ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2018)

Simultaneous inference for misaligned multivariate functional data

Niels Lundtorp Olsen, Bo Markussen and Lars Lau Raket

University of Copenhagen, Denmark

[Received August 2016. Final revision February 2018]

Summary. We consider inference for misaligned multivariate functional data that represents the same underlying curve, but where the functional samples have systematic differences in shape. We introduce a class of generally applicable models where warping effects are modelled through non-linear transformation of latent Gaussian variables and systematic shape differences are modelled by Gaussian processes. To model cross-covariance between sample co-ordinates we propose a class of low dimensional cross-covariance structures that are suitable for modelling multivariate functional data. We present a method for doing maximum likelihood estimation in the models and apply the method to three data sets. The first data set is from a motion tracking system where the spatial positions of a large number of body markers are tracked in three dimensions over time. The second data set consists of longitudinal height and weight measurements for Danish boys. The third data set consists of three-dimensional spatial hand paths from a controlled obstacle avoidance experiment. We use the method to estimate the cross-covariance structure and use a classification set-up to demonstrate that the method outperforms state of the art methods for handling misaligned curve data.

Keywords: Curve alignment; Functional data analysis; Non-linear mixed effects models; Template estimation

1. Introduction

Whereas the literature and available methods for statistical analysis of univariate functional data have been rapidly increasing during the last two decades, multivariate functional data have been a largely overlooked topic. Extension of univariate methodology to multivariate functional data is often considered a trivial task but is rarely done in practice. As a result, the non-trivial parts of extending methodology, such as temporal modelling of cross-covariance or warping of misaligned multi-dimensional signals, have received only little attention.

A wide range of methods for aligning curves is available. For general reviews of the literature on curve alignment, we refer to Ramsay and Silverman (2005), Kneip and Ramsay (2008) and Wang *et al.* (2015). Curve alignment is a non-linear problem so, for the vast majority of methods, one cannot generally expect to align data in a globally optimal way. In the multitude of available methods for univariate functional data, the quality of the results that are obtained with the available implementations is very variable. Often, good implementations of simple methods outperform far more advanced methods with less polished implementations, even if the advanced methods should be more suitable to the data at hand. From the perspective of multivariate functional data, a major issue is that only very few methods with publicly available implementations support alignment of multivariate curves.

Address for correspondence: Niels Lundtorp Olsen, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, København Ø 2100, Denmark.
E-mail: niels.olsen@math.ku.dk

 0035–9254/18/68000

Although misaligned multivariate functional data have been underrepresented in the statistics literature, similar problems have had a central role in other fields. Analysis of misaligned curves in multiple dimensions is fundamental in the shape analysis literature (Younes, 1998; Sebastian *et al.*, 2003; Manay *et al.*, 2006), where for example closed planar shapes can be thought of as functions $f : [0, 1] \to \mathbb{R}^2$ with $f(0) = f(1)$. In much shape data, we do not observe the parameterization of these functions, and for closed shapes the start and end points (0 and 1) of the parameterization are arbitrary in terms of the observed data. As an example, consider data consisting of cells' outlines obtained from two-dimensional images that have been manually annotated. Here the first annotated point on a cell does not bear any significance—in fact the orientation of the cell is most likely to be completely random in the image. For this reason, a fundamental direction of theory in the shape analysis literature is built around invariance to parameterization of the function (Younes, 1998) as well as other classical shape invariances such as translation, scaling and rotation (Kendall, 1989; Dryden and Mardia, 1998).

In recent years, the idea of using invariances similarly to the shape analysis literature has been introduced as a general tool to analyse functional data (Vantini, 2012). The most notable class of methods is based on elastic distances for functional data analysis (Srivastava *et al.*, 2011; Kurtek *et al.*, 2012; Tucker *et al.*, 2013; Srivastava and Klassen, 2016). The fundamental idea underlying these methods is to represent data in terms of square-root velocity functions and to take advantage of the invariance properties of distance on the associated function space, in particular that distances are not affected by warping of the domain in the observed representation. An elastic distance between two curves f_1 and f_2 can be defined as the minimal distance between the square-root velocity functions that are associated with f_1 and $f_2 \circ v$ where the minimum is taken over all possible warps v of f_2 (in the original representation). This approach has proven very successful compared with many conventional approaches, and efficient high quality implementations for various types of data and types of analyses are available (see `http://ssamg.stat.fsu.edu/software/`, and Tucker (2017)).

The vast majority of available methods for handling misaligned functional data are heuristic in the sense that they are based on some choice of data similarity measure that is typically not chosen because it fits well with important characteristics of the data. Rather, the typical rationale is computational convenience and/or incremental improvements over other methods. In the shape literature, methods are perhaps less heuristic and more idealistic, in the sense that they are derived from principles of how a distance between shapes should ideally be. This ideal behaviour is typically specified through invariance properties such as those described above. In contrast with these approaches for handling misalignment, we propose a full simultaneous statistical model for the fundamental types of variation in misaligned multivariate curves. In particular, we propose to treat amplitude variation and warping variation equally by modelling them as random effects on their respective domains.

Only few works have previously considered the idea of simultaneously modelling amplitude and warping as random effects. An early example of an integrated statistical model that modelled curve shifts as random Gaussian effects was presented in Rønn (2001). The simultaneous inference in the model enables data-driven regularization of the magnitude of the shifts through the estimated variance parameters. The idea has been extended to more general warping functions that are modelled by polynomials (Gervini and Gasser, 2005; Rønn and Skovgaard, 2009), and lately also to include serially correlated noise within the observations of an individual curve (Raket *et al.*, 2014). In addition to the data-driven regularization of the predicted random effects that is achieved through estimation of variance parameters, the use of likelihood-based inference naturally relates the discrete observation points and the underlying continuous model. This relationship avoids many common issues that arise when developing methods for continuous

data in the form of presmoothed curves. In particular, the pinching problem, where areas with large deviations are compressed by warping to minimize the integrated residual, does not exist for these methods. Furthermore, the simultaneous modelling of amplitude and warping effects introduces an explicit maximum likelihood criterion for resolving the identifiability problems that are related to separating warp and amplitude effects (Marron *et al.*, 2015). The maximum likelihood estimates induce a separation of the two effects, namely the most likely given the variation that is observed in the data.

A related class of models with random affine transformations of both warping and amplitude variation has become popular in growth curve analysis (Beath, 2007; Cole *et al.*, 2010). Hadjipantelis *et al.* (2014, 2015) provide an extension to this in terms of a simultaneous mixed effects model for the scores in separate functional principal component analyses of the amplitude and the warping effects. The simultaneous model allows not only for cross-correlation within the amplitude and warping scores, but also across these two modes of variation. The estimation procedure that was used in Hadjipantelis *et al.* (2014, 2015), however, relies on a prealignment of the curves that separates the vertical and the horizontal variation.

The major contribution of this paper is a new class of multivariate models that both eliminates the need for presmoothing and prealignment of samples and also enables estimation of cross-correlation between the co-ordinates of the amplitude effect. In the framework proposed, even if we do not assume any cross-correlation of the amplitude effects, the prediction of warping functions will still take the full multivariate sample into account, and the alignment will thus typically be superior to alignment of the individual co-ordinates.

2. Modelling and inference for misaligned multivariate functional data

Consider the multivariate functional observation in Fig. 1, which displays a walking sequence in three-dimensional space of a person equipped with 41 markers from the Carnegie Mellon

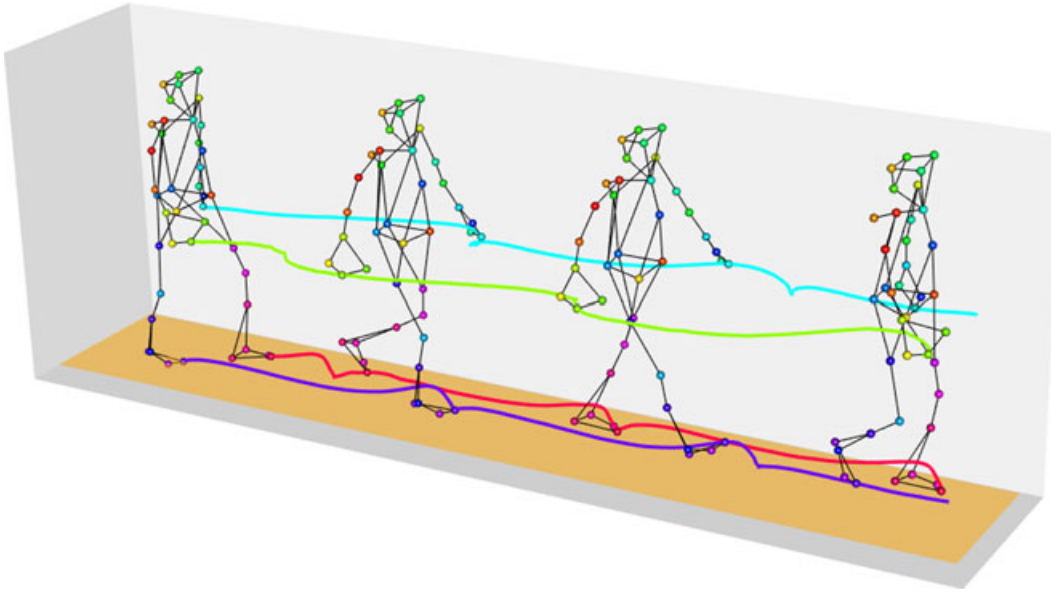

Fig. 1. Data from a motion tracking system where the spatial positions of 41 physical markers are tracked in three dimensions over time: a skeleton model based on the markers is displayed at four temporally equidistant points; the three-dimensional paths of hand and foot markers are displayed

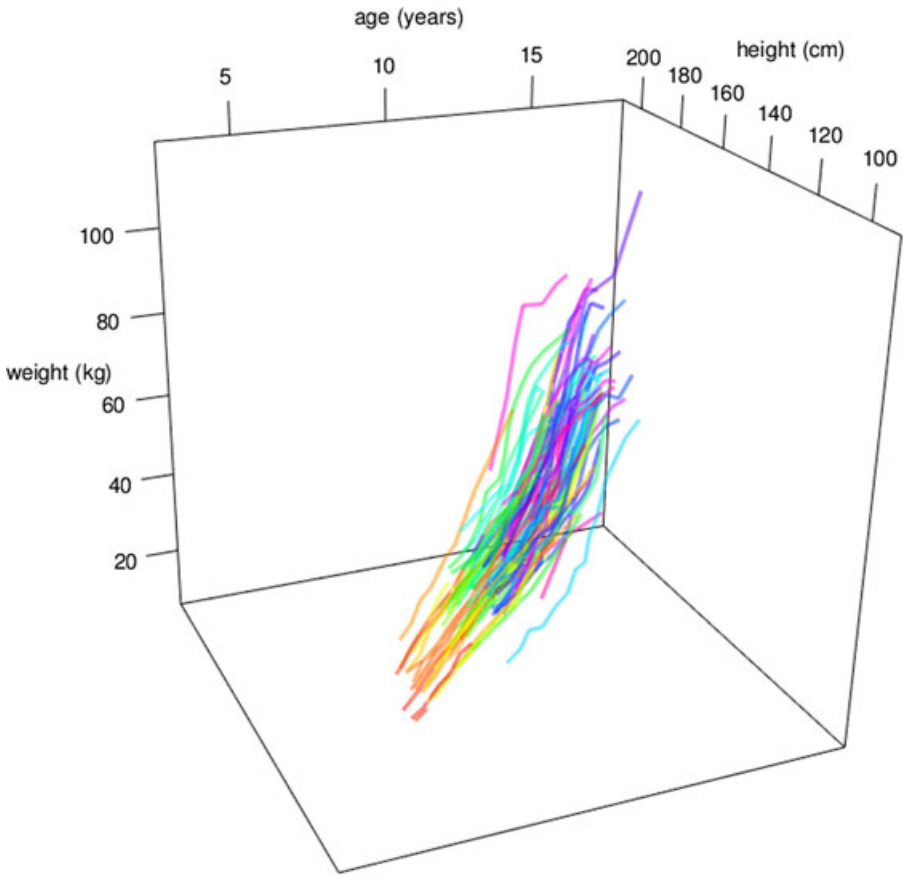

Fig. 2. Height and weight measurements over time for 106 healthy boys from the Copenhagen puberty study: each individual curve indicates a subject

University Graphics Laboratory motion capture database (`http://mocap.cs.cmu.edu/`). The observation is a curve in $\mathbb{R}^{123}$ recorded at 301 time points with a total of 36963 observed values (20 marker positions are missing because of occlusion).

This sample illustrates some of the challenges in analysing multivariate functional data. Firstly, a repetition of the walking cycle would in all likelihood produce a trajectory that is visually very similar to the sample, but it would differ in two aspects: the timing of movement and the movement path would be slightly different. Such differences in timing and path are random perturbations around the person's ideal walking cycle. A natural model for such data is thus a non-linear mixed effects model where movement timing is modelled as a random effect whose effect is only observed through the non-linear transformation of the movement path as a function of time, and the movement path variation is modelled as a stochastic process in $\mathbb{R}^{123}$. However, the very large number of observations in a single functional sample puts strong restrictions on the types of model that can be used. For example, the covariance matrix between the 41 markers at a single time point is 123×123, which in practice makes the problem of estimating a single unstructured covariance (7626 parameters) impossible.

Another example of multivariate functional data is longitudinal measurements of children's height and weight. Fig. 2 displays such data from the Copenhagen puberty study (Aksglaede

et al., 2009; Sørensen *et al.*, 2010). The data reflect the fact that height and weight are generally increasing functions during childhood and adolescence. Again, there will be a non-linear timing effect; observed age is a proxy for a biological or developmental age process of the child, and there will be systematic differences in observation values; taller and heavier children tend to stay taller and heavier than their peers. For height and weight data, one would typically have few observations per child, but the possibility of many children. Thus, the cross-covariance at a given time point could easily be estimated, and one could have a natural interest in inferring possible changes in the correlation between height and weight over time.

These two examples illustrate that the challenges of multivariate functional data can be very different. In what follows we shall introduce a class of models to analyse functional data containing both warp and amplitude variation. To make the model sufficiently flexible, we shall introduce generic models for random warping functions and dynamic cross-correlation structures that can approximate arbitrary structures, and whose resolution of approximation can be coarsened by reducing the number of free parameters.

