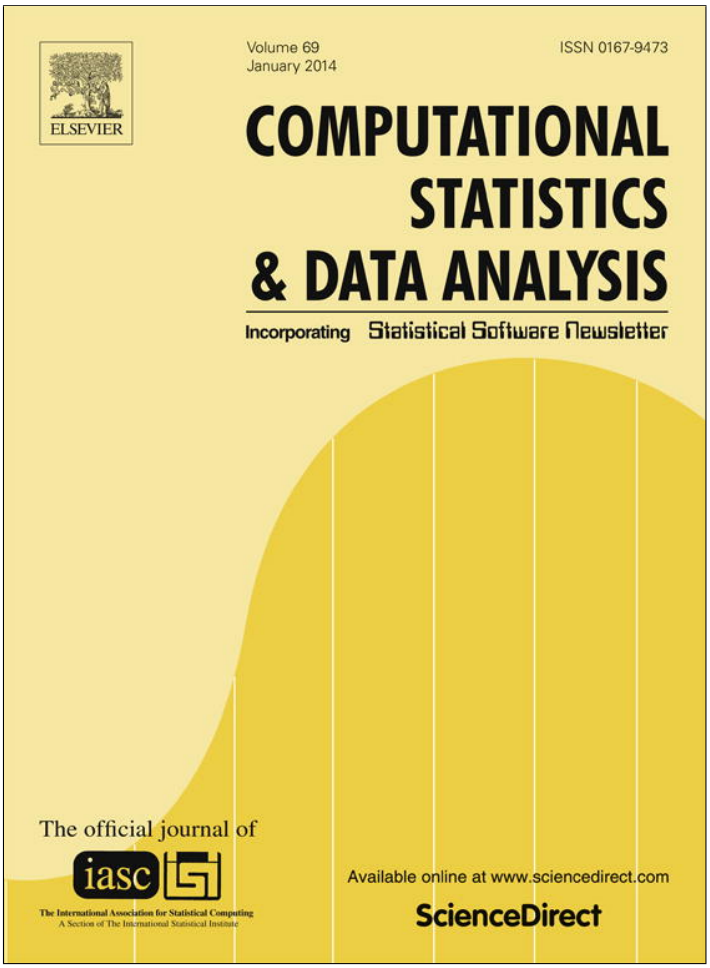

Computational Statistics and Data Analysis 69 (2014) 255–268

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Pairwise dynamic time warping for event data

Ana Arribas-Gil [a,*], Hans-Georg Müller [b]

[a] Departmento de Estadística, Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain
[b] Department of Statistics, University of California, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 7 November 2012
Received in revised form 20 August 2013
Accepted 22 August 2013
Available online 30 August 2013

Keywords:
Alignment
Birth data
French-Canadian fertility
Functional data analysis
On-line auctions
Point process
Registration

ABSTRACT

A new version of dynamic time warping for samples of observed event times that are modeled as time-warped intensity processes is introduced. The approach is developed within a framework where for each experimental unit or subject in a sample, a random number of event times or random locations can be observed. As in this setting the number of observed events differs from subject to subject, usual landmark alignment methods that require the number of events to be the same across subjects are not feasible. This challenge is addressed by applying dynamic time warping, initially by aligning the event times for pairs of subjects, regardless of whether the numbers of observed events within the considered pair of subjects match. The information about pairwise alignments is then combined to extract an overall alignment of the events for each subject across the entire sample. This overall alignment provides a useful description of event data and can be used as a pre-processing step for subsequent analysis. The method is illustrated with a historical fertility study and with on-line auction data.

1. Introduction

Time warping is commonly used in functional data analysis to address the presence of random variation in time in addition to amplitude variation. A major difficulty is the non-identifiability of these two components (Liu and Müller, 2004). This non-identifiability requires that one makes strong assumptions on the nature of the warping mechanism in order to be able to distinguish it from the variation in the random amplitudes. While various solutions have emerged over the years, a gold standard for functional time warping is the landmark method (Kneip and Gasser, 1992; Gasser and Kneip, 1995). In this method, one typically extracts for each random curve X_i in a sample of i.i.d. trajectories $X_1, \ldots, X_n$ a series of landmarks, which correspond to the times or locations where certain features occur, such as the location of a zero or the location of a peak.

One then uses these locations to represent the sample or for subsequent continuous time warping of the random functions, by moving the landmark locations to their cross-sample average locations. Then one can smoothly map the time intervals in between these locations, for example by applying monotone spline transformations. Usually, the landmarks are obtained by presmoothing the often noisily observed functions X_i, obtaining estimates $\hat{X}_i$ and then extracting features such as peak locations by substituting the unknown peak location $\theta_i = \operatorname{argmax}_x X_i(x)$ by the estimated peak location $\hat{\theta}_i = \operatorname{argmax}_x \hat{X}_i(x)$, where a peak is just one of many possible features defining a landmark (Gasser et al., 1984b).

A persistent difficulty with the landmark method has been that in real data applications, landmarks such as peaks can be hard to identify for specific sample curves, due to either noise in the measurements or to the presence of non-standard trajectories, which do not exhibit some of the landmarks or have additional landmarks that do not match with a standard

* Correspondence to: Calle Madrid 126, 28903 Getafe, Spain. Tel.: +34 916245853; fax: +34 916249848.
E-mail address: ana.arribas@uc3m.es (A. Arribas-Gil).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.08.011

set of landmarks. In the analysis of growth curves, for example, one might encounter more than one mid-growth spurt and it is then hard to decide how to position extra peaks within the sequence of landmarks, as the landmark method however requires that the number of identified landmarks matches across all curves and also that the landmarks are in a sequence that is equally interpretable across all subjects in terms of features (often such features are the location of peaks and of zeros and size of peaks for functions and derivatives). In addition, although seemingly straightforward, the computational task of landmark extraction is actually quite burdensome and usually cannot be fully automated, so that identifying and verifying a large number of landmarks even for the modest samples of curves can be a major effort that includes some subjective elements (Gasser et al., 1984a).

For these reasons, non-landmark based methods for function warping have met with increasing interest in recent years in the functional data analysis literature (Silverman, 1995; Wang and Gasser, 1997; Aach and Church, 2001; Rønn, 2001; Gervini and Gasser, 2004; Kneip and Ramsay, 2008; Telesca and Inoue, 2008). As already mentioned, such methods depend explicitly or implicitly on strong assumptions to circumvent the non-identifiability problem. In this paper, we take a different approach. We assume from the beginning that the number of landmarks that can be identified in sample trajectories is random, but that nevertheless their order matters, so that the lower ranked landmarks should be placed closer together in time than the higher ranked landmarks, irrespective of the number of landmarks that are available for each subject.

More generally, the times at which landmarks occur may be viewed as random event times that are observed for each subject. These event times carry information about their order and relative location to each other, but otherwise are not distinguishable. Typical examples for such event data that we use to illustrate our methods are age at child birth per woman in a large French-Canadian historical cohort, where women typically have a large number of offspring, and the timings of bids submitted at online ebay auctions, recorded for each auction in a sample of auctions. For general functional methodology for functional data that correspond to intensities of point processes we refer to Bouzas et al. (2006), Wu et al. (2013).

The problem of aligning event-type data is inherently difficult, because (1) the number of events usually varies across subjects or units, so that landmark methods are not applicable; and (2) the number of events observed per unit or subject is often small. In the historical fertility data that we discuss in Section 4.1 below, many women in this sample have one or two children, so there are subjects for whom only very few birth events are observed. This also means that methods such as quantile normalization (Bolstad et al., 2003; Gallon et al., 2013) or density synchronization (Zhang and Müller, 2011) that assume knowledge about the specific densities for each subject or unit, and at the minimum require large numbers of observations per unit or subject to infer these densities reliably, are not applicable to analyze the kind of event data that we study in this paper.

To the best of our knowledge, there does not yet exist a satisfactory method to deal with typical event data that vary substantially from subject to subject and can be sparse. We therefore believe that the proposed algorithm is innovative in the sense that it addresses data that arise in practice and for which no good solution currently exists. The proposed algorithm provides an efficient solution and we demonstrate its application to real data examples.

Specifically, denote the n_i observed event times for the i-th subject by $t_{i1}, \ldots, t_{in_i}$, and assume that all event times are observed within the interval $\mathcal{T} = [0, T]$. We consider standardized cumulative incidence functions

$$Y_i(t) = \frac{1}{n_i} \sum_{k=1}^{n_i} 1_{\{t_{ik} \le t\}}, \quad t \in \mathcal{T},\ i = 1, \ldots, n,$$

and assume these are generated by an underlying monotone increasing fixed function μ and warping functions h_i, which are i.i.d. realizations of a random variable H taking values on $\mathcal{W}(\mathcal{T}) = \{f \in \mathcal{C}(\mathcal{T}) \mid f \text{ strictly increasing}, f(0) = 0, f(T) = T\}$, such that

$$Y_i(t_{ik}) = \mu(h_i^{-1}(t_{ik})), \quad t_{ik} \in \mathcal{T},\ i = 1, \ldots, n,\ k = 1, \ldots, n_i. \tag{1}$$

Here, $\mu(t)$ is a cumulative distribution function and the constraints on H ensure that $\mu(h_i^{-1}(t))$ also are (random) cumulative distribution functions. This model can be easily extended to the case in which the observed functions represent the absolute number of events, that is $Y_i(t_{ik}) = n_i \mu(h_i^{-1}(t_{ik}))$, or more generally to situations where $Y_i(t_{ik}) = a_i \mu(h_i^{-1}(t_{ik}))$, with a_i i.i.d. unknown random factors. In this framework, registration procedures are defined by estimates of the registration functions h_i. The determination of the time warping functions h_i is of intrinsic interest and also allows to determine the unwrapped versions $X_i(t) = Y_i(h_i(t)), t \in \mathcal{T}$, of the Y_i. For identifiability purposes, it is expedient to assume that $\mathbb{E}[H(t)] = I_{\mathcal{T}}(t)$, where $I_{\mathcal{T}}$ is the identity function on the domain $\mathcal{T}$.

