Biometrics WILEY

Elastic analysis of irregularly or sparsely sampled curves

Lisa Steyer ⬤ | **Almond Stöcker** ⬤ | **Sonja Greven** ⬤

School of Business and Economics, Chair of Statistics, Humboldt-Universität zu Berlin, Berlin, Germany

Correspondence
Lisa Steyer, Humboldt-Universität zu Berlin, School of Business and Economics, Chair of Statistics, Unter den Linden 6, 10099 Berlin, Germany.
Email: lisa.steyer@hu-berlin.de

Abstract

We provide statistical analysis methods for samples of curves in two or more dimensions, where the image, but not the parameterization of the curves, is of interest and suitable alignment/registration is thus necessary. Examples are handwritten letters, movement paths, or object outlines. We focus in particular on the computation of (smooth) means and distances, allowing, for example, classification or clustering. Existing parameterization invariant analysis methods based on the elastic distance of the curves modulo parameterization, using the square-root-velocity framework, have limitations in common realistic settings where curves are irregularly and potentially sparsely observed. We propose using spline curves to model smooth or polygonal (Fréchet) means of open or closed curves with respect to the elastic distance and show identifiability of the spline model modulo parameterization. We further provide methods and algorithms to approximate the elastic distance for irregularly or sparsely observed curves, via interpreting them as polygons. We illustrate the usefulness of our methods on two datasets. The first application classifies irregularly sampled spirals drawn by Parkinson's patients and healthy controls, based on the elastic distance to a mean spiral curve computed using our approach. The second application clusters sparsely sampled GPS tracks based on the elastic distance and computes smooth cluster means to find new paths on the Tempelhof field in Berlin. All methods are implemented in the R-package "elasdics" and evaluated in simulations.

KEYWORDS
curve alignment, Fisher–Rao Riemannian metric, functional data analysis, multivariate functional data, registration, square-root-velocity transformation, warping

1 | INTRODUCTION

In the biomedical sciences, data are increasingly collected that take the form of open or closed curves $\beta : [0,1] \rightarrow \mathbb{R}^d$, $d \in \mathbb{N}$. Examples for such curves in two or three dimensions are (human) movement patterns (e.g., Backenroth et al., 2018), handwritten letters or symbols (e.g., Dryden and Mardia, 2016; Isenkul et al., 2014), protein structures (Srivastava et al., 2010), or the outline of an (e.g., anatomic) object, such as the corpus callosum (Joshi et al., 2013). The two applications we consider in this paper concern a spiral drawing test for the detection of Parkinson's disease, and GPS-recorded movement tracks. In most of the named cases, only the image of the curve represents the object of interest. An "elastic" analysis is then required, that is, a statistical analysis of the curves' image in $\mathbb{R}^d$ that does not take their parameterization over $[0,1]$ into account and is invariant under different parameterizations. Ideally, it should also yield an optimal alignment of different curves to allow point-to-point comparison, as illustrated in the example in Figure 1. As in this example, curves are often observed at a differing number of discrete points. The aim
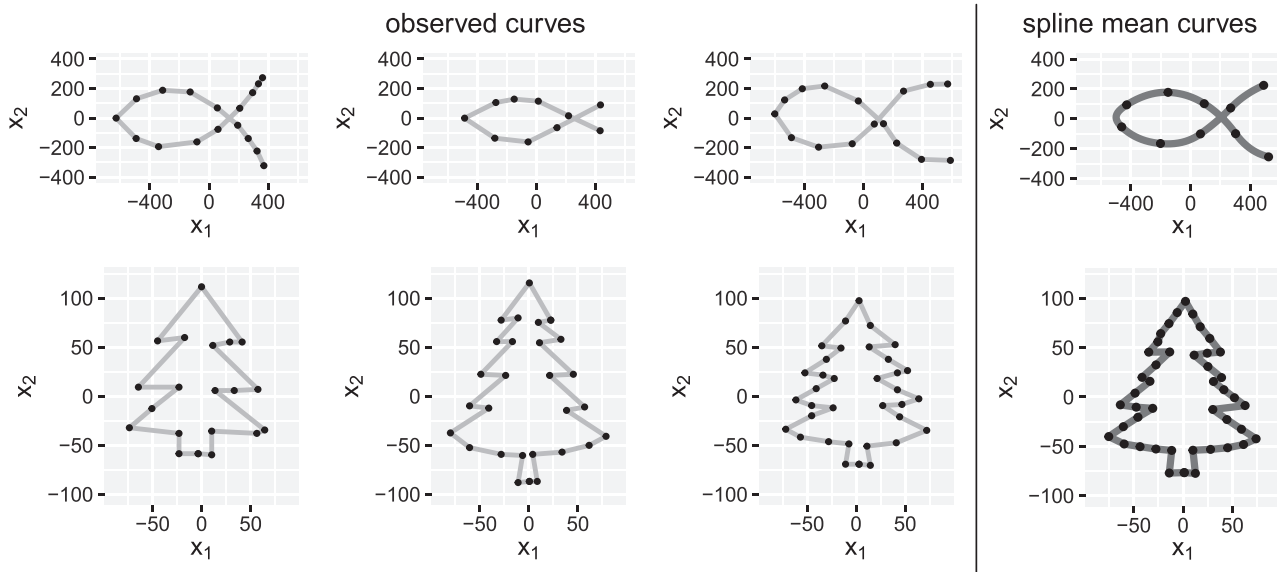
FIGURE 1 Two toy examples of sparsely and irregularly observed curves in $\mathbb{R}^2$ with observed points indicated as black dots and linear interpolation (first three columns). Ideally, the analysis should yield an optimal alignment of different curves to allow comparison of corresponding points such as bumps and other features (the mouth of the fish/the branches of the trees). Smooth or polygonal spline means (last column in dark gray) are computed using our methods, with black dots indicating values at the model-based spline knots