2.1. Statistical model

We consider a set of N discrete observations of q-dimensional curves $\mathbf{y}_1, \ldots, \mathbf{y}_N : [0,1] \to \mathbb{R}^q$ from J subjects. The curves are assumed to be generated according to the model

$$\mathbf{y}_n(t) = \boldsymbol{\theta}_{f(n)}\{v_n(t)\} + \mathbf{x}_n(t), \qquad n = 1, \ldots, N. \tag{1}$$

Here $f : \{1, \ldots, N\} \to \{1, \ldots, J\}$ is a known function that maps sample number to subject number. The unknown fixed effects are subject-specific mean value functions $\boldsymbol{\theta}_j : [0,1] \to \mathbb{R}^q$ for $j = 1, \ldots, J$ that are modelled by using a spline basis assumed to be continuously differentiable. Typical choices are B-spline bases and Fourier bases. The phase variation is modelled by random warping functions $v_n = v(\cdot, \mathbf{w}_n) : [0,1] \to [0,1]$, which are parameterized by independent latent zero-mean Gaussian variables $\mathbf{w}_n \in \mathbb{R}^{m_\mathbf{w}}$ for $n = 1, \ldots, N$ with a common covariance matrix $\sigma^2 C$. Here $v : [0,1] \times \mathbb{R}^{m_\mathbf{w}} \to [0,1]$ is a prespecified function, which is assumed to be continuously differentiable in its second argument, and $m_\mathbf{w} \in \mathbb{N}$ is the dimension of the latent variable. The amplitude variation is modelled by independent zero-mean Gaussian processes $\mathbf{x}_n : [0,1] \to \mathbb{R}^q$ for $n = 1, \ldots, N$ with a common covariance function $\sigma^2 \mathcal{S}$. The unknown variance parameters are thus a scalar $\sigma^2 > 0$, a positive definite matrix $C \in \mathbb{R}^{m_\mathbf{w} \times m_\mathbf{w}}$ and a positive definite function $\mathcal{S} : [0,1] \times [0,1] \to \mathbb{R}^{q \times q}$. In Sections 2.2 and 2.3 we discuss models for the warping functions and the cross-covariance of the amplitude variation that are highly expressive, whereas the number of parameters to be estimated is kept at a moderate level.

We assume that the nth curve is observed at $m_n \in \mathbb{N}$ prefixed time points t_{nk}, which neither need to be equally spaced in time nor to be shared by the N samples. Stacking the m_n temporally discrete observations into a vector we have

$$\vec{\mathbf{y}}_n = \{\mathbf{y}_n(t_{nk}) + \boldsymbol{\varepsilon}_{nk}\}_{k=1}^{m_n} \in \mathbb{R}^{q m_n}, \qquad n = 1, \ldots, N, \tag{2}$$

where the observation noise is given by independent zero-mean Gaussian variables $\boldsymbol{\varepsilon}_{nk} \in \mathbb{R}^q$ with a common variance $\sigma^2 \mathbf{I}_q$. Here $\mathbf{I}_q \in \mathbb{R}^{q \times q}$ denotes the identity matrix.

The major structural difference of model (1) compared with conventional functional mixed effects models (Guo, 2002) is the inclusion of a warping effect. When compared with conventional methods for curve alignment, the model proposed differs by having a random amplitude effect, by modelling warping functions as random effects and by handling all effects simultaneously.

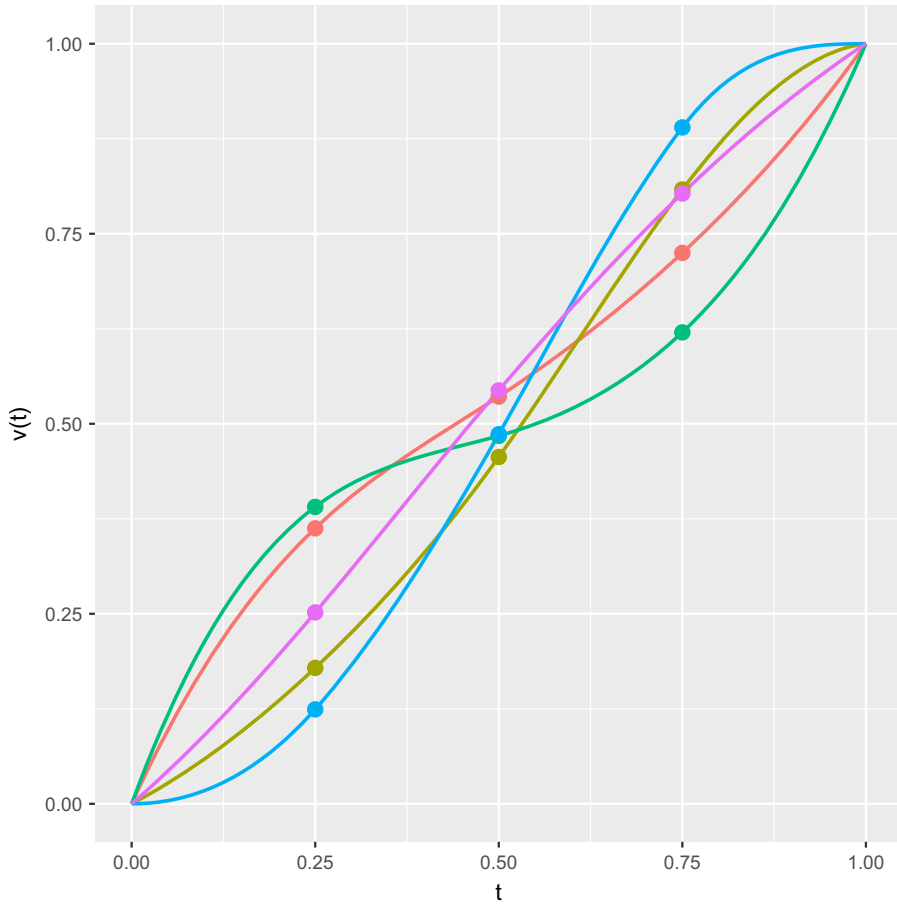

Fig. 3. Simulated warping functions with the covariance given by matrix (4): the warp values at the three interior anchor points are marked by points

2.2. Modelling warping functions

The success of the model relies on its ability to approximate the realizations of the true warping functions. To accomplish this, the warping functions v_n must be sufficiently versatile and able to approximate a large array of different warps. We achieve this by modelling warping functions as the identity mapping plus a deformation modelled by interpolating latent warp variables $\mathbf{w}_n \in \mathbb{R}^{m_\mathbf{w}}$ at prespecified (e.g. equidistant) anchor points t_k for $k = 1, \ldots, m_\mathbf{w}$:

$$v_n(t) = v(t, \mathbf{w}_n) = t + \mathcal{E}_{\mathbf{w}_n}(t), \tag{3}$$

where the interpolation function $\mathcal{E}_\mathbf{w}$ can, for example, be a linear or a cubic spline.

The behaviour of the predicted warping functions will be determined by the combination of interpolation method (and corresponding boundary conditions) and the estimated covariance of the latent variables $\mathbf{w}_n$. Throughout this paper we shall use cubic spline interpolation of the latent variables. If we think of the parameterization of the nth sample, $v_n(t)$, as the internal time of the sample, it is often natural to assume that the internal time is always moving forwards. To ensure this, we shall predict the latent variables $\mathbf{w}_n$ by using constrained optimization such

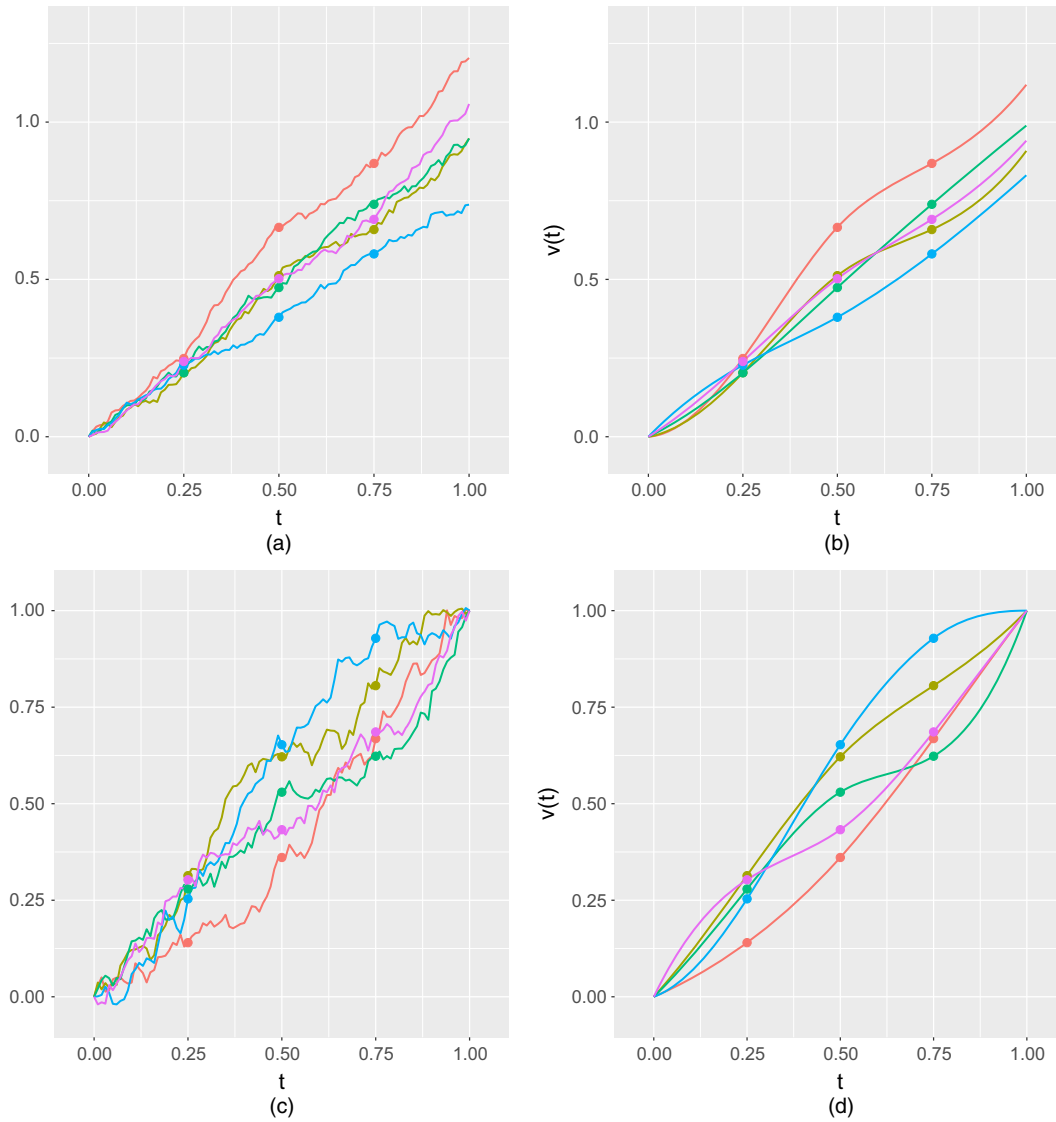

Fig. 4. Constructions of warping functions from stochastic processes with parametric covariances: (a) simulated trajectories of a unit drift Brownian motion with scale 0.1; (b) warping functions by using a unit drift Brownian motion model with $m_{\mathbf{w}} = 3$ interior equidistant anchor points, fixed interpolation at the left boundary and extrapolation of the rightmost deviation at the right end point; (c) simulated trajectories of a unit drift Brownian bridge with scale 0.2; (d) warping functions by using the unit drift Brownian bridge model with $m_{\mathbf{w}} = 3$ interior equidistant anchor points and fixed interpolation at the boundary

that the sequence will be increasing along the corresponding anchor points. But, for cubic interpolation, a sequence of increasing values at the interpolation points is not sufficient to ensure a monotone interpolation function. To force increasing warping functions we shall use the Hyman filter (Hyman, 1983) to ensure that the entire warping function is increasing. For some types of data, it may be meaningful to have warps that can go backwards in time, or it may be useful to include this option to account for uncertainty in the model if the observed signals

contain features where the matching is highly ambiguous. Such types of warp models will not be considered in this paper.

The covariance matrix of the latent variables will determine the regularity of the predicted warping functions. When the number of latent variables $m_{\mathbf{w}}$ is small compared with the number of functional samples N and the number of sampling points $m_1, \ldots, m_N$ for the functional samples, we can assume an unstructured covariance and estimate the corresponding $(m_{\mathbf{w}}^2 + m_{\mathbf{w}})/2$ variance parameters. If the structure of the warping functions is of key interest, we may be able to study the underlying mechanism by estimating an unstructured covariance matrix. Consider for example the simulated warping functions that are shown in Fig. 3. These warping functions use the increasing cubic spline construction detailed above with $m_{\mathbf{w}} = 3$ interior equidistant anchor points, fixed boundary points and covariance matrix

$$\begin{pmatrix} 0.005 & 0 & -0.004 \\ 0 & 0.001 & 0 \\ -0.004 & 0 & 0.005 \end{pmatrix}. \tag{4}$$

The interpretation of the strong negative covariance between the first and third anchor point suggests a burn-out type of process where samples that are ahead initially slow down towards the end and vice versa. The low variance of the middle anchor point suggests that the individual samples are largely synchronized around this time.

In many cases, one can choose a specific interpolation method and specify a reasonable parametric covariance for the latent variables based on properties of the data. It is, for example, often natural to think of warping processes as accumulations of small errors causing desynchronization of the set of observed trajectories that all started in the same state. Thinking of Gaussian processes, Brownian motion with linear unit drift would offer a simple model for phenomena where errors are accumulating and increasing the desynchronization of samples over time. Simulations of unit drift Brownian motions are shown in Fig. 4(a) and the corresponding simulations of warping functions from $m_{\mathbf{w}} = 3$ interior equidistant anchor points, fixed left boundary point and linear extrapolation of the deviation of the rightmost anchor point at the right boundary point are shown in Fig. 4(b).

Suppose that we are analysing longitudinal data of children's heights where we could think of the warping function as the developmental (height) age of the child. At conception (approximately -9 months of age), where the child is merely a fertilized egg, all children are the size of a grain of sand and their developmental ages are synchronized. As the children become older the desynchronization of their developmental ages increases. This can, for example, be seen by the vast variation between the age of onset of puberty. The unit drift Brownian motion warp model seems like a very suitable model for this desynchronization.

Other types of data may give rise to other models. Consider an experiment that records repetitions of a walking sequence such as the data in Fig. 1, and assume that all sequences start from the same pose and end after two completed gait cycles. For such data, the desynchronization is not increasing over time since beginning and end poses are synchronized, but we would expect maximum desynchronization around the middle of the gait cycle window. In this setting, a more suitable model would be a unit drift Brownian bridge as illustrated in Figs 4(c) and 4(d).

Like other hyperparameters, the number of anchor points is a choice of modelling. However, a low number of anchor points (e.g. 3–5) will generate a class of warp functions that is sufficiently flexible for many applications; we used $m_{\mathbf{w}} = 3$ in all applications that are presented in this paper. If, however, local variation is very strong and complex and the observed functional samples carry sufficiently clear information about the systematic shapes to recover such complex warps, a higher number of anchor points should be used.