2. Methodology

2.1. Pairwise registration

For pairwise registration, we adopt the approach proposed by Rong and Müller (2008). The idea is to obtain a global or overall alignment for a sample of random functions, from the information about pairwise alignments. Using this device, the task of global registration is reduced to the often more manageable task of constructing pairwise or "relative" alignments between pairs of random trajectories.

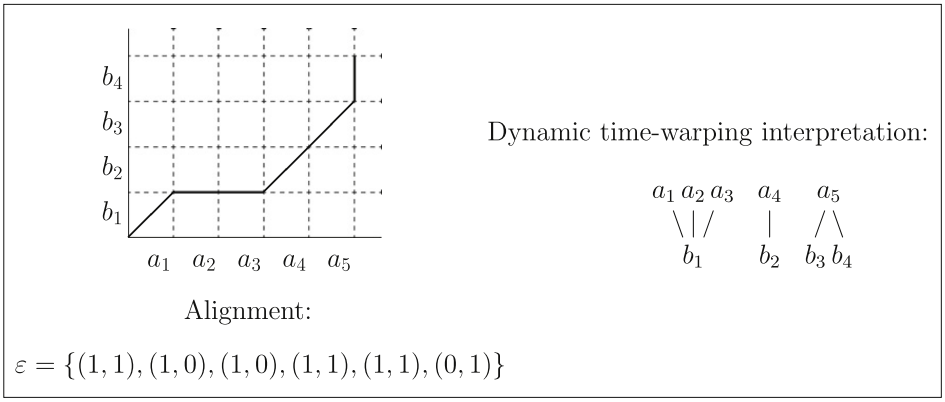

Fig. 1. Graphical representation of a dynamical time warping alignment between two sequences $\mathbf{a} = (a_1, \ldots, a_\ell)$ and $\mathbf{b} = (b_1, \ldots, b_m)$. Lines represent the correspondence between elements of the two sequences.

The initial goal is thus to obtain pairwise synchronizations for all or many pairs of curves Y_i and Y_j, as defined in (1). For this purpose, Rong and Müller (2008) define pairwise synchronization functions

$$g_{ji}(t) = h_j(h_i^{-1}(t)), \tag{2}$$

which provide a transformation of the time scale of the curve Y_j towards that of Y_i, and show that $\mathbb{E}[h_j(h_i^{-1}(t))|h_i^{-1}(t)] = h_i^{-1}(t)$. This motivates the estimators

$$\hat{h}_i^{-1}(t) = \frac{1}{n}\sum_{j=1}^{n} \hat{g}_{ji}(t). \tag{3}$$

The alignment problem is then reduced to the problem of defining appropriate estimators $\hat{g}_{ji}(t)$, that properly reflect the particular nature of the data, in our case event data. While Rong and Müller (2008) proposed the minimization of a penalized L^2-distance to obtain appropriate estimates of $g_{ji}(t)$, in the situation we study here, dynamic time warping (DTW) provides a promising alternative.

2.2. Dynamic time warping

Dynamic time warping (DTW) is a series of alignment algorithms originally developed in the 1970s in the context of speech recognition (see Kruskal and Liberman, 1983 for an overview). These are dynamic programming algorithms that consist in aligning two sequences of outcomes by minimizing the total distance between them, computed as the sum of distances between each pair of points along the aligned positions in a suitable metric. Dynamic time warping is similar to the algorithms used for the alignment of biological sequences, such as the Needleman–Wunsch algorithm (Needleman and Wunsch, 1970).

In this context, given two sequences $\mathbf{a} = (a_1, \ldots, a_\ell)$ and $\mathbf{b} = (b_1, \ldots, b_m)$, with values on a certain feature space $\mathcal{A}$, an alignment of $\mathbf{a}$ and $\mathbf{b}$ is a sequence $\{\varepsilon_k\}_{k=1}^{|\varepsilon|}$ of variable length $|\varepsilon|$, taking values on $\{(0,1),(1,0),(1,1)\}$ such that $\sum_{k=1}^{|\varepsilon|}\varepsilon_k = (\ell, m)$. That is, $(L_r, M_r) = \sum_{k=1}^{r}\varepsilon_k$, $r = 1, \ldots, |\varepsilon|$, defines a path from $(0,0)$ to (ℓ, m) in the two-dimensional integer lattice. We say that a_i is aligned to b_j if there exists an $1 \le r \le |\varepsilon|$ such that $\sum_{k=1}^{r}\varepsilon_k = (i,j)$. Thus in dynamic time warping one allows one element in one sequence to be aligned to more than one element in the other sequence, therefore this method is particularly suitable for sequences of event times with unequal number of events in each sequence. The general approach is illustrated in Fig. 1.

In our setting of event data, we observe pairs of curves $\mathbf{a} = (Y_i(t_{i1}), \ldots, Y_i(t_{in_i}))$ and $\mathbf{b} = (Y_j(t_{j1}), \ldots, Y_j(t_{jn_j}))$ at event times $\mathbf{t} = (t_{i1}, \ldots, t_{in_i})$ and $\mathbf{s} = (t_{j1}, \ldots, t_{jn_j})$, respectively. Then, to align curves Y_i and Y_j one needs to determine, first, which elements of the sequence $\mathbf{a}$ should be mapped to which elements of the sequence $\mathbf{b}$, and second, the time points at which these aligned values should be placed. For the first determination, a discrete DTW approach, which is described next, is used. The second determination is addressed in Section 2.3.

In the following we use the notation $\mathbf{a} = (a_1, \ldots, a_\ell)$ and $\mathbf{b} = (b_1, \ldots, b_m)$ to refer to the observed values, without any temporal reference to any pair of curves. We now describe the DTW algorithm that allows us to find an optimal mapping between these two sequences.

Let $d : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ be a suitable distance function defined on the feature space. Then, for any alignment ε, its alignment distance is given by

$$D_\varepsilon(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^{|\varepsilon|} w(k) \cdot d(a_{L_k}, b_{M_k}), \tag{4}$$

where $w(\cdot)$ represents time weights that account for the length of the alignment. Different definitions of these weights are possible, even $w(\cdot) = 1$ (see Wang and Gasser, 1997 for the study of weight functions in the continuous time DTW problem). In this work we consider the definition given by Kruskal and Liberman (1983), $w(k) = (t_{L_k} - t_{L_{k-1}} + s_{M_k} - s_{M_{k-1}})/2$, which is quite standard in the DTW literature.

Then, the optimal alignment is

$$\varepsilon^{ab} = \underset{\varepsilon \in \mathcal{E}_{\ell,m}}{\arg\min}\, D_\varepsilon(\mathbf{a}, \mathbf{b}), \tag{5}$$

where $\mathcal{E}_{\ell,m}$ is the set of all possible alignments of sequences with lengths ℓ and m. The dynamic time-warping algorithm used to retrieve ε_{ab} is as follows:

1. Initialize:

$$\begin{aligned}
&D_{(1,1)} = 0, \quad I_{(1,1)} = (1,1),\\
&D_{(i,1)} = D_{i-1,1} + d(a_i, b_1)\cdot(t_i - t_{i-1})/2, \qquad I_{(i,1)} = (1,0), \quad i = 2,\ldots,\ell,\\
&D_{(1,j)} = D_{1,j-1} + d(a_1, b_j)\cdot(s_j - s_{j-1})/2, \qquad I_{(1,j)} = (0,1), \quad j = 2,\ldots,m.
\end{aligned}$$

2. Iterate for $i = 2,\ldots,\ell, j = 2,\ldots,m$:

$$D_{(i,j)} = \min \begin{cases} D_{(i-1,j-1)} + d(a_i, b_j)\cdot \dfrac{(t_i - t_{i-1} + s_j - s_{j-1})}{2} \\ D_{(i-1,j)} + d(a_i, b_j)\cdot \dfrac{(t_i - t_{i-1})}{2}\cdot \mathbb{1}I_{(i-1,j)} \neq (0,1) + M\mathbb{1}I_{(i-1,j)} = (0,1) \\ D_{(i,j-1)} + d(a_i, b_j)\cdot \dfrac{(s_j - s_{j-1})}{2}\cdot \mathbb{1}I_{(j-1,i)} \neq (1,0) + M\mathbb{1}I_{(j-1,i)} = (1,0) \end{cases}$$

$$I_{(i,j)} = \begin{cases} (1,1) & \text{if } D_{(i,j)} = D_{(i-1,j-1)} + d(a_i, b_j)\cdot \dfrac{(t_i - t_{i-1} + s_j - s_{j-1})}{2} \\ (1,0) & \text{if } D_{(i,j)} = D_{(i-1,j)} + d(a_i, b_j)\cdot \dfrac{(t_i - t_{i-1})}{2} \\ (0,1) & \text{if } D_{(i,j)} = D_{(i,j-1)} + d(a_i, b_j)\cdot \dfrac{(s_j - s_{j-1})}{2}. \end{cases}$$

3. $d_{\varepsilon^{ab}}(\mathbf{a}, \mathbf{b}) = D_{(\ell,m)}$.
4. Trace back: $\varepsilon^{ab}_{|\varepsilon^{ab}|} = I_{(\ell,m)}$, $i = \ell - I_{(\ell,m)}(1)$, $j = m - I_{(\ell,m)}(2)$, $k = |\varepsilon^{ab}|$ and while $i \geq 1$ and $j \geq 1$ do:

$$\begin{aligned}
&\varepsilon^{ab}_k = I_{(i,j)}, \quad k = k - 1,\\
&i = i - I_{(i,j)}(1), \quad j = j - I_{(i,j)}(2).
\end{aligned} \tag{6}$$

Note that in step 1, the initialization $D_{(1,1)} = 0$, $I_{(1,1)} = (1,1)$, forces the two first values of the two sequences to be aligned together. In step 2, we prevent having a vertical movement followed by an horizontal one, or vice versa, by choosing a sufficiently large constant M. This means that if for instance $a_i, \ldots, a_{i+k}$, $k > 0$, are all mapped to b_j, then b_{j+1} cannot be mapped again to a_{i+k}. As for the computational cost, it is $\mathcal{O}(\ell \cdot m)$, which is low, since ℓ and m, the number of events per curve, are usually quite small.

