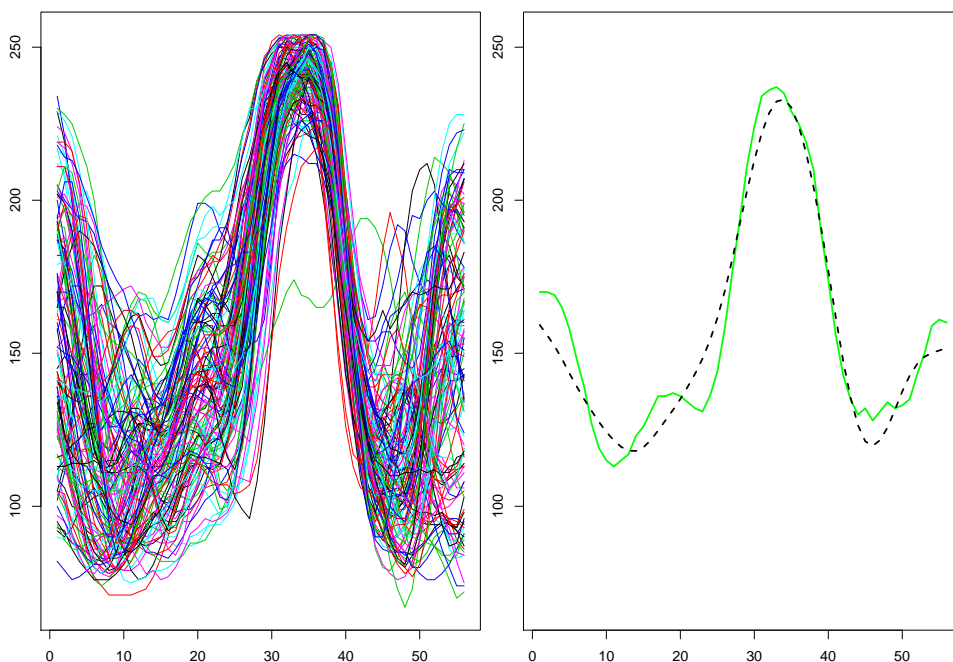


FIG 5. On the left, the first landscape data. On the right, the CAM representative estimation (black dashed line) and the Manifold Warping estimation (green solid line).

Our aim is to build a classification procedure. For this, we will use a labeled set of curves and extract from each group of similar landscape a representative profile. Then, we will allocate a new curve to the group whose representative curve will be the closest. That is the reason why it is important to obtain a pattern which captures the structure of the curves. We will use three different ways to get a representative group of curves, the mean