2.3. Dynamic covariance structures

In the previous section we modelled the covariance structure of smooth warping functions and saw how we could use domain-specific knowledge of the data to choose models with few parameters. Even though the nature of the additive amplitude variation components $\mathbf{x}_n$ from model (1) is different, we can extend these ideas to construct parametric, low dimensional cross-covariance structures that are sufficiently expressive to model a wide array of cross-covariance structures over time.

Proposition 1. Let $f : [0,1] \times [0,1] \to \mathbb{R}_+$ be a positive definite function on the temporal domain $[0,1]$. Let $0 = t_1 < \ldots < t_l = 1$ be anchor points, let $A_1, \ldots, A_l \in \mathbb{R}^{q \times q}$ be a set of symmetric positive definite matrices, and for each $t \in [0,1]$ define $B_t \in \mathbb{R}^{q \times q}$ as the unique positive definite matrix satisfying

$$B_t^{\mathrm{T}} B_t = \frac{t_{k+1} - t}{t_{k+1} - t_k} A_k + \frac{t - t_k}{t_{k+1} - t_k} A_{k+1} \qquad \text{for } t \in [t_k, t_{k+1}]. \tag{5}$$

For all $s, t \in [0,1]$, define $K(s,t) = f(s,t) B_s^{\mathrm{T}} B_t \in \mathbb{R}^{q \times q}$. Then the function $K : [0,1] \times [0,1] \to \mathbb{R}^{q \times q}$ is positive definite.

Proof. First we remark that, since the space of positive definite matrices is a convex cone, the linear interpolation $B_t^{\mathrm{T}} B_t$ is also positive definite, and we may take B_t as the positive square root. To prove that $K : [0,1] \times [0,1] \to \mathbb{R}^{q \times q}$ is positive definite it suffices to show that the associated finite dimensional marginal matrices are positive definite. Thus, given $s_1, \ldots, s_m \in [0,1]$ we let the block matrix $\mathbf{V} \in \mathbb{R}^{qm \times qm}$ be defined by

$$\mathbf{V} = \begin{pmatrix} B_{s_1}^{\mathrm{T}} f(s_1, s_1) B_{s_1} & B_{s_1}^{\mathrm{T}} f(s_1, s_2) B_{s_2} & \cdots & B_{s_1}^{\mathrm{T}} f(s_1, s_m) B_{s_m} \\ B_{s_2}^{\mathrm{T}} f(s_2, s_1) B_{s_1} & B_{s_2}^{\mathrm{T}} f(s_2, s_2) B_{s_2} & \ldots & \\ \vdots & & \ddots & \vdots \\ B_{s_m}^{\mathrm{T}} f(s_m, s_1) B_{s_1} & B_{s_m}^{\mathrm{T}} f(s_m, s_2) B_{s_2} & \ldots & B_{s_m}^{\mathrm{T}} f(s_m, s_m) B_{s_m} \end{pmatrix}. \tag{6}$$

By straightforward calculations we have $\mathbf{V} = \mathbf{B}^{\mathrm{T}} (\mathbf{F} \otimes \mathbf{I}_q) \mathbf{B}$, where $\mathbf{B} \in \mathbb{R}^{qm \times qm}$ is the block diagonal matrix of $\{B_{s_1}, \ldots, B_{s_m}\}$ and

$$\mathbf{F} = \begin{pmatrix} f(s_1, s_1) & \cdots & f(s_1, s_m) \\ \vdots & \ddots & \vdots \\ f(s_m, s_1) & \cdots & f(s_m, s_m) \end{pmatrix}. \tag{7}$$

For $z \in \mathbb{R}^{qm} \setminus \{0\}$ we must show that $z^{\mathrm{T}} \mathbf{V} z > 0$. Setting $u = \mathbf{B} z \neq 0$ and using that $\mathbf{F}$ is positive definite by assumption we have $z^{\mathrm{T}} \mathbf{V} z = u^{\mathrm{T}} (\mathbf{F} \otimes \mathbf{I}_q) u > 0$. □

Proposition 1 gives a general framework for constructing dynamical covariance functions, and it is simple to construct parametric models that enable estimation of time varying cross-correlations in a statistical setting. In the statement of proposition 1 we assumed a common marginal covariance function f along all co-ordinates. The idea of modelling a cross-covariance structure by linearly interpolating cross-covariances at specific points seamlessly extends to multivariate diagonal covariance functions (i.e. no cross-covariances), such that the individual co-ordinates of the functional samples may be modelled by using different types of covariance function or different parameters.

3. Estimation

Direct likelihood inference in model (1) is not feasible as the model contains non-linear latent variables in combination with possibly very large data sizes. Instead we propose a maximum likelihood estimation procedure based on iterative local linearization (Lindstrom and Bates, 1990). The procedure is a multivariate extension of the estimation procedure that was described in Raket *et al.* (2014), however, with an improved estimation of fixed effects.

The estimation procedure consists of alternating steps of

(a) estimating fixed effects (i.e. spline coefficents) and predicting the most likely warp variables given the data and current parameter estimates and
(b) estimating variance parameters from the locally linearized likelihood function around the maximum *a posteriori* predictions $\mathbf{w}_1^0, \ldots, \mathbf{w}_N^0$ of the warp variables.

The linearization in the latent Gaussian warp parameters $\mathbf{w}_1, \ldots, \mathbf{w}_N$ means that we approximate the non-linearly transformed probability density by the density of a linear combination of multivariate Gaussian variables. The estimation procedure is thus a Laplace approximation of the likelihood, and the quality of the approximation is approximately second order (Wolfinger, 1993).

3.1. Predicting warps

In the first step of the estimation procedure we want to predict the most likely warps from model (1) given the current parameter estimates. The negative log-posterior for a single functional sample is proportional to

$$(\vec{\gamma}_{\mathbf{w}_n} - \vec{\mathbf{y}}_n)^{\mathrm{T}} (\mathbf{I}_{qm_n} + S_n)^{-1} (\vec{\gamma}_{\mathbf{w}_n} - \vec{\mathbf{y}}_n) + \mathbf{w}_n^{\mathrm{T}} C^{-1} \mathbf{w}_n \tag{8}$$

where $\vec{\gamma}_{\mathbf{w}_n} \in \mathbb{R}^{qm_n}$ is the stacked vector $\{\boldsymbol{\theta}_{f(n)}\{v(t_{nk}, \mathbf{w}_n)\}\}_{k=1}^{m_n}$ and $S_n \in \mathbb{R}^{qm_n \times qm_n}$ is the amplitude covariance $\{\mathcal{S}(t_{nj}, t_{nk})\}_{j,k=1,\ldots,m_n}$ at the sample points. The issue of predicting warps is thus a non-linear least squares problem that can be solved by conventional methods.

3.2. Estimating variance parameters

Since $\boldsymbol{\theta}_{f(n)} \circ v(t_{nk}, \cdot)$ are smooth functions for all $n = 1, \ldots, N$ and $k = 1, \ldots, m_n$ we can linearize model (1) around a given prediction $\mathbf{w}_n^0$ by using the first-order Taylor expansion. The linearization is given by

$$\boldsymbol{\theta}_{f(n)}\{v(t_{nk}, \mathbf{w}_n)\} \approx \boldsymbol{\theta}_{f(n)}\{v(t_{nk}, \mathbf{w}_n^0)\} + \partial_t \boldsymbol{\theta}_{f(n)}\{v(t_{nk}, \mathbf{w}_n^0)\} (\nabla_{\mathbf{w}} v(t_{nk}, \mathbf{w}_n^0))^{\mathrm{T}} (\mathbf{w}_n - \mathbf{w}_n^0). \tag{9}$$

For the discrete observation of the nth curve this gives a linearization of model (1) as a vectorized linear mixed effects model of the form

$$\vec{\mathbf{y}}_n \approx \vec{\gamma}_{\mathbf{w}_n^0} + Z_n(\mathbf{w}_n - \mathbf{w}_n^0) + \vec{\mathbf{x}}_n + \vec{\varepsilon}_n, \qquad n = 1, \ldots, N, \tag{10}$$

where $\vec{\gamma}_{\mathbf{w}_n^0}, \vec{\mathbf{x}}_n, \vec{\varepsilon}_n \in \mathbb{R}^{qm_n}$ are the stacked vectors

$$\begin{aligned} \vec{\gamma}_{\mathbf{w}_n^0} &= \{\boldsymbol{\theta}_{f(n)}\{v(t_{nk}, \mathbf{w}_n^0)\}\}_{k=1}^{m_n}, \\ \vec{\mathbf{x}}_n &= \{\mathbf{x}_n(t_{nk})\}_{k=1}^{m_n}, \\ \vec{\varepsilon}_n &= \{\varepsilon_{nk}\}_{k=1}^{m_n}, \end{aligned}$$

and $Z_n \in \mathbb{R}^{qm_n \times m_{\mathbf{w}}}$ is the rowwise stacked matrix

$$Z_n = \{\partial_t \boldsymbol{\theta}_{f(n)}\{v(t_{nk}, \mathbf{w}_n^0)\} \nabla_{\mathbf{w}} v(t_{nk}, \mathbf{w}_n^0)\}_{k=1}^{m_n}.$$

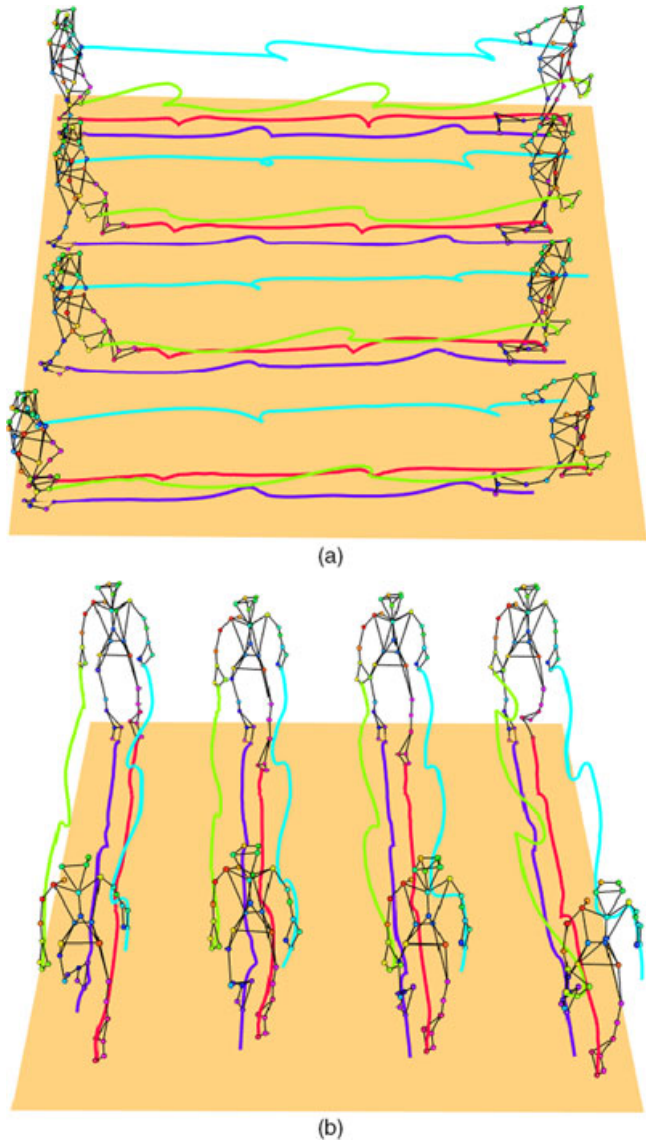

Fig. 5. (a) Side and (b) frontal view of the motion trajectories of four walking sequences performed by the same participant

In the approximative model (10) twice the negative profile log-likelihood $l(\sigma^2, C, \mathcal{S})$ for the variance parameters is given by

$$\sum_{n=1}^{N} [qm_n \log(\sigma^2) + \log\{\det(V_n)\} + \sigma^{-2}(\vec{\mathbf{y}}_n - \vec{\gamma}_{\mathbf{w}_n^0} + Z_n \mathbf{w}_n^0)^{\mathrm{T}} V_n^{-1} (\vec{\mathbf{y}}_n - \vec{\gamma}_{\mathbf{w}_n^0} + Z_n \mathbf{w}_n^0)], \tag{11}$$

where $V_n = Z_n C Z_n^{\mathrm{T}} + S_n + \mathbf{I}_{qm_n}$ with $S_n = \{\mathcal{S}(t_{nj}, t_{nk})\}_{j,k=1,\ldots,m_n}$. In particular, the profile maximum likelihood estimate for σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{qm} \sum_{n=1}^{N} (\vec{\mathbf{y}}_n - \vec{\gamma}_{\mathbf{w}_n^0} + Z_n \mathbf{w}_n^0)^{\mathrm{T}} V_n^{-1} (\vec{\mathbf{y}}_n - \vec{\gamma}_{\mathbf{w}_n^0} + Z_n \mathbf{w}_n^0)$$

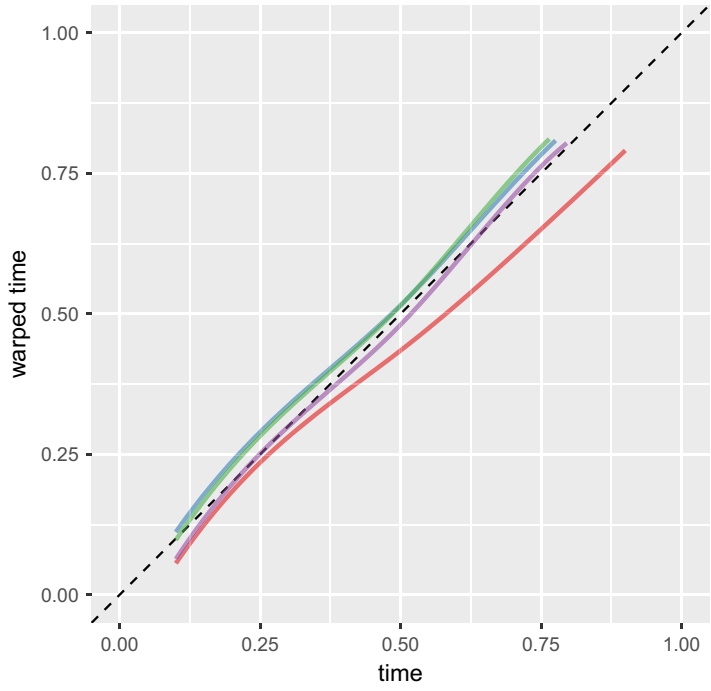

Fig. 6. Predicted warping functions for the motion capture data: ——, repetition 1; ——, repetition 2; ——, repetition 3; ——, repetition 4

where $m = \Sigma_{n=1}^{N} m_n$ is the total number of observations. Estimation of the variance parameters C and $\mathcal{S}$ related to the warping and amplitude effects is done by using the profile likelihood $l(\hat{\sigma}^2, C, \mathcal{S})$.