Once we obtain the optimal alignment, ε^{ab}, of sequences $\mathbf{a}$ and $\mathbf{b}$, what we get is a correspondence between values $a_1, \ldots, a_\ell$ and $b_1, \ldots, b_m$, without any timing reference. That is, at this stage we do not tackle the problem of designating the time points at which two (or more) aligned values should be placed. This is a major difference between our approach and previous work using DTW related algorithms in the context of time warping, as for example Wang and Gasser (1997, 1999). Indeed, these works deal with continuous features and time synchronization problems in which finding similarities between points in different curves and assigning temporal references cannot be tackled separately. In Bigot (2006), a dynamic programming algorithm similar to (6) is used to find a correspondence between two previously identified sets of landmarks from two curves. Then, a penalized least squares minimization problem is solved to estimate the warping functions that register the two curves through the alignment of their landmarks.

Our approach is quite different since we focus on the global alignment of the whole set of curves for which pairwise alignment maps constitute just a preliminary stage. That is, we use DTW to determine the pairwise correspondence of events, and this information about pairwise correspondences is then combined (see Section 2.3) to obtain the global alignment, i.e., to register the entire sample. This is a key difference from previous applications of DTW, as the main disadvantage of DTW based registration methods is that the generalization from two to a larger number of curves is not straightforward. Indeed, extensions to the alignment of a larger set of curves usually rely on defining a common set of landmarks that is obtained from individual landmark vectors or alternatively from a template curve to which then the entire sample of curves is aligned. However, this standard approach requires that a vector of well-defined landmarks or a template curve is available in the first place, and furthermore suffers from a loss of information and efficiency, as it requires computationally intensive algorithms.

In the last decade, DTW has been used for the alignment of expression profiles from time-course microarray experiments (Aach and Church, 2001; Clote and Straubhaar, 2006). In this context, a n-dimensional time series, containing the expression

levels of a group of n genes recorded at the same time points is aligned to another n-dimensional time series corresponding to a different microarray experiment of the same n genes. That is, DTW is used to perform the pairwise alignment of two n-dimensional sequences. This is different from using DTW to align n different trajectories. Indeed, although multiple dynamic time warping is theoretically possible and algorithmically equivalent to the pairwise case, it is computationally unfeasible even for a relatively small number of sequences, since its computing time is proportional to ℓ^n, where n represents here the total number of sequences and ℓ their length, assuming they all have similar lengths.

2.3. Pairwise dynamic time warping

The proposed strategy is to obtain the pairwise alignment maps in a first step, by combining DTW with pairwise registration. The goal is to extract information about global alignment from the pairwise correspondences. Then, given a sample of n curves, we aim at estimating $\hat{g}_{ij}(t)$ and $\hat{g}_{ji}(t)$ for any $i, j = 1, \ldots, n$. The observations in model (1) are (t_{ik}, y_{ik}), $k = 1, \ldots, n_i$, $i = 1, \ldots, n$, where t_{ik} is the time at which the k-th event is observed for individual i and y_{ik} is the observed value of Y_i at time t_{ik}.

In the following, we focus on a given pair of curves y_i and y_j, and for the sake of simplicity denote the corresponding observations as (t_k, y_{ik}), $k = 1, \ldots, n_i$, and (s_k, y_{jk}), $k = 1, \ldots, n_j$. Applying algorithm (6) we obtain an optimal mapping, ε^{ij}, of the values y_i and y_j. While it is one thing to decide which values on both sequences should be aligned together, it is another matter to determine the time points corresponding to the aligned records. Recalling that the alignment of pairs of curves here is just a preliminary step to arrive at a global alignment of the entire sample of curves and is not an objective by itself, it is useful to note that from the estimated pairwise alignments one also obtains $\hat{g}_{ji}(t_k) = s_h$ and $\hat{g}_{ij}(s_h) = t_k$ for matched pairs of values y_{ik} and y_{jh}. However, if more than one value in one curve is mapped to the same single value in the other curve, additional considerations are needed for the definition of $\hat{g}_{ji}(t)$ and $\hat{g}_{ij}(t)$.

We now describe the detailed procedure to obtain $\hat{g}_{ij}(t)$ and $\hat{g}_{ji}(t)$ from observed sequences (t_k, y_{ik}), $k = 1, \ldots, n_i$, and (s_k, y_{jk}), $k = 1, \ldots, n_j$:

- find ε^{ij}, the optimal alignment between the two sequences applying algorithm (6) with d being the Euclidean distance in $\mathbb{R}$. Denote $p = |\varepsilon^{ij}|$.
- $(L_{1:p}, M_{1:p}) = \left\{\sum_{k=1}^{r} \varepsilon_k^{ij}\right\}_{r=1,\ldots,p}$.
- $\alpha_1 = M_1$ and $\alpha_{2:n_i} = \{M_s, s = 2, \ldots, \ell, \text{ s.t. } L_s \neq L_{s-1}\}$.
  $\beta_1 = L_1$ and $\beta_{2:n_j} = \{L_s, s = 2, \ldots, \ell, \text{ s.t. } M_s \neq M_{s-1}\}$.
- for $k = 1, \ldots, n_i$:

$$
\begin{aligned}
&t_k^\star = s_{\alpha_k} \quad \text{if } \alpha_{k-1} \neq \alpha_k \neq \alpha_{k+1} \text{ (only one of these if } k = 1 \text{ or } n_i)\\
&t_{k+h}^\star = s_{\alpha_k} - l + \frac{2l}{H-1}h, \quad h = 0, \ldots, H-1, \text{ with } l = \frac{\delta(t_{k+H-1} - t_k)}{2}\\
&\qquad \text{if } \alpha_{k-1} \neq \alpha_k = \alpha_{k+1} = \cdots = \alpha_{k+H-1} \neq \alpha_{k+H}.
\end{aligned} \tag{7}
$$

- for $k = 1, \ldots, n_j$:

$$
\begin{aligned}
&s_k^\star = t_{\beta_k} \quad \text{if } \beta_{k-1} \neq \beta_k \neq \beta_{k+1} \text{ (only one of these if } k = 1 \text{ or } n_j)\\
&s_{k+h}^\star = t_{\beta_k} - l + \frac{2l}{H-1}h, \quad h = 0, \ldots, H-1, \text{ with } l = \frac{\delta(s_{k+H-1} - s_k)}{2}\\
&\qquad \text{if } \beta_{k-1} \neq \beta_k = \beta_{k+1} = \cdots = \beta_{k+H-1} \neq \beta_{k+H}.
\end{aligned} \tag{8}
$$

- define

$$
\hat{g}_{ji}(t_k) = t_k^\star, \quad k = 1, \ldots, n_i \qquad \hat{g}_{ij}(s_k) = s_k^\star, \quad k = 1, \ldots, n_j. \tag{9}
$$

Note that expressions (7) and (8) correspond to an interpolation step in the cases in which more than one value of y_i has been assigned to the same single value of y_j and vice versa. In those cases, we spread those values over an interval, rather than bringing them together to the same time point. That is, if values $y_{ik}, \ldots, y_{i,k+H-1}$ are all mapped to y_{j,α_k}, instead of defining $\hat{g}_{ji}(t_{k+h}) = s_{\alpha_k}$, $h = 0, \ldots, H-1$, we equally distribute these points around w_{α_k} in such a way that the slope of $\hat{g}_{ji}$ between t_k and t_{k+H-1} is equal to δ, where δ is a parameter that needs to be chosen in advance. In this way, we guarantee that the estimates $\hat{g}_{ji}(t)$ and $\hat{g}_{ij}(t)$ are strictly increasing. This is illustrated in Fig. 2.