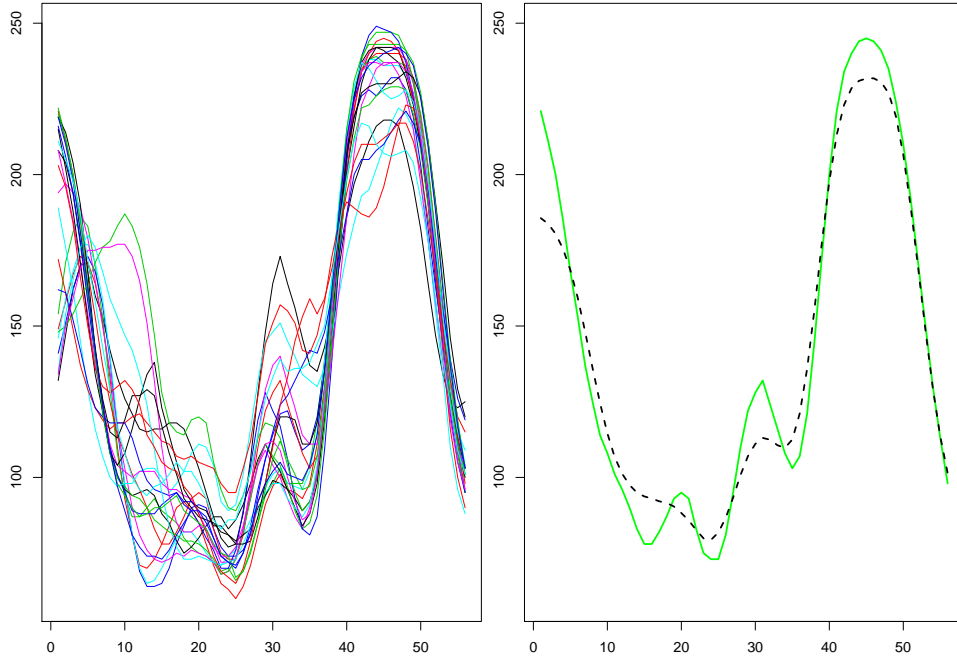


FIG 6. *On the left, the second landscape data. On the right, the CAM representative estimation (black dashed line) and the Manifold Warping estimation (green solid line).*

curve, the CAM method and our method, referred to as the manifold warping. We will compare their classification performance together with a usual classification procedure : the classical k -nearest neighbours.

In Figure 5, we observe that the CAM average oversmooths the peaks of activity at times 12 and 22 to make them almost nonexistent. This is a clear defect since, according to the experts of landscape remote sensing, these peaks of activity are representative of the nature of landscape. Indeed, these peaks convey essential informations which determines, among other things, the type of landscape. On the other hand, these changes are very well rendered by the pattern obtained by Manifold Warping. The same conclusions can be drawn in Figure 6 for an other landscape. In this application domain, extracting a curve by Manifold Warping is best able to report data as reflecting their structure and thus to obtain a better representative.

So, if we try to identify "unknown" landscapes by classifying each curve. We find the nearest representative by Euclidean distance, we obtain the confusion matrices displayed in Table 1. Note that here we have sought to classify the two landscapes plots at the earliest in the season. To benchmark

our procedure, we compare our performance to the method of the k -nearest neighbors classification. We classified our data by Euclidean distance in our initial data sets and we get the confusion matrix of Table 2. We get a much better discrimination of landscapes, working on parts of curves, with the method of estimating a representative by Manifold Warping than by the CAM method or by classical mean.

Pixel	Manifold classification		CAM classification	
Reference	Landscape1	Landscape2	Landscape1	Landscape2
Landscape1	21	0	12	9
Landscape2	1	19	1	19

TABLE 1

Manifold Warping and CAM confusion matrices.

Pixel	Mean classification		k -nn classification	
Reference	Landscape1	Landscape2	Landscape1	Landscape2
Landscape1	12	9	15	6
Landscape2	0	20	2	18

TABLE 2

Classical mean and k -nearest neighbors confusion matrices.

6. Conclusion. By using an ISOMAP inspired strategy, we have extracted from a pattern of curves, a curve which, playing the role of the mean, serves as a pattern conveying the information of the data. In some cases, in particular when the structure of the deformations entails that the curve can be embedded into a manifold regular enough, we have shown that this corresponds to finding the structural expectation of the data, developed in [DLM11], which improves the performance of other *mean* extraction methods. This enables to derive a classification strategy that assigns a curve to the group, whose representative curve is the closest, with respect to the chosen distance. Of course, the performance of this allocation rule deeply relies on the good performance of the pattern extraction.

One of the major drawbacks of this methodology are that first a high number of data are required in order to guarantee a good approximation of the geodesic distance at the core of this work. Actually, note that the number of observations, i.e the sampling rate of the manifold highly depends on the regularity of the manifold such that the assumption that the euclidean path between two observations follow approximatively the geodesic path. Hence, the data set should be carefully chosen for the manifold to be smooth enough. We point out that an enhancement could come from a prior registration procedure first applied to the curve and then the manifold

warping procedure applied to the registered data. The second drawback which may also be viewed as an advantage, is the following : the extracted pattern is a curve that belong to the observations. One the one hand, it may contains noise if the data are noisy observations but on the other hand it thus guarantees that the pattern shares the mean properties and specifies of the observations. A solution when the noise must be removed is either to directly smooth the resulting pattern or to consider the neighbourhood of the extracted pattern with respect to the approximated geodesic distance and then use a kernel estimator with these observations to obtain a regularized *mean curve*.

Nevertheless, we promote this procedure when a large amount of data are available and when the sets of similar curves share a common behaviour which fully characterizes the observations, coming from an economic, physical or biological model for instance. This methods has been applied with success to a large amount of cases. Numerical packages for R or Matlab are available on request.