3.3. *Estimating fixed effects*

As the fixed effects are given by spline bases, estimation of these can be handled within the framework of linear Gaussian models, remembering that basis functions should be evaluated at warped time points $v_n(t_{nk})$. Since $v_n(t_{nk}) = v(t_{nk}, \mathbf{w}_n)$ changes with $\mathbf{w}_n$, we are required to recalculate the spline basis matrix for each new prediction of $\mathbf{w}_n$. This estimation improves that of Raket *et al.* (2014), which used a pointwise estimation based on the inverse warp that ignored the amplitude variance of the curves.

There is no closed form expression for the maximum likelihood estimator of the fixed effects in the linearized model, since spline coefficients also enter the variance terms through the matrices Z_n, as can be seen in equation (11). However, by construction Z_n is linear in the spline coefficients so estimation can be done by using an expectation–maximization (EM) algorithm. The details of these calculations can be found in Appendix C.

In practice, the estimation in the linearized model can be approximated by estimating from the posterior likelihood (8) which gives a computationally efficient closed form solution. The difference between these two approaches is that the EM algorithm takes the uncertainty in prediction of $\mathbf{w}_n$ into account and is guaranteed to decrease the linearized likelihood (11). However, for a moderate number of warp parameters, there should only be a small conditional variance on $\mathbf{w}_n$.

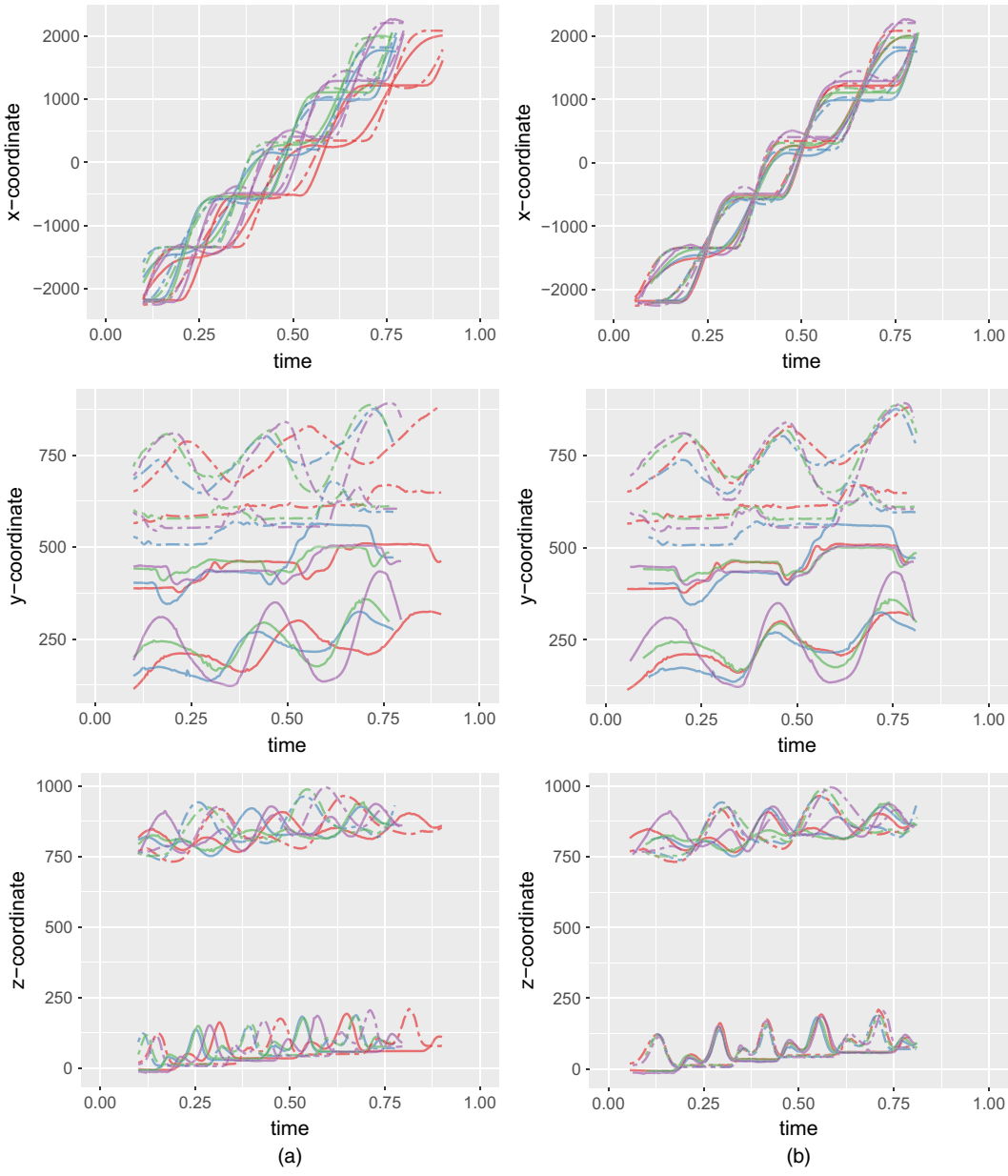

Fig. 7. (a) Observed and (b) aligned curves from the motion capture data (data values are the raw values from the tracking system): ——, repetition 1; ——, repetition 2; ——, repetition 3; ——, repetition 4; - - - -, left-hand side; ——, right-hand side

In the data applications that are presented in the following sections, we estimated fixed effects from the posterior likelihood. In the last application on hand movements, these posterior likelihood estimates were used to initialize the likelihood optimization which were subsequently fine tuned by the EM algorithm with a single update per warp prediction. This was done to evaluate whether improved likelihood estimates could be obtained, but the EM algorithm offered only a very slight improvement in linearized likelihood.

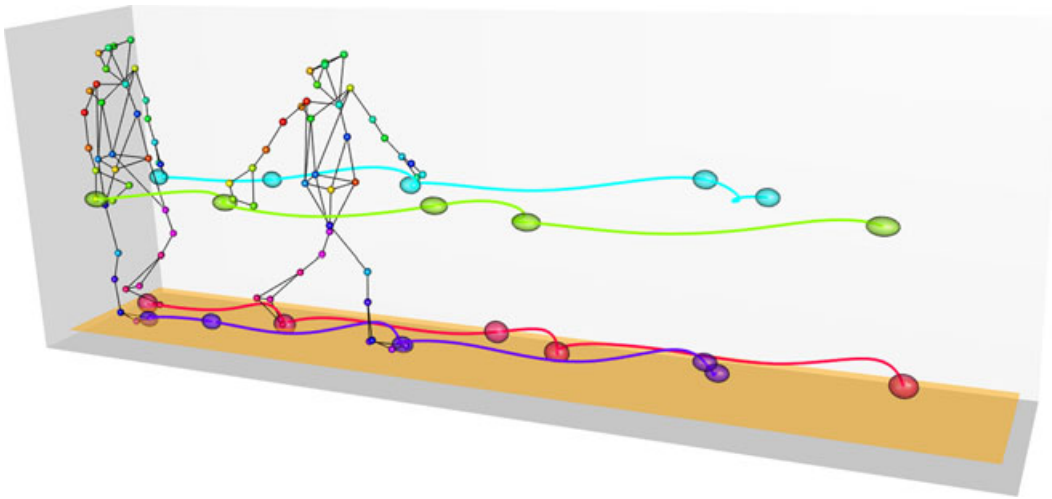

Fig. 8. Estimated mean trajectories with five temporally equidistant ellipsoids indicating 95% (marginal) confidence areas: two of the intermediate body poses of the fourth sample have been added as a reference

4. Applications

4.1. Motion capture data

4.1.1. Data and model

The motion data consist of four 12-dimensional functional objects. The curves consist of a total of 1284 temporal observations in $\mathbb{R}^{12}$. As can be seen in Fig. 5, the trajectories start and end at different places during the gait cycle. To handle this structure, time was scaled to the interval $[0, 1]$ such that all samples began at 0.1, and such that the temporally longest trajectory ended at 0.9. We included random-shift parameters s_n in our warping functions to model these different temporal onsets of the gait cycle. The shifts s_n were modelled as Gaussian random variables. The full model is

$$\mathbf{y}_n(t) = \boldsymbol{\theta}\{v(t, \mathbf{w}_n, s_n)\} + \mathbf{x}_n(t) \tag{12}$$

where $\boldsymbol{\theta} : [0, 1] \to \mathbb{R}^{12}$ is the mean curve for the observations (modelled by using a three-dimensional B-spline basis with 30 interior anchor points) and the warping function v is given by

$$v(t, \mathbf{w}_n, s_n) = t + s_n + \mathcal{E}_{\mathbf{w}_n}(t)$$

where $\mathcal{E}_{\mathbf{w}_n}$ is an increasing cubic spline interpolation (Hyman filtered) of $\mathbf{w}_n$ at $m_{\mathbf{w}} = 3$ equidistant anchor points. No subject-specific effects were included as all responses were recorded from the same individual. The amplitude effect $\mathbf{x}_n$ was modelled as a Gaussian process with a Matérn covariance $f_{\text{Matérn}(2,\kappa)}(s, t)$ with second-order smoothness, assuming independent co-ordinates and a common range parameter κ (see equation (16) in Appendix B). We assumed different scaling parameters for each of the 12 co-ordinates of $\mathbf{x}_n$. Since the data are roughly cut to include two gait cycles, we would expect high synchronization of the start and end poses in percentual time when corrected for the different onsets. Therefore, latent variables $\mathbf{w}_n$ were modelled as discretely observed Brownian bridges with a single scale parameter.

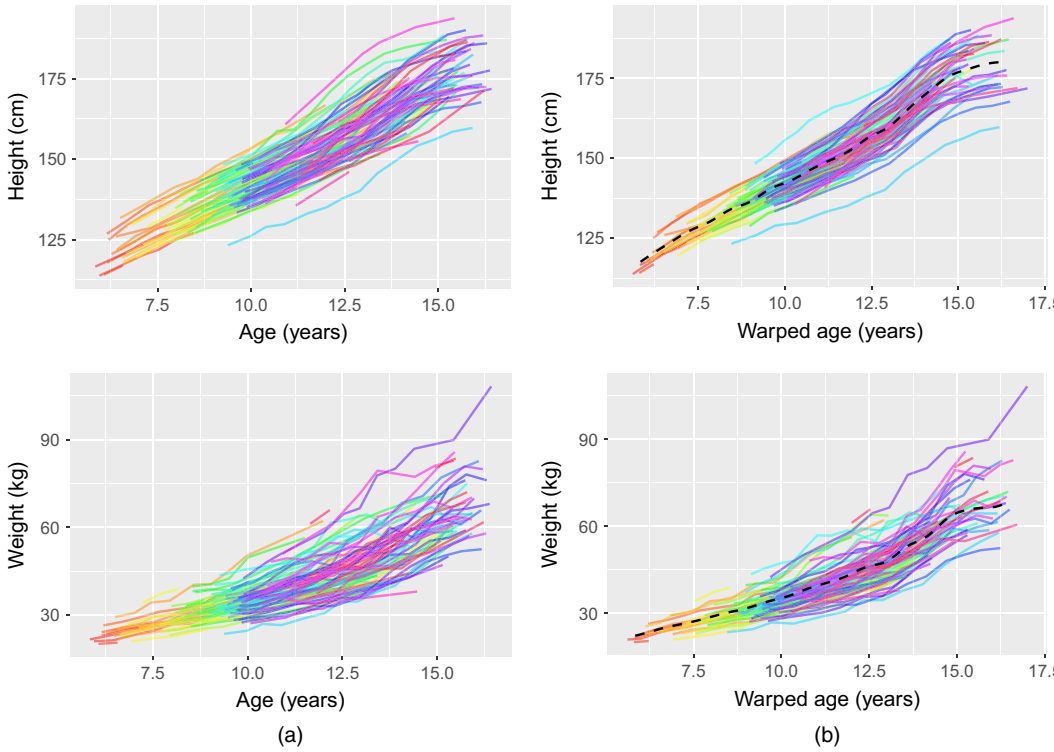

Fig. 9. (a) Observed and (b) aligned height and weight curves from the Copenhagen puberty study: – – –, estimated template curves

4.1.2. Results

The predicted warping functions are shown in Fig. 6, and the corresponding aligned samples are shown in Fig. 7. The samples are nicely aligned; in particular, the regular elevation profiles of the left and right feet seem very well aligned. The remaining signals have their key features aligned, with the residual variation evenly spread out across the co-ordinates. This is a feature of the simultaneous multivariate fitting, where the best alignment given the variation in the different co-ordinates is found. Individual alignment of the co-ordinates would produce warping functions that overfitted the individual aspects of the movement. In Fig. 8, we have displayed the estimated mean trajectories $\boldsymbol{\theta}$ and illustrated the uncertainty after alignment by 95% prediction ellipsoids for the amplitude effect $\mathbf{x}_n$.

4.2. Height and weight data

Consider the height and weight measurements from the Copenhagen puberty study (Aksglaede *et al.*, 2009; Sørensen *et al.*, 2010) that are shown in Fig. 2. The data contain 960 pairs of height and weight measurements for 106 healthy Danish boys. The individual amplitude effects in the data set are clearly visible in the form of systematic deviations from the mean. The data also contain warping variation in the sense that age is a proxy for developmental age; each boy has his own internal clock that determines, for example, the onset of puberty. Alignment for this warping effect would then align the pubertal growth spurts that are visible as steep height increase in the individual boys occurring in the period 11–14 years.