Once we have discrete time estimates of $g_{ji}(t)$ and $g_{ij}(t)$, for any pair $i, j = 1, \ldots, n$, following Rong and Müller (2008) we define:

$$
\hat{h}_i^{-1}(t_k) = \frac{1}{n}\sum_{j=1}^{n} \hat{g}_{ji}(t_k), \quad k = 1, \ldots, n_i, \ i = 1, \ldots, n. \tag{10}
$$

We note here that each estimated warping function, $\hat{h}_i^{-1}$, is only defined over the grid of observed event times of its corresponding curve, y_i. The same applies for the estimated registered curves $\hat{x}_i = y_i \circ \hat{h}_i$. Then, any further step involving

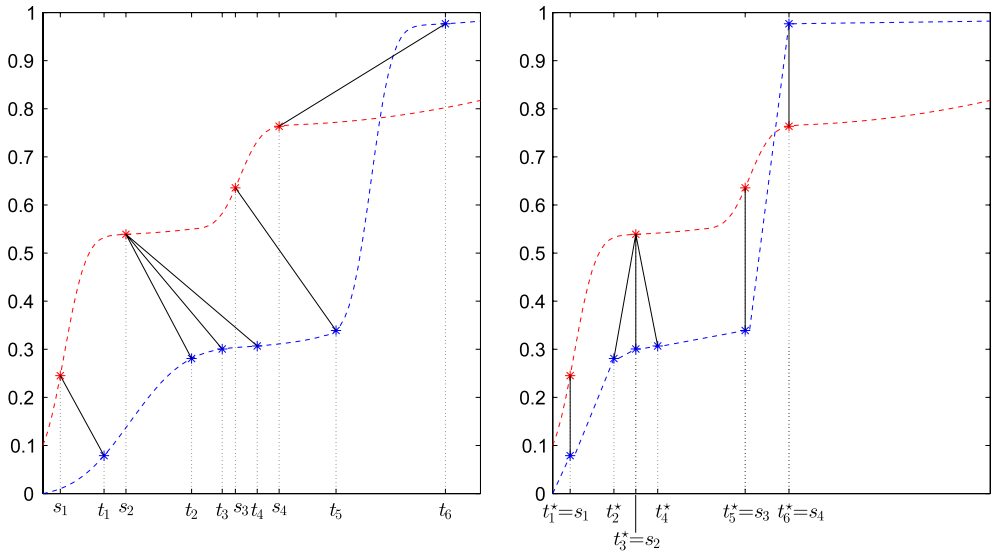

Fig. 2. Left: blue stars are points (t_k, y_{ik}) and red stars are points (s_k, y_{jk}). Solid black lines represent the alignment obtained by DTW. Imaginary continuous trajectories are displayed in dashed lines, although these are not observed in practice. Right: alignment of y_i towards y_j. Blue stars are now points $(t_k^\star = \hat{g}_{ji}(t_k), y_{ik})$. Points (t_2, y_{i2}), (t_3, y_{i3}), (t_4, y_{i4}) are all mapped to (s_2, y_{j2}) and therefore are spread in a small interval around s_2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

computations with either the estimated warping functions or registered curves, such us calculating the sample mean of the registered curves for instance, requires an additional interpolation step aiming to anchor the curves at a common grid of time points.

As for the computational complexity of the pairwise dynamic time warping algorithm to produce an alignment of a set of n curves, this is proportional to $\binom{n}{2}\ell^2 \approx (n \cdot \ell)^2$, ($\ell$ standing for some average number of events). This represents a very important reduction compared to the dynamic time warping approach for multiple alignment, which has a computational cost of $\mathcal{O}(\ell^n)$. The cost for pairwise warping can be further reduced through the subsampling of pairs for which the pairwise warping is computed (Rong and Müller, 2008).

2.4. Fine-tuning the algorithm

The pairwise dynamic time warping (PDTW) algorithm described in the previous section may be used with any choice of metric d in the dynamic time warping step and any choice of regularization slope δ in the pairwise registration step. In this section, we provide some intuition on how these parameters can be chosen in practice.

The tuning parameter δ defines the slope of the pairwise synchronization function $\hat{g}_{ji}$ for interpolating between any two points of y_i that have been aligned by dynamic time warping to the same point y_j. A small positive slope is needed for the interpolation to ensure that the pairwise synchronization functions and the warping functions are invertible and well defined. Larger values of δ will lead to increased separation between these time points in the alignment of y_i to y_j.

We found that in practice different values of δ yield very similar global alignments so that the precise choice of δ is less relevant. The separation in the values of the pairwise warping function $\hat{g}_{ji}$ induced by $\delta > 0$ is always bounded by the distance between the time points of y_j. In Fig. 2 for instance, the separation between $t_2^\star$, $t_3^\star$ and $t_4^\star$ is constrained by the distance between s_1 and s_2 and s_2 and s_3, so higher values of δ will increase the separation of these points only up to a limit, which becomes smaller as the number of events gets larger. In practice it is best to keep δ at a small positive value, such as $\delta = 10^{-5}$, the value which we use throughout this paper.

In cases where the observed data do not represent events or empirical cumulative event functions, metrics other than the Euclidean distance may be of interest. For instance, one could be interested in aligning segments of the observed curves that have similar slopes. Besides considering the alignment of the differentiated curves, one could directly address the problem on the original curves by considering in (6) a metric based on slopes differences such as:

$$d(a_i, b_j) = \left(\frac{a_i - a_{i-1}}{m_a(t_i - t_{i-1})} - \frac{b_j - b_{j-1}}{m_b(s_j - s_{j-1})} \right)^2, \quad m_a = \max_i \frac{a_i - a_{i-1}}{t_i - t_{i-1}}, m_b = \max_j \frac{b_j - b_{j-1}}{s_j - s_{j-1}}.$$

Of course, different metrics will lead to different final alignments, and one should choose the metric to use with dynamic time warping according to the nature of the data and to the expected type of alignment.

3. Simulation study

In this section, we conduct a simulation study to evaluate the performance of our method at recovering registered curves from a sample of observed time-warped curves. We generate curves according to two different models. Observations are

Table 1
Mean values and standard deviations of RMSE over 100 simulation runs for different sample sizes.

	n	PDTW euclidean distance	PDTW slopes distance	Quantile normalization
Model 1	15	0.0085 (0.0094)	0.0489 (0.0158)	0.0127 (0.0143)
	25	0.0066 (0.0049)	0.0492 (0.0121)	0.0081 (0.0072)
	50	0.0051 (0.0038)	0.0493 (0.0092)	0.0041 (0.0054)
	100	0.0038 (0.0022)	0.0477 (0.0072)	0.0024 (0.0029)
Model 2	15	0.1320 (0.0231)	0.0178 (0.0109)	0.1249 (0.0270)
	25	0.1315 (0.0195)	0.0160 (0.0079)	0.1253 (0.0266)
	50	0.1330 (0.0119)	0.0130 (0.0037)	0.1282 (0.0181)
	100	0.1339 (0.0078)	0.0118 (0.0019)	0.1297 (0.0111)

(t_{ij}, Y_{ik}) where $Y_{ik} = \mu_i(h_i^{-1}(t_{ik}))$, $t_{ik} \in [0, T]$, $i = 1, \ldots, n$, $k = 1, \ldots, n_i$, $t_{i1} = 0$, $t_{in_i} = T$. The numbers of events per curve, $n_i - 2$, are independent and uniformly distributed on $\{1, 2, \ldots, N\}$, N being the maximum number of events, where we fix $T = 10$, $N = 25$ and select the number of subjects or units n from $\{15, 25, 50, 100\}$.

The warping functions $h_i(t)$ are equal to t^{β_i} with probability $1/2$ and to $2t - t^{\beta_i}$ with probability $1/2$, with $\beta_i \sim U(1, 2)$ i.i.d. The intensity functions for the two models are

Model 1: $\mu_i(t) = \min\{f(t, N), n_i\}$

Model 2: $\mu_i(t) = a_i f(t, n_i)$ if $t \leq T - 1$ and $\mu_i(t) = a_i n_i$ if $t > T - 1$, with $a_i \sim U(1.5, 2.5)$ i.i.d., where

$$f(t, m) = \frac{m}{2.25}\left\{1.25 + \arctan\left(\frac{t}{T-1}(\tan(1) - \tan(-1.25)) + \tan(-1.25)\right)\right\}.$$

In Fig. 3 we present some of the curves generated with the two models. Note that, in model 1 all the intensity functions coincide up to the point at which each curve reaches its event number. Also, observed values under that model correspond to the number of events. In model 2, the intensity functions are different for different curves and the observed values do not correspond to the number of events but to a transformation of them. To assess the performance of the methods for aligning curves generated under the two models, we use two different metrics, the Euclidean distance and the distance based on slopes (see Section 2.4). To enhance the comparisons, the curves are also registered using the quantile normalization method (Bolstad et al., 2003; Gallon et al., 2013) that is described later in this section.

For each sample size we have generated 100 samples and for each sample element we have computed the relative mean square error between the estimated and the original intensity functions,

$$\mathrm{RMSE}(k) = \frac{1}{n}\sum_{i=1}^{n}\frac{\int_0^T (\mu_i(t) - \hat{\mu}_i^k(t))^2 dt}{\int_0^T \mu_i(t)^2 dt}.$$

In Table 1 we present the mean values and standard deviations of the values of RMSE. These results are illustrated in Fig. 3 for four particular simulation runs.

As expected, for model 2 the PWDT with distances based on the slope metric works better than any of the other methods since it is able to recover the original structure of the intensity curves. Both the PDTW with Euclidean distance and the quantile normalization method tend to estimate intensity functions that are closer to each other than the *true* ones (see Fig. 3, third column). For model 1 we observe the opposite behavior. Here the PDTW with distance based on slopes aligns the points at which similar slope changes occur and fails at recovering similar values in different curves. In this model the PDTW with Euclidean distance and the quantile normalization method achieve good results. In terms of RMSE quantile normalization is better than PDTW for larger sample sizes. Indeed, for small sample sizes (relative to the number of events) the quantile normalization method may not work at all producing intensity functions that are not even well defined (due to non-increasing warping function estimates). This effect tends to disappear as the number of curves increases, however, we still observe it even with $n = 50$ or $n = 100$ in some simulation runs.