7. Appendix.

PROOF OF THEOREM 1. Take $\mu = X(\alpha)$ with $\alpha \in]b, c[$, we can write

$$\begin{aligned} \hat{\mu}_I^1 &= \arg \min_{X(\alpha) \in \mathcal{C}} \sum_{i=1}^n \delta(X(A_i), X(\alpha)) \\ &= \arg \min_{X(\alpha) \in \mathcal{C}} \sum_{i=1}^n D(A_i, \alpha) = \arg \min_{X(\alpha) \in \mathcal{C}} C(\alpha) \end{aligned}$$

where D is the following distance on $]b, c[$:

$$D(A_i, \alpha) = \left| \int_{A_i}^{\alpha} \|X'(a)\| da \right|.$$

In the following, let $(A_{(i)})_i$ the ordered parameters. That is

$$A_{(1)} < A_{(2)} < \dots < A_{(n)}.$$

Then, for a given $\alpha \in]b, c[$ such that $A_{(j)} < \alpha < A_{(j+1)}$, we get that

$$\begin{aligned} C(\alpha) &= jD(\alpha, A_{(j)}) + \sum_{i=1}^{j-1} iD(A_{(i)}, A_{(i+1)}) \\ &+ (n-j)D(\alpha, A_{(j+1)}) + \sum_{i=j+1}^{n-1} (n-i)D(A_{(i)}, A_{(i+1)}). \end{aligned}$$

For the sake of simplicity, let $n = 2q + 1$. It follows that $\widehat{\text{med}}(A) = A_{(q+1)}$. Moreover, let $\alpha = A_{(j)}$ with $j < q + 1$. By symmetry, the case $j > q + 1$ will hold. Then, we rewrite $C(\alpha)$ as

$$C(\alpha) = \sum_{i=1}^{j-1} iD(A_{(i)}, A_{(i+1)}) + \sum_{i=j}^{n-1} (n-i)D(A_{(i)}, A_{(i+1)})$$

and, by introducing $A_{(q+1)}$, we get that

$$\begin{aligned} C(\alpha) &= \sum_{i=1}^{j-1} iD(A_{(i)}, A_{(i+1)}) + \sum_{i=j}^q iD(A_{(i)}, A_{(i+1)}) \\ &+ \sum_{i=j}^q (n-2i)D(A_{(i)}, A_{(i+1)}) + \sum_{i=q+1}^{n-1} (n-i)D(A_{(i)}, A_{(i+1)}). \end{aligned}$$

Finally, we notice that

$$C(\alpha) = C(A_{(q+1)}) + \sum_{i=j}^q (n-2i)D(A_{(i)}, A_{(i+1)}) > C(A_{(q+1)}).$$

And the result follows since

$$\widehat{\mu}_I^1 = \arg \min_{X(\alpha) \in \mathcal{C}} C(\alpha) = X(A_{(q+1)}) = X(\widehat{\text{med}}(A)) = \widehat{f}_{\text{SM}}.$$

□

References.

- [BLV10] Jérémie Bigot, Jean-Michel Loubes, and Myriam Vimond. Semiparametric estimation of shifts on compact lie groups for image registration. *Probab. Theory Relat. Fields*, Published online, 2010.
- [BP03] Rabi Bhattacharya and Vic Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.*, 31(1):1–29, 2003.
- [BS02] Mukund Balasubramanian and Eric L. Schwartz. The isomap algorithm and topological stability. *Science*, 295, 2002.
- [CL09] I. Castillo and J.-M. Loubes. Estimation of the distribution of random shifts deformation. *Math. Methods Statist.*, 18(1):21–42, 2009.
- [CN06] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal, Complex Systems*:1695, 2006.
- [DLM11] Jean-Franois Dupuy, Jean-Michel Loubes, and Elie Maza. Non parametric estimation of the structural expectation of a stochastic increasing function. *Stat Comput*, 21(1):121–136, 2011.
- [dST03] Vin de Silva and Joshua B. Tenenbaum. Unsupervised learning of curved manifolds. In *Nonlinear estimation and classification (Berkeley, CA, 2001)*, volume 171 of *Lecture Notes in Statist.*, pages 453–465. Springer, New York, 2003.