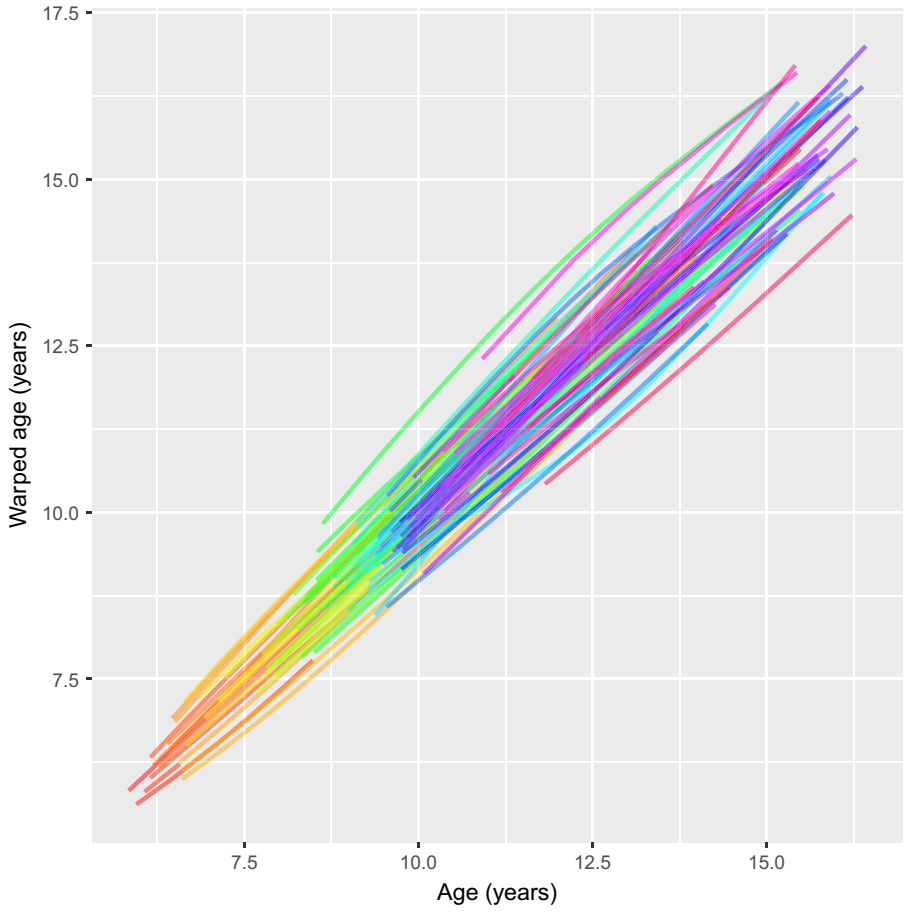

Fig. 10. Predicted warping function corresponding to the data in Fig. 9

4.2.1. Modelling

Whereas height is a naturally increasing function of age, weight is not necessarily. However, looking at the 2014 Danish weight reference Tinggaard *et al.* (2014), we see a convex increase in the cross-sectional mean weight curve in the relevant age interval. Based on this, we modelled $\boldsymbol{\theta}$ by using an increasing spline (integrated quadratic B-splines) basis with 20 equidistant internal knots in the age interval $[5, 17]$ years in both dimensions. The warping functions (3) were modelled as increasing cubic (Hyman filtered) splines with $m_{\mathbf{w}} = 3$ equidistant internal anchor points in the age interval $[5, 20]$ years and extrapolation at the right boundary point as in Fig. 4(b). The latent variables $\mathbf{w}_n$ were modelled as discretely observed Brownian motions with a single scale parameter. The temporally increasing variance of the Brownian motion seems a good model for developmental age where we would expect high initial synchronization, and up to several years desynchronization at the onset of puberty.

To model the amplitude variation, we used a dynamic cross-covariance with equidistant knots at $\{5, 10, 15, 20\}$ years as described in proposition 1, i.e.

$$\mathcal{S}(s,t) = f_{\text{Matérn}(2,\kappa)}(s,t) B_s^{\mathrm{T}} B_t.$$

The temporal covariance structure $f_{\text{Matérn}(2,\kappa)}(s,t)$ is the Matérn covariance function with fixed

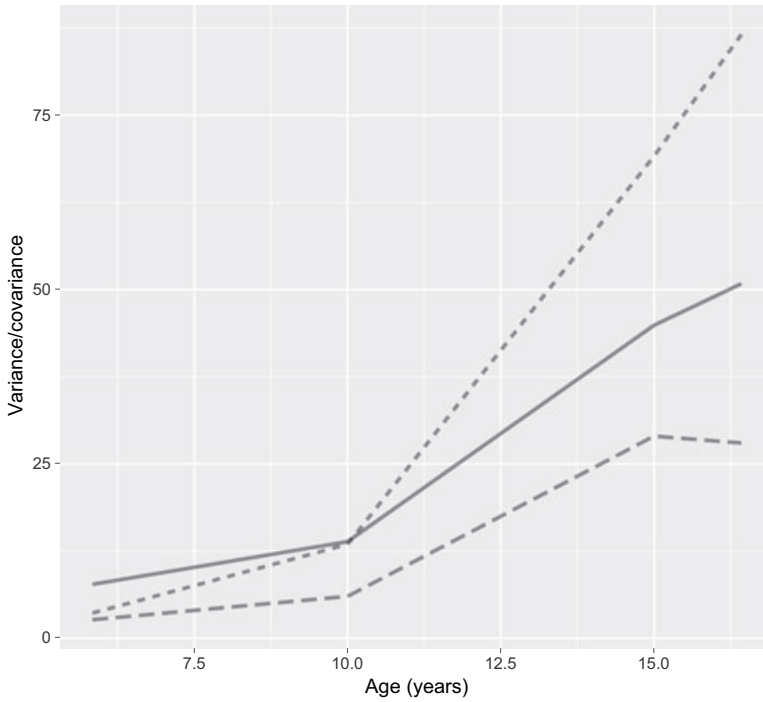

Fig. 11. Estimated marginal variances and cross-covariance functions of age for the height and weight data in Fig. 9 (the marginal variances also include the error variance): ———, height variance; - - - - - -, weight variance; — — —, height–weight covariance

smoothness parameter $\alpha = 2$ and unknown range parameter κ; see equation (16) in Appendix A. This implies twice-differentiable sample paths of $\mathbf{x}_i$, which is a reasonable assumption given the nature of the data. Furthermore, since we expected heterogeneous variances of the measurement error $\boldsymbol{\varepsilon}_{nk}$ on height and weight in equation (2), we extended the model with a parameter $\rho > 0$ such that

$$\operatorname{var}(\boldsymbol{\varepsilon}_{nk}) = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix}.$$

This gives a total of 14 parameters describing the cross-covariance model.

4.2.2. Results

The aligned samples and estimated means are displayed in Fig. 9(b), and the corresponding predicted warping functions can be found in Fig. 10. We see that the individual growth curves are now aligned more tightly than before; in particular the pubertal height spurts seem to be well aligned. Although the shapes of the curves are well aligned, the model still allowed for considerable amplitude variation to be left after warping. This is as it should be; for increasing curves such as these a perfect fit could be achieved by warping, but the result would be meaningless and indicate that developmental age could be perfectly determined from a single measurement of a child's height. Given the proposed model-based separation of amplitude and warping effects that are induced by the maximum likelihood estimates, the information that is contained in a child's longitudinal data about the child's developmental age can be quantified through the posterior distribution of the warping effects.

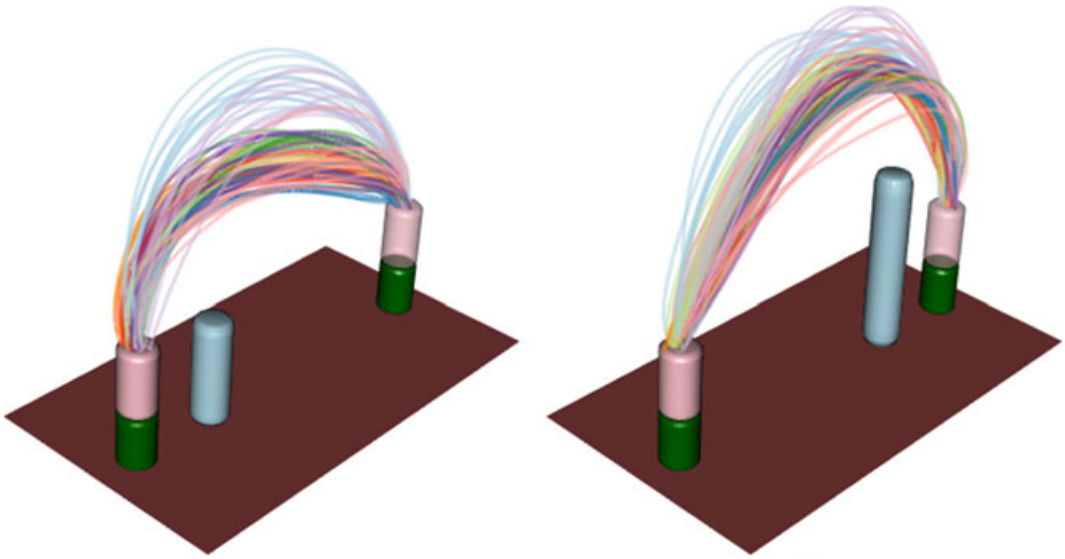

Fig. 12. Recorded movement paths in the experiment with (a) a small obstacle 15 cm from the starting position and (b) a tall obstacle 45 cm from the starting position: ——, participant 1; ——, participant 2; ——, participant 3; ——, participant 4; ——, participant 5; ——, participant 6; ——, participant 7; ——, participant 8; ——, participant 9; ——, participant 10

The estimated covariance structure is shown in Fig. 11. As we would expect, height and weight variances increase with age. The covariance increases at a slower rate and has a slight decrease after 15 years, giving a correlation of 0.42 at 16.5 years.

4.3. Arm movement data

Our third example is an analysis of human arm movements in obstacle avoidance tasks. Hand movement paths in two experimental conditions are displayed in Fig. 12. In each experimental condition, a wooden cylindrical object (pink) at a starting position (the green cylinder) was to be moved 60 cm forwards and placed on a target cylinder. Between the starting and target positions, a cylindrical obstacle was placed. The obstacle height (small, medium or tall) and obstacle position (five equidistant positions between the starting and target positions) varied with experimental condition. A total of 15 obstacle avoidance conditions were performed plus a control condition with no obstacle. 10 right-handed participants performed 10 repetitions of each experimental condition, and the spatial position of the hand was recorded at a sampling rate of 110 Hz. The data set thus consists of 1600 functional samples with a total of $m = 175\,535$ three-dimensional sampling points giving a total sample size of 526 605 observations. The present data set is described in detail in Grimme (2014), and the experiment is a refined version of the experiment that was described in Grimme *et al.* (2012). The data set is available through a public repository (`https://github.com/larslau/Bochum_movement_data`).

4.3.1. Data processing and modelling

We analysed the data separately for the 16 experimental conditions. Following the convention for modelling human motor control data, time was modelled as percentual time rather than

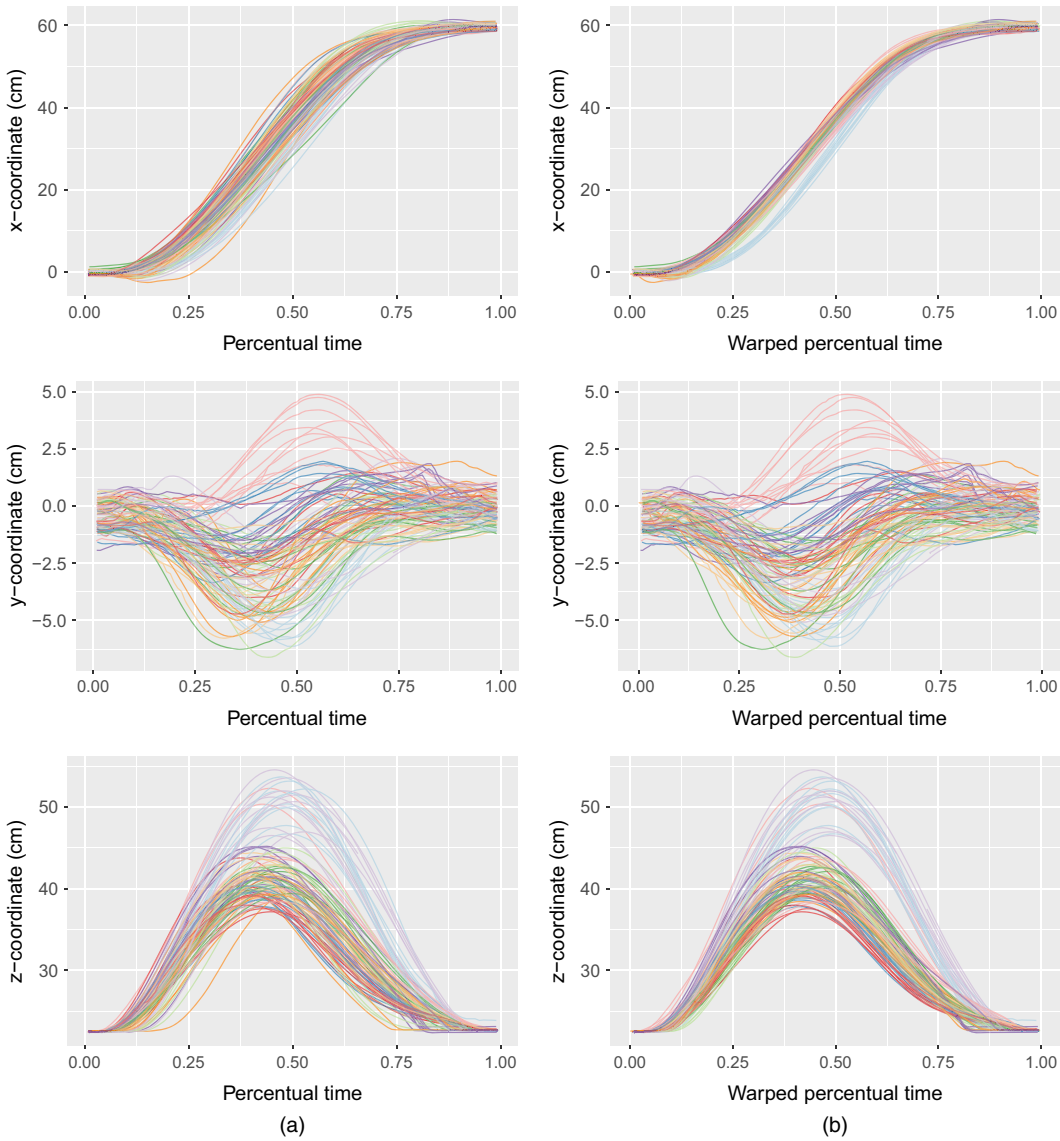

Fig. 13. Data from the experiment with a small obstacle 30 cm from the starting position plotted in (a) percentual time and (b) warped percentual time: the colouring follows the colouring in Fig. 12

observed time. This means that all movement time intervals were scaled to $[0, 1]$, such that 0 corresponds to the onset of the movement and 1 corresponds to the end of the movement. We used model (1) to model the data separately for the 16 different experimental conditions. The mean path $\boldsymbol{\theta}_j$ for the jth participants was modelled in a cubic B-spline basis with 21 interior knots. We modelled the warping functions (3) as increasing cubic spline interpolations (Hyman filtered) with $m_{\mathbf{w}} = 3$ equidistant anchor points. The choice of three knots was evaluated, and found optimal, in terms of the cross-validation set-up that is described in the classification study below. The latent variables $\mathbf{w}_n$ were modelled as discretely observed Brownian bridges with a single scale parameter, because of the fixed end points of the data.