For aligning event data, for which the observed curves represent absolute number of events, the *quantile normalization method* (Bolstad et al., 2003; Gallon et al., 2013; Zhang and Müller, 2011) that is included in the above simulation comparisons proceeds by defining the warping functions as

$$\hat{h}_i(\bar{t}_k) = t_{ik}, \quad i = 1, \ldots, n, \ k = 1, \ldots, n_i, \tag{11}$$

where

$$\bar{t}_k = \frac{1}{|I_k|}\sum_{i \in I_k}^{t_{ik}}, \quad I_k = \{i, n_i \geq k\},$$

is the average k-th event time for those curves with at least k events, $k = 1, \ldots, N$ ($N = \max\{n_i\}$). The estimators resulting from this method for event data registration are not well defined in general. Indeed, nothing ensures that $\bar{t}_k < \bar{t}_{k+1}$, and the warping function estimate given in (11) is not guaranteed to be monotone.

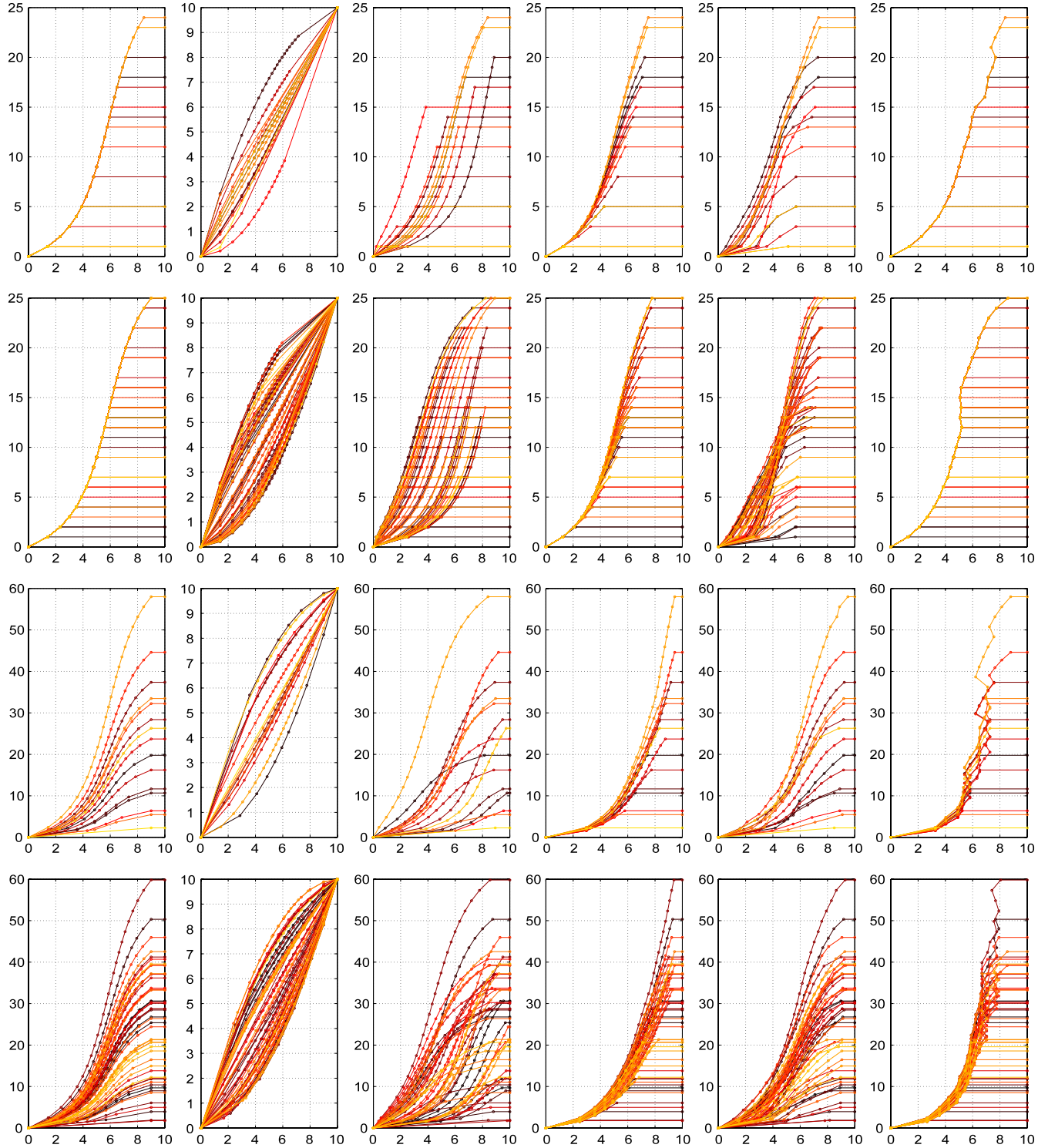

Fig. 3. First three columns, from left to right: individual intensity functions; individual warping functions; observed time-warped curves. Last three columns, from left to right: estimated intensity functions by PDTW with Euclidean distance, by PDTW with distance based on slopes and by quantile normalization. From top to bottom: simulated curves under model 1 with $n = 15$; simulated curves under model 1 with $n = 50$; simulated curves under model 2 with $n = 15$; simulated curves under model 2 with $n = 50$.

4. Applications

4.1. Historical fertility data

We study fertility from a well-documented 17/18th century cohort of 1877 native born French-Canadian women who lived past age 50. This data contain the number of children and the age at the different births for each woman in the sample.

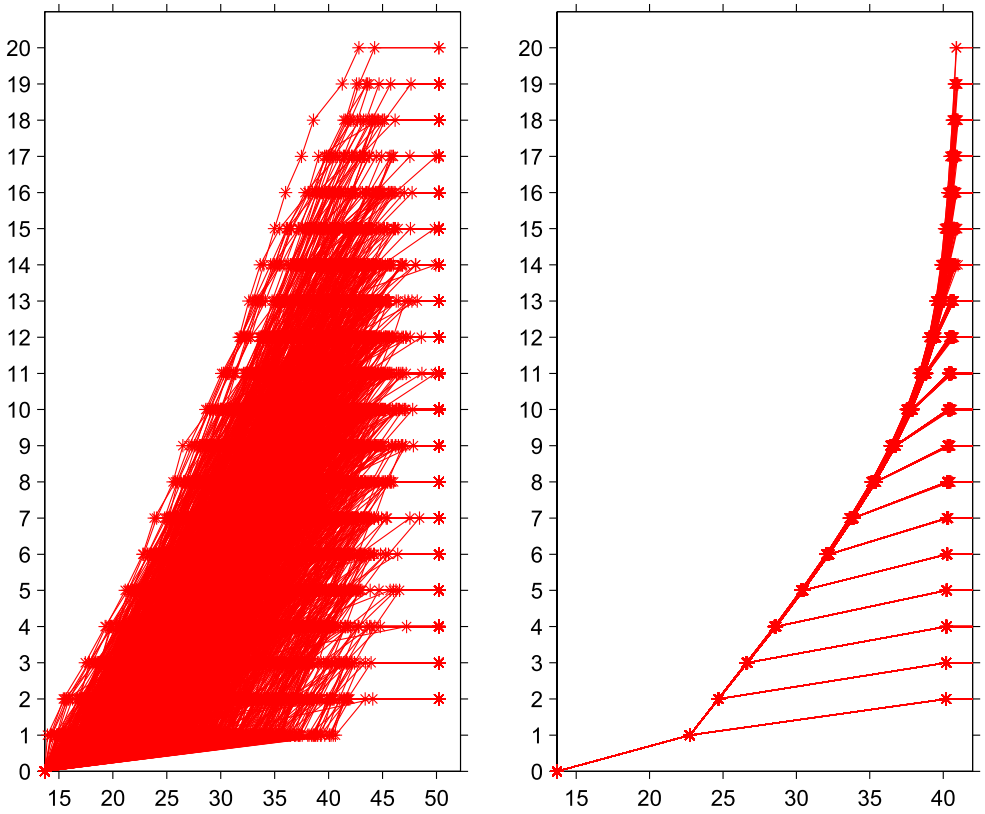

Fig. 4. Left: absolute number of children along time for 1810 women. Stars represent the observed data, solid lines are obtained by connecting births within each record. Right: absolute number of children versus registered births times for 1810 women.

We consider only those women with more than one child; the resulting data consist of 1810 individuals. For more details on this data set see Müller et al. (2002) and the references therein.

To study the relation between the total number of children and the dynamics of the giving birth process, we transform the data to obtain a curve (t_{ik}, y_{ik}), $i = 1, \ldots, n_i$ for each woman, where t_{ik} is the age of the i-th woman at her k-th birth, $y_{ik} = k$ and n_i is her total number of children. The time domain is $\mathcal{T} = [13.7, 50.2)$, units are years. To anchor all sample curves on this interval, we add points $(13.7, 0)$ and $(50.2, n_i)$ at the beginning and at the end of each curve. Since this introduces an artificial constant fragment at the end of each curve, and an important reference is the age at the last birth for each woman and not the end of the interval, we force points (t_{in_i}, y_{in_i}), $i = 1, \ldots, n$, to be mapped together.

The entire sample before and after registration by PDTW with Euclidean distance is presented in Fig. 4. Note that the registered sample includes contracted time domains for some women so that the periods between births may appear to be shorter than 9 months. This is a consequence of the forced alignment at the first and last births across all women and is also reflected in some of the mean group intensity functions (averages of the registered curves for women with the same number of offspring) displayed in Fig. 5. Nevertheless, the mean intensity function for the whole sample presents a good approximation of the standard fertility process and captures a linear trend along with the mean ages at the first and last birth, which are 22.4 and 40.2 years respectively. In the case of the sample mean of the original trajectories (Fig. 5), these ages are under-, respectively, over-estimated.

After registration, the estimated warping functions can be used for classification purposes.

For classification, we consider a distance-based approach by defining individual distances derived from the warping functions as follows,

$$d(i,j) = \int_{\mathcal{T}} \left(\hat{h}_i^{-1}(t) - \hat{h}_j^{-1}(t)\right)^2 dt, \quad i,j = 1, \ldots, n. \tag{12}$$

Indeed, since the warping functions are strictly increasing and map $\mathcal{T}$ to $\mathcal{T}$ bijectively, the area between two of these functions is a good indicator of the difference in the degree of time distortion that they introduce in the respective individual curves. We expect that this distance will be useful in discriminating between different fertility dynamics. This distance matrix needs to be computed numerically from the discrete-valued warping function estimates.

Unlike common approaches in functional data analysis, we propose to perform clustering directly from the distance matrix without any previous step to reduce the dimensionality of the data. The clustering method is a modified version of the K-means algorithm in which the centroid of a group G, is calculated as its Fréchet mean, $c_G = \arg\min_{x \in G} \sum_{y \in G} d^2(x, y)$. To choose the number of clusters we use the *silhouette* coefficient (Kaufman and Rousseeuw, 1990) that for each individual provides a measure of how well classified it is. The *silhouette* of individual i is $(b_i - a_i)/\max\{a_i, b_i\}$, where a_i is the mean

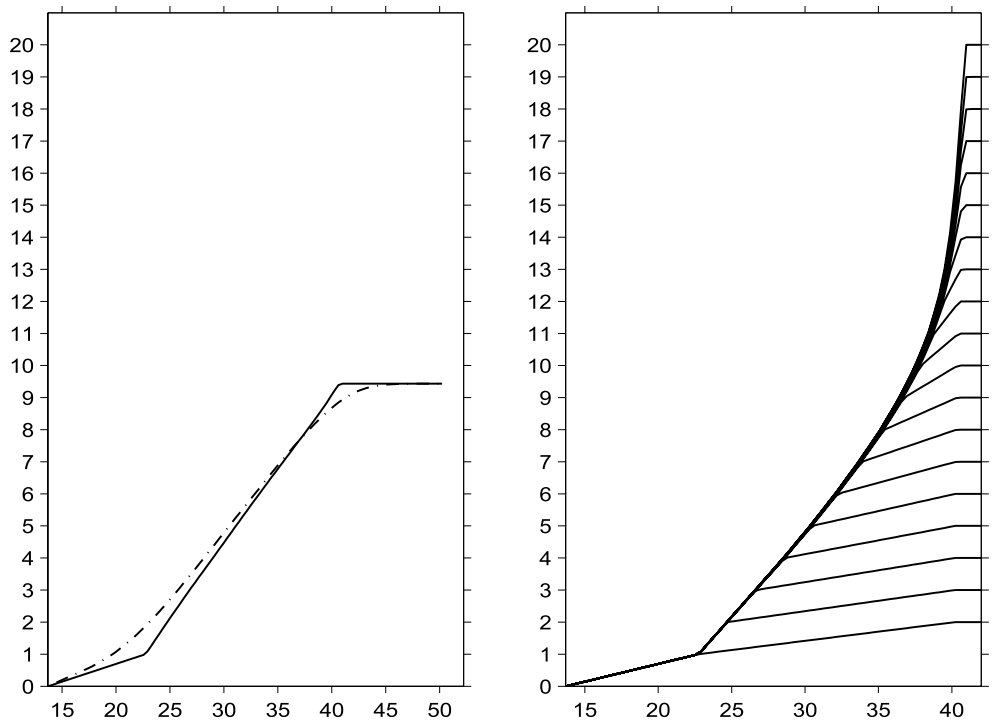

Fig. 5. Left: mean intensity function before (dashed–dotted line) and after registration (solid line). Right: group mean intensity functions. The groups are defined by the women with the same number of children.

distance of i to the rest of individuals in the same cluster and b_i is the mean distance of i to the individuals of the closest cluster (besides the one it is assigned to). Finally, the silhouette coefficient of a clustering is the mean silhouette across individuals. Then, the choice of the number of clusters is done by maximizing the silhouette coefficient. For this data set, the maximum was found for $k = 2$ clusters, with a value of 0.60. According to Kaufman and Rousseeuw (1990), a value of this coefficient between 0.51 and 0.70 indicates that a reasonable structure has been found in the data.

The results for $k = 2$ clusters are displayed in Fig. 6. The two clusters can be defined as women with late birth trajectories, and women with regular birth trajectories. A regular birth trajectory is almost linear, the first birth occurring in average at 20.8 years and the last birth at 39.7 years. In contrast, a late birth trajectory is piecewise linear with two differentiated pieces, the first one with lower slope than the second one. In fact the slope of the second fragment is similar to that of a regular trajectory. The average ages at the first and last birth for women in this category are 29.4 and 41.8 years. Then, we can say that the main difference between the women in the two clusters is the age at which the first child is born, since after the first birth, the fertility process looks similar in both cases. If we now look at the distribution of the number of children in these two clusters we observe, as expected, that the total number of children is generally lower for those women who had their first child at an older age.

4.2. Online auction data

Modeling of price paths in on-line auction data has received a lot of attention in recent years (Shmueli and Jank, 2005; Jank and Shmueli, 2006; Shmueli et al., 2007; Liu and Müller, 2008). One of the reasons is the availability of huge amounts of data that can be obtained from the on-line auction and shopping website eBay.com, which has become a global market place in which millions of people worldwide buy and sell products. The price evolution during an auction can be thought as a continuous process which is observed discretely and sparsely only at the instants in which bids are placed. In fact, bids tend to concentrate at the beginning and at the end of the auction, corresponding to two typically observed phenomena, "early bidding" and "bid sniping" (a situation in which "snipers" place their bids at the very last moment). Here, we analyze a set of 158 eBay auctions for Palm M515 Personal Digital Assistants (PDA), of a fixed duration of seven days, that took place between March and May, 2003. This data set is publicly available at http://www.rhsmith.umd.edu/digits/statistics/data.aspx and has been previously studied by Shmueli and Jank (2005) and Liu and Müller (2008), among others. As in Liu and Müller (2008) we restrict our analysis to 156 auctions after removing two irregular recordings.

Our interest here is not focused on the price process, but rather on the intensity functions that quantify the bid arrivals along the time axis, where each bid arrival generates an event time. This approach may be useful to understand the different bidding behaviors and for characterizing the bid arrivals process. A characteristic of these data is that the bidding trajectories for different auctions are quite dissimilar. While most of them reflect very low bidding activity at the beginning of the auction and intense bidding when the auction is near its end, in other auctions one observes different patterns. It seems reasonable

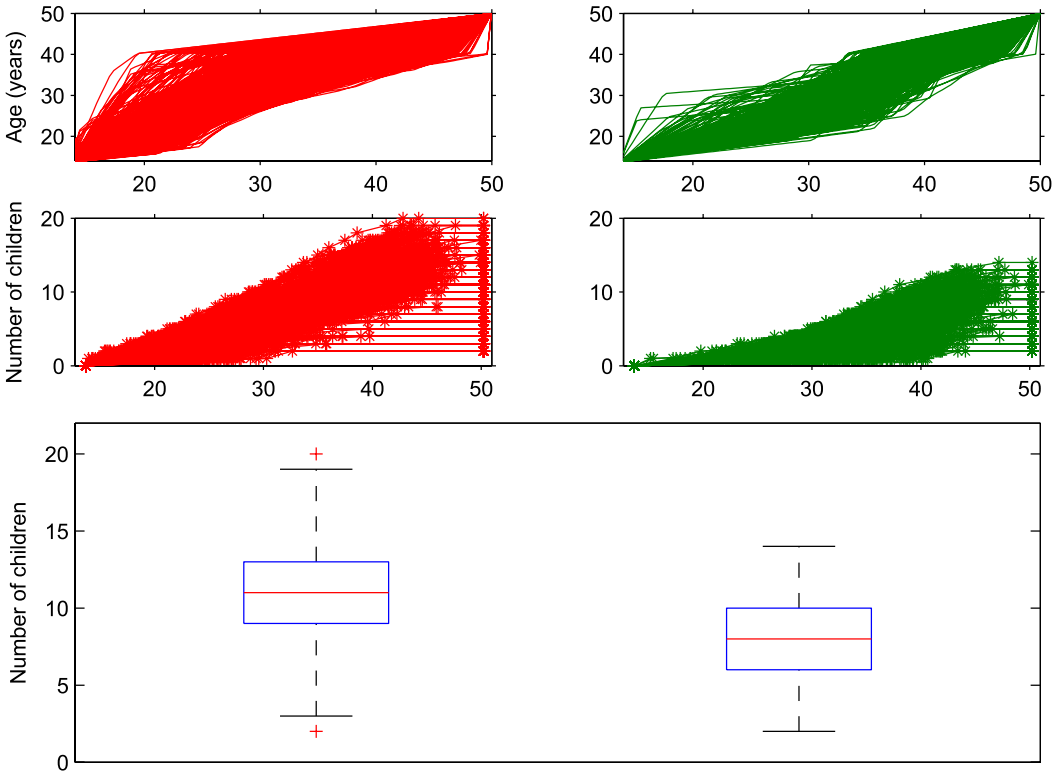

Fig. 6. Warping function estimates ($\hat{h}_i^{-1}$), fertility curves and box plots of the number of children for the two clusters.

to view these trajectories as accelerated and decelerated versions of an underlying bidding intensity function and therefore the proposed time warping model seems suitable for this analysis.