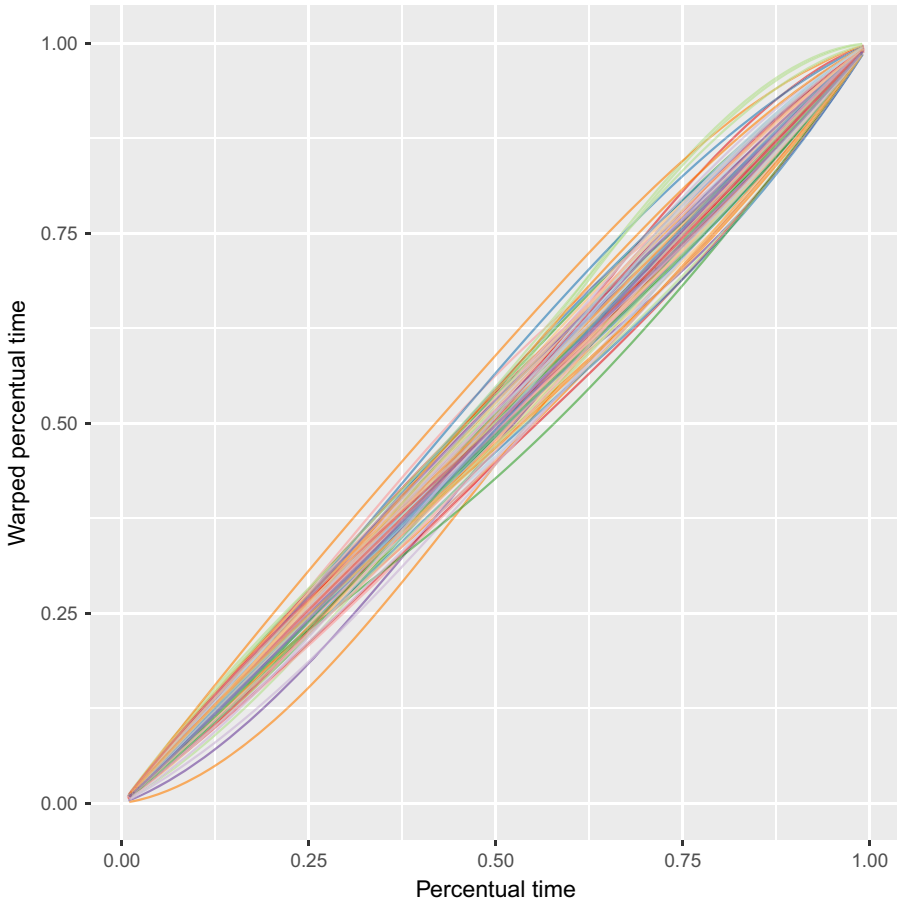

Fig. 14. Predicted warping functions corresponding to the alignment in Fig. 13

The amplitude variation was modelled by using a dynamic cross-correlation model with knots at $\{0, 0.4, 0.6, 1\}$ as described in proposition 1, i.e.

$$\mathcal{S}(s,t) = f_{\text{mixture}(a)}(s,t)\, f_{\text{Matérn}(\alpha,\kappa)}(s,t)\, B_s^{\mathrm{T}} B_t.$$

The temporal covariance structure is given as a combination of stationary and bridge Matérn serial correlation with mixture parameter a, smoothness parameter α and range parameter κ. The details of this covariance structure are described in equations (15) and (16) in Appendix B. This dynamic cross-correlation structure has 27 free parameters.

The knot positions $\{0, 0.4, 0.6, 1\}$ were chosen such that we could model a change in cross-correlation structure around the middle of the movement in percentual time, in particular the change that happens when the movement progresses from lift to descend. The concept of isochrony (Grimme *et al.*, 2012) suggests that the times where the peak heights are reached are largely invariant to obstacle height and placement, and for the given data the peak heights generally occur for $t \in (0.4, 0.6)$; see for example Fig. 13.

Fig. 13(a) displays the observed x-, y- and z-co-ordinates in a single experimental condition as functions of percentual time. Fig. 13(b) displays the co-ordinates in predicted warped percentual time. We see that the x- and z-co-ordinates are very well aligned within participant, and that

Table 1. Parameter estimates for the arm movement data

d (cm)	*Obstacle*	σ	α	κ	a	$\sigma\tau$
15.0	Small	0.0012	1.432	0.157	19.56	0.0519
	Medium	0.0012	1.749	0.120	24.60	0.0525
	Tall	0.0013	1.627	0.124	22.54	0.0502
22.5	Small	0.0013	1.788	0.128	25.13	0.0531
	Medium	0.0011	1.638	0.139	58.20	0.1177
	Tall	0.0012	1.679	0.121	26.37	0.0528
30.0	Small	0.0012	1.773	0.121	20.96	0.0549
	Medium	0.0014	1.663	0.139	21.31	0.0518
	Tall	0.0012	1.687	0.128	26.63	0.0643
37.5	Small	0.0012	1.481	0.155	17.69	0.0622
	Medium	0.0013	1.658	0.125	19.80	0.0596
	Tall	0.0010	1.633	0.121	34.29	0.0563
45.0	Small	0.0013	1.761	0.123	19.10	0.0504
	Medium	0.0016	1.760	0.119	13.09	0.0668
	Tall	0.0010	1.670	0.121	37.46	0.0548
No obstacle	—	0.0009	1.786	0.142	47.04	0.0561

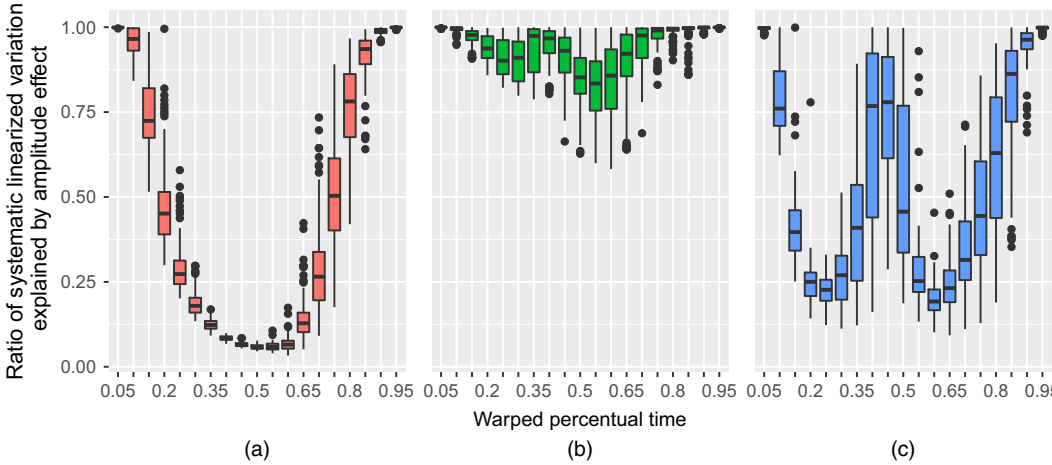

Fig. 15. Co-ordinatewise boxplot of the temporal development of the ratio of S_n to $S_n + Z_n C Z_n^{\mathsf{T}}$ for the 100 samples in the experiment with a small obstacle 30 cm from the starting position: (a) x; (b) y; (c) z

the alignment of the y-co-ordinate seems to contain a relatively larger proportion of amplitude variation after alignment than the x- and z-co-ordinates. We note that the alignment procedure does not change the movement path in (x, y, z) space. The predicted maximum *a posteriori* warping functions are displayed in Fig. 14.

4.3.2. Parameter estimates

The common variance parameter σ and the Matérn parameters α and κ varied little with experiment. In contrast the relative weight a of the stationary covariance and the bridge covariance varied considerably across experiments. However, a was large in all cases, meaning that a large majority of the variance is captured by the stationary part. Table 1 gives all the parameter estimates.

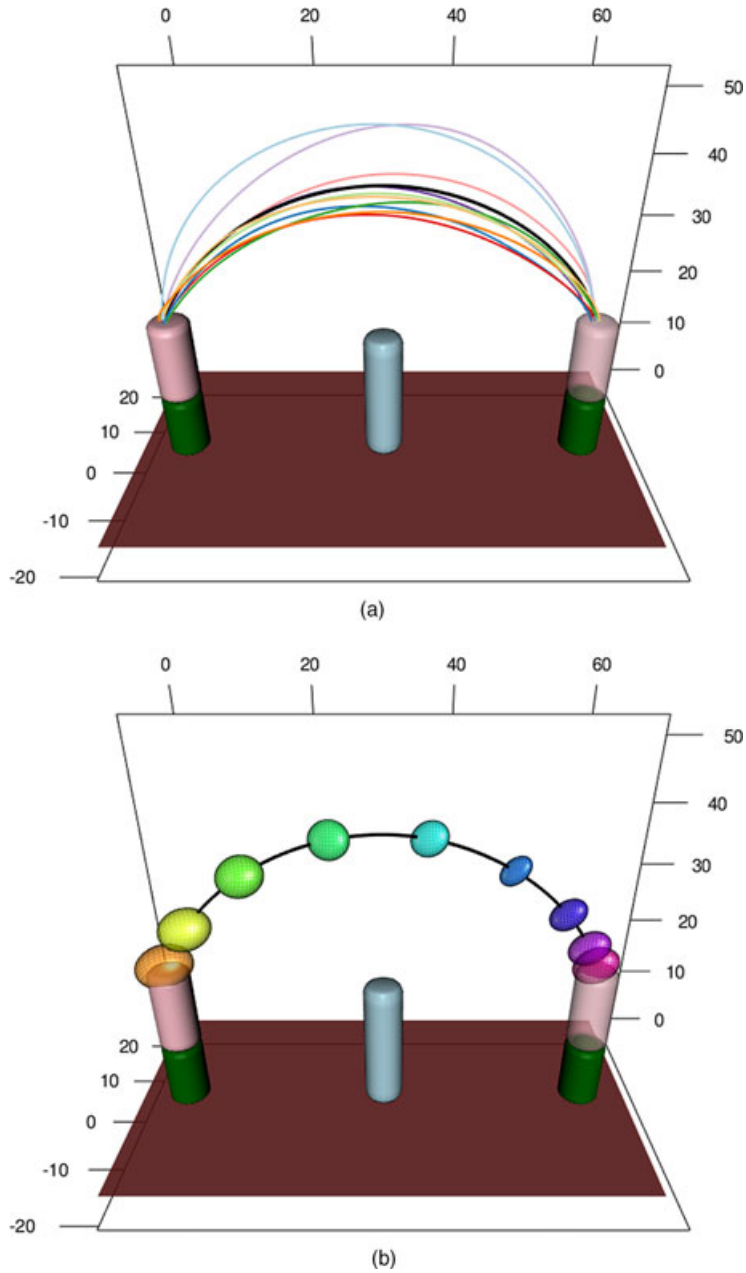

Fig. 16. Estimated experiment-specific curve (———) and participant-specific curves for the experimental set-up with (a) a small obstacle 30 cm from the starting position and (b) estimated 95% predictions ellipsoids for the systematic amplitude effect in the same set-up (the ellipsoids are displayed temporally equidistant around the mean trajectory for the experimental set-up): ———, participant 1; ———, participant 2; ———, participant 3; ———, participant 4; ———, participant 5; ———, participant 6; ———, participant 7; ———, participant 8; ———, participant 9; ———, participant 10

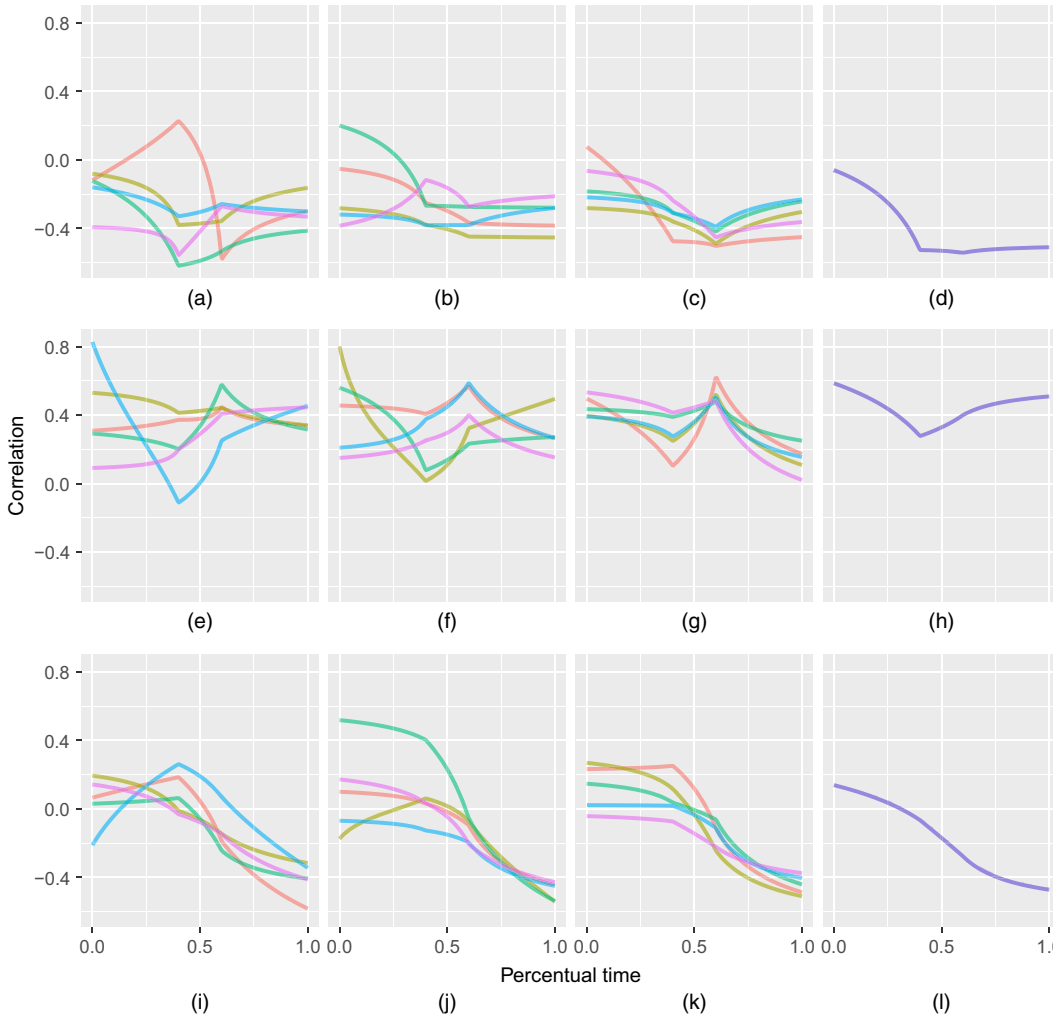

Fig. 17. Correlation functions over time as estimated by the proposed model in all 16 experimental set-ups (——, obstacle 15 cm from start; ——, obstacle 22.5 cm from start; ——, obstacle 30.0 cm from start; ——, obstacle 37.5 cm from start; ——, obstacle 45.0 cm from start): (a) *x*–*y*, 20.0 cm; (b) *x*–*y*, 27.5 cm; (c) *x*–*y*, 35 cm; (d) *x*–*y*, no obstacle; (e) *x*–*z*, 20.0 cm; (f) *x*–*z*, 27.5 cm; (g) *x*–*z*, 35 cm; (h) *x*–*z*, no obstacle; (i) *y*–*z*, 20.0 cm; (j) *y*–*z*, 27.5 cm; (k) *y*–*z*, 35 cm; (l) *y*–*z*, no obstacle

4.3.3. Variance and cross-correlations

The amplitude variation was assumed to be generated from Gaussian processes $\mathbf{x}_n$ and white noise $\varepsilon_n \sim N(0, \sigma^2 \mathbf{I}_{3m_n})$. Since the observed curves are very smooth the estimated contributions from the white noise terms were very small.

Fig. 15 shows the ratios of systematic amplitude variance to linearized systematic variance (amplitude and linearized warp) as estimated by the model. At the end points all variance was captured by the serially correlated amplitude effect. In the *y*-direction almost all variation was captured by the amplitude variance which fits well with the aligned *y*-co-ordinates of the movement path in Fig. 13. The warp-related variance accounted for a larger part of the variation in the *x*- and *z*-directions. The temporal structure of the *x*-co-ordinate reveals that the warp effect explained the majority of the variance around the middle of the movement, whereas for the *z*-

co-ordinate it explained the majority of the variance during lift and descend. Thus, the model predicted warping functions by using a trade-off where the (percentual) temporal midpoints of the transport component and the lift and descend components had highest influence when measuring the alignment of samples.

The individual participant's estimated mean trajectories and the systematic amplitude variation are illustrated in Fig. 16. In Fig. 16(b), the prediction ellipsoids in the middle are relatively small considering that this is the region with most variation. This is because most of the variation was captured by the participant-specific mean curves and the warping effect, as we would expect. The amplitude variance around the end points seems somewhat overestimated, which suggests that the chosen anchor points provided too coarse a model for the dynamics of the true covariance function around the end points.

Of particular interest is the correlation for the three axes (i.e. x–y, x–z and y–z) and how it varies over time as seen in Fig. 17. From the results, it is clear that the correlations vary over time, which Fig. 16 also illustrates. The variation of correlation with respect to time is moderate for the x–y- and x–z-correlations, but for the y–z-correlations there is a clear trend for all experimental set-ups that the correlation goes from positive values to negative values. This is a surprising and perhaps unexpected feature since all experimental set-ups are symmetric in the y-co-ordinate. A plausible explanation is that lifting a centrally placed object with the right hand is generally associated with moving that hand to the right (in our set-up, a positive y-value). When the object is raised we observe a positive correlation in the y–z-plane (faster initial movement timing amplifies the effect), and when the object is lowered again we observe corresponding negative correlation.

4.3.4. Classification

To compare different models objectively, we can fit the models to a subset of the samples and compare their fits in terms of their classification accuracies of participant on the remaining data, i.e., for a given functional sample that was not used to fit the model, we wish to determine which of the participants performed the movement. The primary objective of such an exercise is to compare similar generative models, but not so as to obtain the highest possible classification accuracy—a higher score could probably be achievable by standard machine learning methods that would reveal little about the structure of the problem. A similar classification-based approach was used to evaluate the hierarchical 'pavpop' model that was described in Raket *et al.* (2016), which was applied to the one-dimensional acceleration magnitude profiles of the three-dimensional arm movement data set.

The present classification was done in a chronological fivefold cross-validation set-up (the first fold consisted of the two first repetitions for each person, the second fold of the third and fourth and so forth). Different models were fitted on the five training sets, each leaving out one of the folds (the test set). For each test set, the samples were classified by using the model estimates from the corresponding training set. The classification accuracy was then computed as the average classification accuracy across the five folds for each experiment.

In what follows the method proposed is denoted by simultaneous inference for misaligned multivariate curves, SIMM. The following models were used in the comparison.

(a) *Nearest centroid, NC*: the centroids for each person were estimated as the pointwise means in the training set. The classification was done by using the minimal Euclidean distance to the estimated centroid (using linear interpolation).
(b) *Nearest centroid weighted, NC-W*: the centroids were computed similarly to the NC method, but the classification was done by using a distance with weighted co-ordinates, the weights for the x-, y- and z-co-ordinates were 0.1, 0.7 and 0.2.

(c) *Fisher–Rao L^2, FR-L^2*: pointwise template functions were estimated by using groupwise elastic function alignment and principal component analysis extraction for modelling amplitude variation (Tucker *et al.*, 2013; Tucker, 2017). The standard setting of using three principal components was used. The elastic curve approach for functional data is widely considered the state of the art framework for handling misaligned functional data (Marron *et al.*, 2015). The template functions were estimated separately for each of the three value co-ordinates of the trajectories. Classification was done by using minimal Euclidean distance to the estimated template functions.

(d) *Fisher–Rao elastic, FR_{E}*: template functions were estimated similarly to method FR-L^2, but classification was done by using an elastic distance that both measures co-ordinatewise distances as a sum of phase (Tucker *et al.* (2013), section 3.1) and amplitude directions (Tucker *et al.* (2013), definition 1). The weighting between phase and amplitude distances was 0.16/0.84.

(e) *Fisher–Rao elastic weighted, FR_{E}-W*: template functions and classification were done similarly to method FR_{E}, except that we include a weighting of the three elastic distances corresponding to each value co-ordinate. The weighting between phase and amplitude distances was 0.14/0.86 and the weights for the x-, y- and z-components of the elastic distance were 0.3, 0.2 and 0.5.

(f) *Elastic curve metric, ECM*: the multivariate elastic distance between curves is defined as the geodesic distance on $L^2([0,1];\mathbb{R}^3)/\Gamma$, where Γ is the closure of the set of positive diffeomorphisms on $[0,1]$. In the quotient space $L^2([0,1];\mathbb{R}^3)/\Gamma$, all temporal features are removed and a comparison of curves is done by using only their image in $\mathbb{R}^3$, but in a way that is consistent with reparameterizations of the original curves (Srivastava and Klassen, 2016). Templates were estimated as the pointwise averages of samples aligned to the Karcher mean in $L^2([0,1];\mathbb{R}^3)/\Gamma$ computed by using the `fdasrvf` R package (Tucker, 2017). Classification was done by using a weighted sum of multivariate elastic distance and phase distance (defined as for method FR_{E}). The weighting between the elastic and phase distances was 0.24/0.76.

(g) *SIMM*: the person-specific templates are estimated by using the proposed model with a diagonal cross-covariance structure (i.e. no cross-covariance). Classification is done by using the nearest posterior distance under the maximum likelihood estimates as a function of the unknown sample.

(h) *SIMM-CC*: estimation and classification are done similarly to method SIMM, but using the full dynamic cross-covariance structure that was described in the previous sections.

All weights described in the above methods were chosen by cross-validation on the accuracies for the three experimental set-ups with $d = 30.0$ cm. The grids that were used for determining the parameters are given in Appendix A.

The classification accuracies are available in Table 2. If we first consider the NC-type methods that do not model any warping effect, we see a marked increase in accuracy when weighting the different co-ordinates in the classification, and thus emulating a constant diagonal cross-covariance structure. If we consider the basic elastic model FR-L^2 based on the Fisher–Rao metric, we see similar results to those of the simple NC model, even though the FR-L^2 method also accounts for a warping effect when estimating the template. When classifying by using an elastic distance, as was done in FR_{E}, we see a great increase in classification accuracy. The phase distance contributes considerably to these improvements. When considering only elastic amplitude distance (i.e. weighting phase/amplitude distances 0/1) the average classification accuracy is 0.576. Taking the deformation distance into account in the classification, and thus paying a price

Table 2. Classification accuracies of various methods†

d (cm)	*Obstacle*	*Results for the following models:*							
		NC	*NC-W*	*FR-L²*	*FR*$_E$	*FR*$_E$*-W*	*ECM*	*SIMM*	*SIMM-CC*
15.0	Small	0.62	0.71	0.58	0.77	0.79	0.77	0.80	**0.85**
	Medium	0.60	0.63	0.62	0.64	0.68	0.77	0.80	**0.83**
	Tall	0.52	0.57	0.54	0.58	0.58	0.77	**0.84**	0.81
22.5	Small	0.51	0.58	0.50	0.68	0.66	**0.77**	0.69	**0.77**
	Medium	0.52	0.64	0.56	0.62	0.73	0.70	**0.75**	0.72
	Tall	0.50	0.62	0.49	0.64	0.73	0.73	0.74	**0.79**
30.0	Small	0.53	*0.59*	0.53	*0.69*	*0.72*	***0.76***	0.70	**0.76**
	Medium	0.45	*0.47*	0.48	*0.65*	*0.68*	*0.70*	**0.79**	0.75
	Tall	0.58	*0.63*	0.56	*0.65*	*0.73*	*0.78*	**0.86**	0.83
37.5	Small	0.51	0.55	0.52	0.67	0.72	0.70	**0.77**	0.76
	Medium	0.45	0.50	0.43	0.68	0.65	0.69	**0.68**	**0.68**
	Tall	0.50	0.53	0.54	0.67	0.73	0.72	**0.80**	**0.80**
45.0	Small	0.49	0.54	0.51	0.66	0.71	0.75	0.69	**0.76**
	Medium	0.48	0.53	0.44	0.66	0.70	0.71	**0.78**	0.73
	Tall	0.50	0.54	0.50	0.71	0.75	0.74	0.82	**0.83**
No obstacle	—	0.48	0.56	0.52	0.68	0.72	**0.80**	0.64	0.70
Average		0.515	0.574	0.520	0.666	0.705	0.741	0.761	**0.773**

†Bold indicates the best result(s); italics indicates that the given experiments were used for training.

for warping the templates, we see a great increase in classification accuracy. The heuristic idea of having to pay a price for large warps in many ways emulates the proposed idea of modelling the warping functions as random effects. Finally, method FR$_E$-W includes a weighting of the combined phase and amplitude distances across the x-, y- and z-co-ordinates of the observed trajectories, which again increases the accuracy.

The elastic metric has many similarities with the Fisher–Rao metric but is multivariate in nature. The ECM method has higher accuracies than the similar FR$_E$ and FR$_E$-W methods. An exploratory comparison of results suggested that this was caused by more appropriate warping across all co-ordinates, leading to both better estimates of templates and in turn more accurate phase distances.

The SIMM model is the proposed model described above, but without a dynamic cross-correlation structure. Instead we have three scale parameters that describe the weighting of the marginal variances in the three value co-ordinates. The model is thus both comparable with FR$_E$-W and ECM, both of which are outperformed in terms of accuracy. It is important to note that, whereas FR$_E$-W and ECM required cross-validation on a subset of the test data to estimate the parameters, the SIMM model estimates all variance parameters that are used in the weighting of the different aspects of the movement from the training data. The final model, SIMM-CC, includes a full dynamic cross-covariance structure. Even though we might expect that this model was much more prone to overfitting to the training data (the model includes 27 free amplitude variance parameters compared with the six parameters of model SIMM), we see a slight increase in accuracy of the method. We remark that the ECM, SIMM and SIMM-CC methods, which make a joint warp of the three spatial co-ordinates, had the best accuracies among the methods in consideration. This strongly supports the idea of modelling multivariate signals with a joint warping of all value co-ordinates.

5. Discussion

In this paper we have proposed a new class of models for simultaneous inference for misaligned multivariate functional data. We fitted these types of model to three different data sets and applied them in one classification scenario.

The idea behind the approach is simultaneously to model the predominant effects in functional data sets, misalignment and amplitude variation, as random effects. The simultaneous modelling allows separation of these effects in a data-driven manner, namely by maximum likelihood estimation. In particular, we saw that this separation resulted in nicely behaving warping functions that did not seem to overalign the functional samples.

The models enable estimation of dynamic correlation functions between the individual co-ordinates of the amplitude variation. We demonstrated that we can achieve superior fits and better classification by using the parametric construction from proposition 1, even when the number of free parameters is high relative to the number of functional samples. By fitting the model to two large functional data sets related to human movement, we also demonstrated the computational feasibility of maximum likelihood inference with such models.

The parametric model class for dynamic covariance structures proposed is very general, but other modelling approaches could be better suited in some situations. For example, instead of using a fixed number of parameters to describe each marginal variance and cross-covariance function, one would often prefer to do this in a data-driven manner. One possibility could be to model the multivariate amplitude covariance function by using a multivariate functional factor analysis model, e.g. a multivariate extension of the rank-reduced model of James *et al.* (2000), where the number of parameters describing the covariance is fixed, and the covariance is described in terms of functional principal components. However, such amplitude effects cannot be effectively fitted by using conventional optimizers for the likelihood and would require the development of specialized efficient fitting methods (e.g. generalizing the methods of Peng and Paul (2009)). Another relevant approach would be simultaneous warping of fixed effects and amplitude variation, and one could also consider extending the domain of feasible warping functions by modelling the latent warp variables w as more general functional objects (e.g. stochastic processes) instead of elements belonging to $\mathbb{R}^{m_{\mathbf{w}}}$ for some $m_{\mathbf{w}}$. We shall leave these extensions as future work.

Acknowledgements

We thank the referees and the Associate Editor for their careful reading and suggested improvements.

Appendix A: Cross-validation grids

The cross-validation that was used to determine the parameters of the methods NC-W, $\mathrm{FR_E}$, $\mathrm{FR_E}$-W and ECM in Section 4.3 were given as follows. The possible weights between the three value co-ordinates were $\{\mathbf{w} \in \mathbb{R}^3 : w_i \in \{0, 0.1, \ldots, 1\}, w_1 + w_2 + w_3 = 1\}$ and the possible weights between amplitude and phase distance were $\{\mathbf{w} \in \mathbb{R}^2 : w_i \in \{0, 0.02, \ldots, 1\}, w_1 + w_2 = 1\}$. Method NC-W uses only weighting between value co-ordinates and $\mathrm{FR_E}$ and ECM use only weighting between the amplitude and phase distance.

For model SIMM-CC we explored adding more than three knots to the warp model ($m_{\mathbf{w}} = 3, 4, 5$), but $m_{\mathbf{w}} = 3$ gave the best cross-validation score.

Appendix B: Covariance functions

Below we list the covariance functions that are used in the three data examples.