The time domain for observations is $\mathcal{T} = [0, 168)$, where the time units are hours. As in the previous example, we have added points $(0, 0)$ and $(168, n_i)$ at the beginning and at the end of each trajectory. The minimum and maximum observed numbers of bids for auctions in the sample are 8 and 51, respectively. Applying PDTW with Euclidean distance, with results shown in Fig. 7, indicates that the registered (non-standardized) intensity functions show an exponential-type pattern. We obtain an "almost common" intensity function irrespective of the number of bids. Indeed, the function is the same for all auctions except for the time period after the last bid, which is represented by a constant straight line at a height that depends on the total number of bids. This time period has been expanded (resp. contracted) during the registration process in those auctions with a low (resp. high) number of bids. Due to the low variability in the registered sample, the group mean intensity functions (Fig. 8) are quite similar to the registered trajectories. Also in Fig. 8, the overall mean intensity function after registration is compared to the sample mean of the observed curves. "Early bidding" and "bid sniping" can thus be described as phenomena that are due to time warping of the overall mean intensity function.

Aiming to differentiate between the different bidding behaviors, we apply the warping-based clustering procedure described in Section 4.1. Here, again the number of clusters was found to be $k = 2$, with a *silhouette* coefficient of 0.65. In Fig. 9 we present the warping function estimates, bidding trajectories and final price distributions of the two clusters. An interesting question is whether there is any possible relation between the bidding dynamics and the closing price of an auction. The clusters are defined by auctions with late ("L") and regular–early ("R–E") bidding activity, according to the shape of the warping function estimates corresponding to each group.

In contrast to the information discernible in the warping functions, it is quite difficult to ascertain differences between the observed bidding trajectories of the two clusters. The first cluster, "L", is composed by auctions with a very slow start, and contains most of the auctions with lowest final prices. Indeed the median final price of cluster "L" is slightly lower than that of "R–E". However it also contains auctions with high closing prices (ranging between 240 and 260$), although these systematically correspond to auctions with unusually high opening bids (ranging between 50 and 179.99$). The second cluster, "R–E", contains both auctions with bidding trajectories that are similar to the mean intensity function obtained after registration and auctions presenting very intense early bidding activity.

The bidding trajectories of both clusters seem to present a similar final shape that corresponds to a substantial increase in the bidding activity during the last day of the auction and is reflected by the terminal vertical shift. That is, bidder's behaviors during the last day of an auction seems to be more or less the same, independently of the previous bidding activity. It might thus appear that the early bidding activity determines the outcome of an auction. However, the situation is more complex. Within the cluster "R–E" one can distinguish two very different starting behaviors, which do not correspond to different price distributions. Indeed, if one chooses to divide the data into 3 groups, cluster "R–E" is split into an "R" and an "E" cluster for which the closing price distributions are very similar. As a conclusion, a very low activity during the first 3/4 of an auction

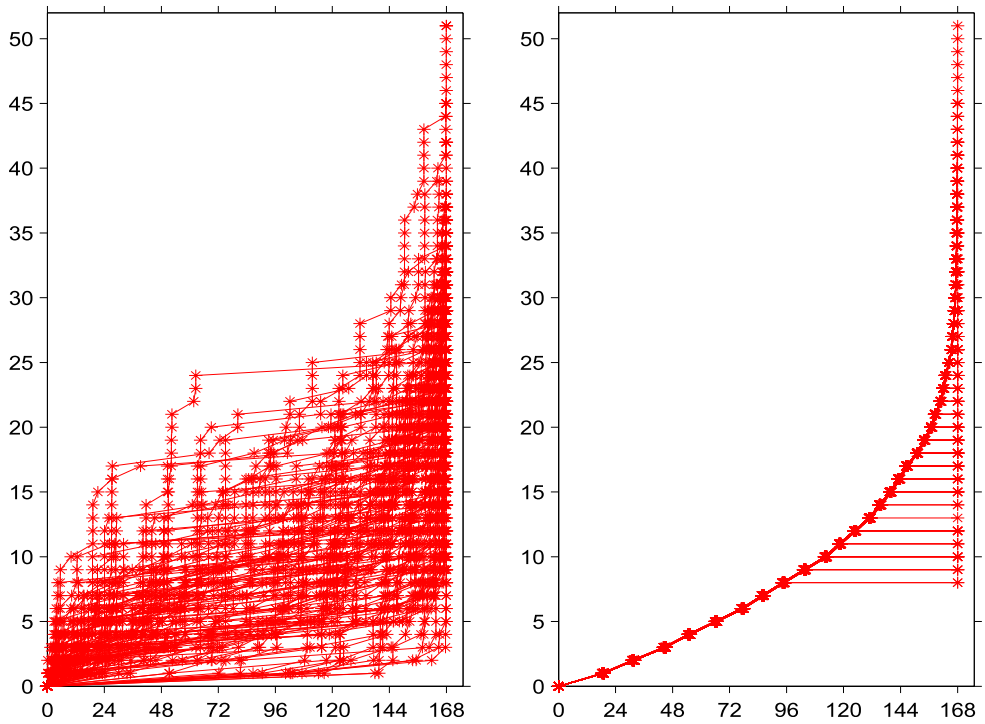

Fig. 7. Left: observed bidding trajectories Y_i for 156 auctions. Stars represent the observed data, solid lines are connecting the stars. Right: registered bidding trajectories.

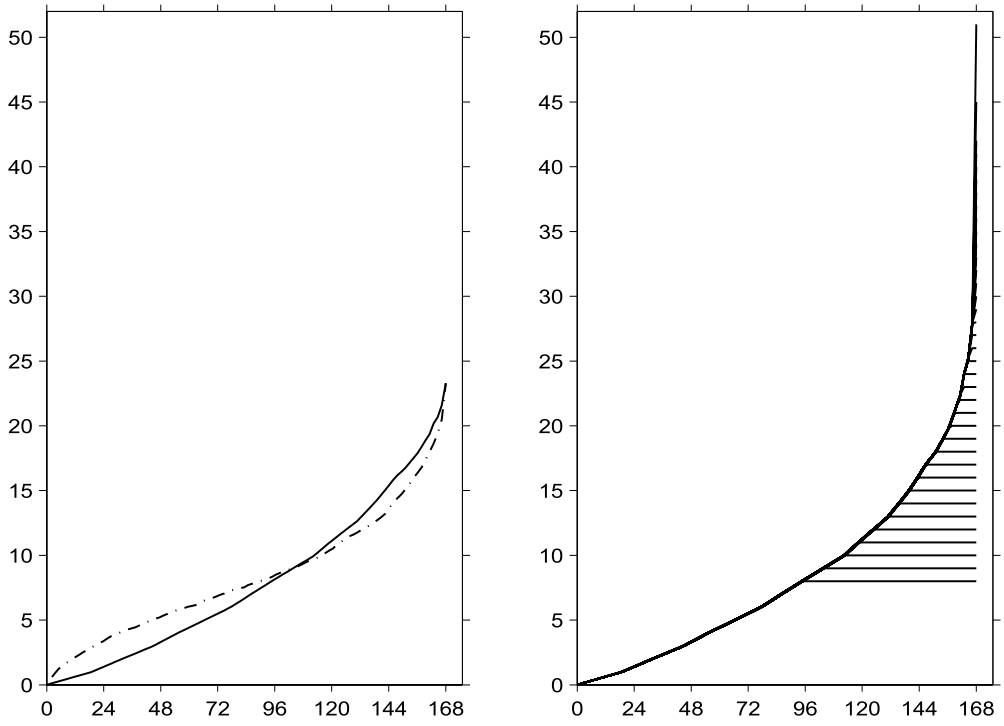

Fig. 8. Left: mean intensity function before (dashed–dotted line) and after registration (solid line). Right: group mean intensity functions. The grouping is defined by the number of bids.

seems to portend lower closing prices. In contrast, auctions with an early intense bidding activity do not necessarily lead to higher final prices.