Schur's theorem states that the pointwise product of covariance functions yields a valid covariance function (Schur, 1911). This property is used in the arm movement example.

(a) *Brownian bridge*: the covariance function for the Brownian bridge defined on the temporal domain $[0, 1]$ is given by

$$f_{\text{bridge}}(s,t) = \tau^2 \min(s,t)\{1 - \max(s,t)\} = \tau^2\{\min(s,t) - st\}, \qquad s,t \in [0,1], \tag{13}$$

where $\tau > 0$ is a scale parameter.

(b) *Brownian motion*: the covariance function for the Brownian motion defined on the domain $[0, \infty)$ is given by

$$f_{\text{motion}}(s,t) = \tau^2 \min(s,t), \qquad s,t \geqslant 0, \tag{14}$$

where $\tau > 0$ is a scale parameter.

(c) *Mixing stationary and bridge covariances*: the combination of a stationary and bridge covariance with mixtures a and b is given by

$$f_{\text{mixture}(a,b)}(s,t) = a + b\{\min(s,t) - st\}.$$

In our analysis the parameter b is redundant, so we use

$$f_{\text{mixture}(a)}(s,t) = a + \min(s,t) - st. \tag{15}$$

Note that the bridge covariance is not the same construction as when conditioning a stochastic process X on its end point value.

(d) *Matérn covariance function*: the covariance function for the Matérn covariance with smoothness parameter α and range parameter κ is given by

$$f_{\text{Matérn}(\alpha,\kappa)}(s,t) = \frac{2^{1-\alpha}}{\Gamma(\alpha)} \left(\frac{|s-t|}{\kappa} \right)^{\alpha} K_\alpha \left(\frac{|s-t|}{\kappa} \right), \qquad s,t \in \mathbb{R}. \tag{16}$$

Here K_α is the modified Bessel function of the second kind. A Gaussian process with Matérn covariance is stationary, and conversely any stationary continuous Gaussian process with mean 0 has a covariance function that up to scale is given by a Matérn covariance function (Rasmussen and Williams, 2006).

Appendix C: Expectation–maximization algorithm for the spline coefficients in the linearized model

First note that by assumption the mean curves $\boldsymbol{\theta}$ are the same, expect for warping, for trajectories belonging to the same subject groups and are independent of other subject groups. Thus, to simplify the notation and to ease argumentation, we shall assume that all trajectories belong to the same subject group.

Let $f = \{f_k\}$ be the spline base function for $\boldsymbol{\theta}$ and let $\mathbf{c}$ be the spline coefficients, i.e. $\boldsymbol{\theta}(t) = f(t)\mathbf{c}$. Consider the linearized model from equation (10):

$$\vec{\mathbf{y}}_n \approx \vec{\boldsymbol{\gamma}}_{\mathbf{w}_n^0} + Z_n(\mathbf{w}_n - \mathbf{w}_n^0) + \vec{\mathbf{x}}_n + \vec{\varepsilon}_n, \qquad n = 1, \ldots, N,$$

with log-likelihood

$$\sum_{n=1}^{N} [q m_n \log(\sigma^2) + \log\{\det(V_n)\} + \sigma^{-2}(\vec{\mathbf{y}}_n - \vec{\boldsymbol{\gamma}}_{\mathbf{w}_n^0} + Z_n \mathbf{w}_n^0)^{\mathrm{T}} V_n^{-1} (\vec{\mathbf{y}}_n - \vec{\boldsymbol{\gamma}}_{\mathbf{w}_n^0} + Z_n \mathbf{w}_n^0)].$$

For the remainder we assume that $\mathbf{w}_n^0 = \{\mathbf{w}_{nl}^0\}_{l=1}^{m_w}$ and all variance parameters $(\mathcal{S}, C, \sigma^2)$ are fixed, and that we have a current estimate of the spline coefficients $\mathbf{c}_0$. The conditional expectation and variance of $\mathbf{w}_n$ given the observations $\mathbf{y}$ under the current parameters will be denoted by $\bar{\mathbf{w}}_n = \{\bar{\mathbf{w}}_{nl}\}_{l=1}^{m_w} \in \mathbb{R}^{m_w}$ and $\bar{\bar{\mathbf{w}}}_n = \{\bar{\bar{\mathbf{w}}}_{nl_1 l_2}\}_{l_1, l_2 = 1}^{m_w} \in \mathbb{R}^{m_w \times m_w}$ respectively. Using this notation the conditional log-likelihood of $\vec{\mathbf{y}}_n$ given $\mathbf{w}_n$ is

$$l_{\vec{\mathbf{y}}_n | \mathbf{w}_n} = (\vec{\mathbf{y}}_n - \vec{\boldsymbol{\gamma}}_{\mathbf{w}_n^0} - Z_n(\mathbf{w}_n - \mathbf{w}_n^0))^{\mathrm{T}} S_n^{-1} (\vec{\mathbf{y}}_n - \vec{\boldsymbol{\gamma}}_{\mathbf{w}_n^0} - Z_n(\mathbf{w}_n - \mathbf{w}_n^0)) + \log\{\det(S_n)\}.$$

The term $\log\{\det(S_n)\}$ does not influence the estimation of $\mathbf{c}$, and hence it will be removed in what follows. The conditional expectation $E[l_{\vec{\mathbf{y}}_n|\mathbf{w}_n}|\vec{\mathbf{y}}_n]$ given the observation hence equals

$$(\vec{\mathbf{y}}_n - \vec{\gamma}_{\mathbf{w}_n^0} - Z(\bar{\mathbf{w}}_n - \mathbf{w}_n^0))^{\mathrm{T}} S_n^{-1} (\vec{\mathbf{y}}_n - \vec{\gamma}_{\mathbf{w}_n^0} - Z(\bar{\mathbf{w}}_n - \mathbf{w}_n^0)) + \mathrm{tr}(S_n^{-1} Z_n \bar{\bar{\mathbf{w}}}_n Z_n^{\mathrm{T}}). \tag{17}$$

Defining $R_n = f\{v(t_k, \mathbf{w}_n^0)\}$ and $R_{nl} = \partial_t f\{v(t_k, \mathbf{w}_n^0)\}\partial_{\mathbf{w}_l} v(t_k, \mathbf{w}_n^0)$ for $l = 1, \ldots, m_w$ we have that $Z_n = \{R_{nl}\mathbf{c}\}_{l=1}^{m_w}$ and thus $Z_n \mathbf{w}_n = (\Sigma_{l=1}^{m_w} \mathbf{w}_{nl} R_{nl})\mathbf{c}$. Using this the trace from expression (17) can be expanded as a double sum:

$$\mathrm{tr}(S_n^{-1} Z \bar{\bar{\mathbf{w}}}_n Z^{\mathrm{T}}) = \sum_{l_1, l_2 = 1}^{m_w} \bar{\bar{\mathbf{w}}}_{nl_1 l_2} \, \mathrm{tr}(S_n^{-1} R_{nl_1} \mathbf{c}\mathbf{c}^{\mathrm{T}} R_{nl_2}^{\mathrm{T}}).$$

Calculating the gradient of expression (17) now gives that $\nabla_{\mathbf{c}} E[l_{\vec{\mathbf{y}}_n|\mathbf{w}_n}|\vec{\mathbf{y}}_n]$ is proportional to

$$-K_n^{\mathrm{T}} S_n^{-1} (\vec{\mathbf{y}}_n - K_n \mathbf{c}) + \sum_{l_1, l_2 = 1}^{m_w} \bar{\bar{\mathbf{w}}}_{nl_1 l_2} R_{nl_2}^{\mathrm{T}} S_n^{-1} R_{nl_1} \mathbf{c},$$

where $K_n = R_n + \Sigma_{l=1}^{m_w} (\bar{\mathbf{w}}_{nl} - \mathbf{w}_{nl}^0) R_{nl}$. From this it follows that the maximization step of the EM algorithm for the spline coefficients $\mathbf{c}$ is given by

$$\mathbf{c}_{\mathrm{new}} = \left(\sum_{n=1}^{N} K_n^{\mathrm{T}} S_n^{-1} K_n + \sum_{l_1, l_2 = 1}^{m_w} \bar{\bar{\mathbf{w}}}_{nl_1 l_2} R_{nl_1} S_n^{-1} R_{nl_2}^{\mathrm{T}} \right)^{-1} \sum_{n=1}^{N} K_n^{\mathrm{T}} S_n^{-1} y_n.$$

References

Aksglaede, L., Sørensen, K., Petersen, J. H., Skakkebæk, N. E. and Juul, A. (2009) Recent decline in age at breast development: the Copenhagen Puberty Study. *Pediatrics*, **123**, no. 5, e932–e939.

Beath, K. J. (2007) Infant growth modelling using a shape invariant model with random effects. *Statist. Med.*, **26**, 2547–2564.

Cole, T. J., Donaldson, M. D. and Ben-Shlomo, Y. (2010) SITAR—a useful instrument for growth curve analysis. *Int. J. Epidem.*, **39**, 1558–1566.

Dryden, I. L. and Mardia, K. V. (1998) *Statistical Shape Analysis*, vol. 4. Chichester: Wiley.

Gervini, D. and Gasser, T. (2005) Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. *Biometrika*, **92**, 801–820.

Grimme, B. (2014) Analysis and identification of elementary invariants as building blocks of human arm movements. *PhD Thesis* (in German). International Graduate School of Biosciences, Ruhr-Universität Bochum, Bochum.

Grimme, B., Lipinski, J. and Schöner, G. (2012) Naturalistic arm movements during obstacle avoidance in 3D and the identification of movement primitives. *Exptl Brain Res.*, **222**, no. 3, 185–200.

Guo, W. (2002) Functional mixed effects models. *Biometrics*, **58**, 121–128.

Hadjipantelis, P. Z., Aston, J. A., Müller, H.-G. and Evans, J. P. (2015) Unifying amplitude and phase analysis: a compositional data approach to functional multivariate mixed-effects modeling of Mandarin Chinese. *J. Am. Statist. Ass.*, **110**, 545–559.

Hadjipantelis, P. Z., Aston, J. A., Müller, H.-G. and Moriarty, J. (2014) Analysis of spike train data: a multivariate mixed effects model for phase and amplitude. *Electron. J. Statist.*, **8**, 1797–1807.

Hyman, J. M. (1983) Accurate monotonicity preserving cubic interpolation. *SIAM J. Scient. Statist. Comput.*, **4**, 645–654.

James, G. M., Hastie, T. J. and Sugar, C. A. (2000) Principal component models for sparse functional data. *Biometrika*, **87**, 587–602.

Kendall, D. G. (1989) A survey of the statistical theory of shape. *Statist. Sci.*, **4**, 87–99.

Kneip, A. and Ramsay, J. O. (2008) Combining registration and fitting for functional models. *J. Am. Statist. Ass.*, **103**, 1155–1165.

Kurtek, S., Srivastava, A., Klassen, E. and Ding, Z. (2012) Statistical modeling of curves using shapes and related features. *J. Am. Statist. Ass.*, **107**, 1152–1165.

Lindstrom, M. J. and Bates, D. M. (1990) Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**, 673–687.

Manay, S., Cremers, D., Hong, B.-W., Yezzi, A. J. and Soatto, S. (2006) Integral invariants for shape matching. *IEEE Trans. Pattn Anal. Mach. Intell.*, **28**, 1602–1618.

Marron, J., Ramsay, J. O., Sangalli, L. M. and Srivastava, A. (2015) Functional data analysis of amplitude and phase variation. *Statist. Sci.*, **30**, 468–484.

Peng, J. and Paul, D. (2009) A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *J. Computnl Graph. Statist.*, **18**, 995–1015.

Raket, L. L., Grimme, B., Schöner, G., Igel, C. and Markussen, B. (2016) Separating timing, movement conditions and individual differences in the analysis of human movement. *PLOS Computnl Biol.*, **12**, no. 9, article e1005092.
Raket, L. L., Sommer, S. and Markussen, B. (2014) A nonlinear mixed-effects model for simultaneous smoothing and registration of functional data. *Pattn Recogn Lett.*, **38**, 1–7.
Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*, 2nd edn. New York: Springer.
Rasmussen, C. E. and Williams, C. K. I. (2006) *Gaussian Processes for Machine Learning*. Cambridge: MIT Press.
Rønn, B. B. (2001) Nonparametric maximum likelihood estimation for shifted curves. *J. R. Statist. Soc.* B, **63**, 243–259.
Rønn, B. B. and Skovgaard, I. M. (2009) Nonparametric maximum likelihood estimation of randomly time-transformed curves. *Braz. J. Probab. Statist.*, **23**, 1–17.
Schur, J. (1911) Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *J. Reine Angew. Math.*, **140**, 1–28.
Sebastian, T. B., Klein, P. N. and Kimia, B. B. (2003) On aligning curves. *IEEE Trans. Pattn Anal. Mach. Intell.*, **25**, 116–125.
Sørensen, K., Aksglaede, L., Petersen, J. H. and Juul, A. (2010) Recent changes in pubertal timing in healthy Danish boys: associations with body mass index. *J. Clin. Endcrin. Metablsm*, **95**, 263–270.
Srivastava, A. and Klassen, E. P. (2016) *Functional and Shape Data Analysis*. Berlin: Springer.
Srivastava, A., Klassen, E., Joshi, S. H. and Jermyn, I. H. (2011) Shape analysis of elastic curves in Euclidean spaces. *IEEE Trans. Pattn Anal. Mach. Intell.*, **33**, 1415–1428.
Tinggaard, J., Aksglaede, L., Sørensen, K., Mouritsen, A., Wohlfahrt-Veje, C., Hagen, C. P., Mieritz, M. G., Jørgensen, N., Wolthers, O. D., Heuck, C., Petersen, J. H., Main, K. M. and Juul, A. (2014) The 2014 Danish references from birth to 20 years for height, weight and body mass index. *Acta Paed.*, **103**, 214–224.
Tucker, J. D. (2017) fdasrvf: elastic functional data analysis. *R Package Version 1.8.3*. Sandia National Laboratories, Albuquerque. (Available from `https://github.com/jdtuck/fdasrvf_R/`.)
Tucker, J. D., Wu, W. and Srivastava, A. (2013) Generative models for functional data using phase and amplitude separation. *Computnl Statist. Data Anal.*, **61**, 50–66.
Vantini, S. (2012) On the definition of phase and amplitude variability in functional data analysis. *Test*, **21**, 676–696.
Wang, J.-L., Chiou, J.-M. and Mueller, H.-G. (2015) Review of functional data analysis. *Preprint arXiv:1507.05135*. University of California, Davis.
Wolfinger, R. (1993) Laplace's approximation for nonlinear mixed models. *Biometrika*, **80**, 791–795.
Younes, L. (1998) Computable elastic distances between shapes. *SIAM J. Appl. Math.*, **58**, 565–586.