5. Discussion

We have proposed an efficient registration algorithm for functional data that can be characterized by a random number of event times, which could correspond to original observations, or alternatively could be derived features from functional

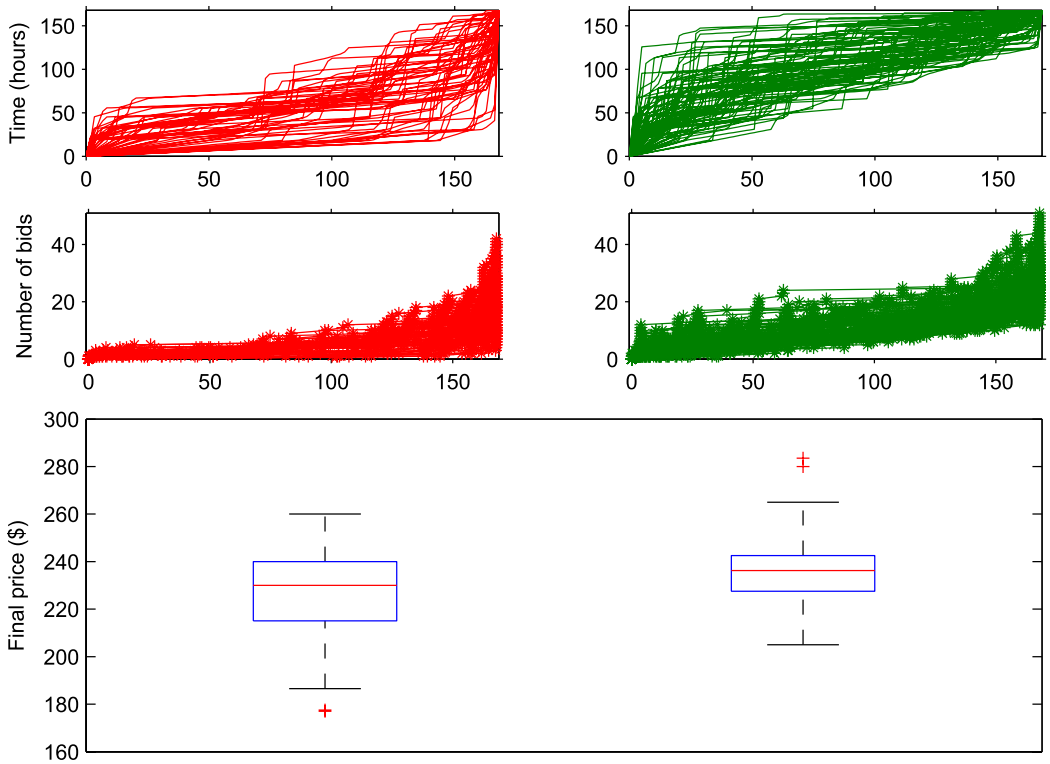

Fig. 9. Warping function estimates ($\hat{h}_i^{-1}$), bidding trajectories and box plots of the final auction price for the two clusters "L" ("Late Bidding", left) and "R–E" ("Regular–Early Bidding", right).

data such as the location of peaks or other landmarks that are not consistently observed across the random trajectories that constitute the functional data sample. One of the advantages of the proposed method is that it can cope with large sets of curves and small as well as varying numbers of observed events per subject.

Unlike the previous DTW-based strategies for continuous time curve registration we focus on discrete time pairwise "local" alignments that subsequently are combined to obtain a whole sample "global" alignment. The proposed methodology can thus be viewed as a computationally feasible extension of DTW for the simultaneous alignment of many curves that are characterized by events that vary from subject to subject.

In addition to the estimation of the warping functions, it is also often of interest to determine their uncertainty. This topic has been largely ignored in the literature on warping and registration of random functions. There are different sources of uncertainty in our proposed approach. The first source for uncertainty is the pairwise registration step that is implemented by DTW. A second source is the determination of the overall warping functions obtained from combining the information on pairwise registration. Two practical approaches can be taken to analyze the aspects of these uncertainties: one is to bootstrap individual event data, by resampling the event times from the empirical distribution that is defined by the observed events, separately for each subject. A bootstrap sample then consists of event data from the same sample of subjects that has been observed but with resampled event times for each of the subjects. The bootstrap distribution of the registration functions can then be assessed. This addresses primarily the uncertainties introduced by pairwise registration. Secondly, random subsamples of the subjects can be repeatedly generated and the registration functions can be determined for each of the subjects in the subsample. Using this device repeatedly will provide information about the uncertainty in obtaining the overall registration functions from the pairwise alignments, although for smaller sized samples the assessed uncertainty may be larger than the one in the actual sample. The two devices can also be combined to arrive at a better assessment of overall uncertainty. A more precise analysis of such practical methods to quantify uncertainty will be of interest but is beyond the scope of this paper.

The method relies on the choice of a metric on the observations space, which is used to find matching values over different curves. The standard choice in many applications is the Euclidean distance, although many other metric could be used, depending on the specifics of the warping problem. Indeed, this is a crucial point: different metrics will provide different alignments. Knowing which values should be aligned together is equivalent to having a notion of what similarity means in each particular application. For example, in the applications of Section 4, we focus on aligning some particular events, such as the birth of the first child, for the fertility data. For this purpose, the Euclidean distance proves to be suitable.

The alternative quantile normalization method that is described in Section 3 leads to anomalies when directly applied to event data with unequal numbers of events, as it has not been devised for such data. Therefore, this method is not particularly suited for the analysis of the type of data that we are targeting in this paper. Nevertheless, when applying it to data and simulations, this method can give reasonable results, especially for cases where the number of subjects is large.

As for the clustering of a sample of curves, different notions of similarity can be suitable for different purposes. In this context, our proposal of clustering based on warping estimates can be seen as a way of defining a distance between individuals which focuses on the similarity of their temporal behavior.

Acknowledgments

This research was supported by NSF grants DMS-1104426 and DMS-1228369, and by Spanish grants "José Castillejo" JC2010-0057, MTM2010-17323 and ECO2011-25706 (Spanish Ministry of Science and Innovation).

References

Aach, J., Church, G.M., 2001. Aligning gene expression time series with time warping algorithms. Bioinformatics 17, 495–508.
Bigot, J., 2006. Landmark-based registration of curves via the continuous wavelet transform. Journal of Computational and Graphical Statistics 15, 542–564.
Bolstad, B.M., Irizarry, R.A., Astrand, M., Speed, T.P., 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19, 185–193.
Bouzas, P.R., Valderrama, M., Aguilera, A.M., Ruiz-Fuentes, , 2006. Modeling the mean of a doubly stochastic Poisson process by functional data analysis. Computational Statistics and Data Analysis 50, 2655–2667.
Clote, P., Straubhaar, J., 2006. Symmetric time warping, Boltzmann pair probabilities and functional genomics. Journal of Mathematical Biology 53, 135–161.
Gallon, S., Loubes, J.M., Maza, E., 2013. Statistical properties of the quantile normalization method for density curve alignment. Mathematical Biosciences 242, 129–142.
Gasser, T., Kneip, A., 1995. Searching for structure in curve samples. Journal of the American Statistical Association 90, 1179–1188.
Gasser, T., Köhler, W., Müller, H.G., Kneip, A., Largo, R., Molinari, L., Prader, A., 1984a. Velocity and acceleration of height growth using kernel estimation. Annals of Human Biology 11, 397–411.
Gasser, T., Müller, H.G., Köhler, W., Molinari, L., Prader, A., 1984b. Nonparametric regression analysis of growth curves. The Annals of Statistics 12, 210–229.
Gervini, D., Gasser, T., 2004. Self-modeling warping functions. Journal of the Royal Statistical Society, Series B 66, 959–971.
Jank, W., Shmueli, G., 2006. Functional data analysis in electronic commerce research. Statistical Science 21, 155–166.
Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley-Interscience.
Kneip, A., Gasser, T., 1992. Statistical tools to analyze data representing a sample of curves. The Annals of Statistics 16, 82–112.
Kneip, A., Ramsay, J., 2008. Combining registration and fitting for functional models. Journal of the American Statistical Association 103, 1115–1165.
Kruskal, J., Liberman, M., 1983. The symmetric time-warping problem: from continuous to discrete. In: Kruskal, J.B., Sankoff, D. (Eds.), Time Warps, String Edits, and Mocromolecules: The Theory and Practice of Sequence Comparison. Adison-Wesley, pp. 125–161.
Liu, X., Müller, H.G., 2004. Functional convex averaging and synchronization for time-warped random curves. Journal of the American Statistical Association 99, 687–699.
Liu, B., Müller, H.G., 2008. Functional data analysis for sparse auction data. In: Jank, W., Shmueli, G. (Eds.), Statistical Methods in E-Commerce Research. Wiley, New York, pp. 269–290.
Müller, H.G., Chiou, J.M., Carey, J.R., Wang, J.L., 2002. Fertility and lifespan: late children enhance female longevity. The Journals of Gerontology: Biological Sciences 57, 202–206.
Needleman, S.B., Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequences of two proteins. Journal of Molecular Biology 48, 443–453.
Rong, T., Müller, H.G., 2008. Pairwise curve synchronization for functional data. Biometrika 95, 875–889.
Rønn, B.B., 2001. Nonparametric maximum likelihood estimation for shifted curves. Journal of the Royal Statistical Society, Series B 63, 243–259.
Shmueli, G., Jank, W., 2005. Visualizing online auctions. Journal of Computational and Graphical Statistics 14, 299–319.
Shmueli, G., Russo, R.P., Jank, W., 2007. The BARISTA: a model for bid arrivals in online auctions. The Annals of Applied Statistics 1, 412–441.
Silverman, B., 1995. Incorporating parametric effects into functional principal components analysis. Journal of the Royal Statistical Society, Series B 57, 673–689.
Telesca, D., Inoue, L.Y.T., 2008. Bayesian hierarchical curve registration. Journal of the American Statistical Association 103, 328–339.
Wang, K., Gasser, T., 1997. Alignment of curves by dynamic time warping. The Annals of Statistics 25, 1251–1276.
Wang, K., Gasser, T., 1999. Synchronizing sample curves nonparametrically. The Annals of Statistics 27, 439–460.
Wu, S., Müller, H.G., Zhang, Z., 2013. Functional data analysis for point processes with rare events. Statistica Sinica 23, 1–23.
Zhang, Z., Müller, H.G., 2011. Functional density synchronization. Computational Statistics and Data Analysis 55, 2234–2249.