of this paper is to extend elastic statistical methodology to such realistic cases where curves are irregularly and sparsely sampled. In particular, we develop suitable elastic spline models for (Fréchet) mean curves of samples of such curves, and show that certain first- and second-order splines meet the identifiability properties required in a modulo parameterization context. These means can be smooth curves, such as shown for the fish in Figure 1, or polygonal curves, better suited for curves with sharp corners like the trees in Figure 1. To this end, we also propose suitable algorithms for alignment and distance computation of irregularly or sparsely sampled curves—necessary for mean computation, but also useful for distance-based analyses such as clustering or classification. In particular, we derive a useful simplification of the warping (reparameterization, alignment) problem when interpreting the observed curves as polygons.

The alignment problem for curves in $\mathbb{R}^d$ is closely related to the registration problem in functional data analysis (Ramsay and Silverman, 2005), which corresponds to the case $d = 1$. For two functions f_1 and f_2, warping has commonly been treated as an optimization problem $\inf_{\gamma \in \Gamma} \|f_1 - f_2 \circ \gamma\|_{L_2}$ on a suitable function space Γ of warping functions γ. This choice is problematic as $\inf_{\gamma \in \Gamma} \|f_1 - f_2 \circ \gamma\|_{L_2}$ does not define a proper distance on the space of curves modulo parameterization. The mapping is not symmetric and can be zero even if f_2 is not a warped version of f_1, which is related to the so-called

"pinching" problem (Marron et al., 2015). Intuitively, this "pushes" the integration mass to parts of the domain where f_1 and f_2 are close. To avoid this "pinching" effect, a regularization term can be added to the loss function (Ramsay and Silverman, 2005). This is done in various dynamic time warping algorithms, where usually large values of the derivative of the warping function are penalized (Sakoe and Chiba, 1978). Alternatively, one can choose a small number of basis functions for the warping or combine both approaches to use penalized basis functions (Ramsay and Li, 1998). Moreover, Bayesian approaches to modeling warping functions have been suggested (e.g., Lu et al., 2017, or Matuk et al., 2021 for sparse one-dimensional functions).

All of these approaches restrict the amount of warping; thus, the analysis is not completely independent of the observed parameterization. This seems more suitable for one-dimensional functions ($d = 1$) where one seeks to separate phase (parameterization) and amplitude (image) but considers both as informative. If we analyze curves in $\mathbb{R}^d$, $d > 1$, however, we are usually only interested in the image representing the curve, that is, the equivalence class of the curve with respect to (w.r.t.) parameterization, which makes penalized, restricted, or Bayesian approaches for the warping less suitable.

Srivastava et al. (2010) propose a proper metric on the resulting quotient space via minimizing the distance between the square-root-velocity (SRV) transformed

curves. For more details on this framework, see Srivastava and Klassen (2016) and Subsection 2.1. Their perspective is focused on the curves as functions (rather than discrete observations) that, in practice, requires interpolating the curves on a regular grid for the mean computation. This works well in the case of densely observed curves. Often, however, for example, in our applications, curves are only observed at a relatively small number of discrete points, where the number differs between curves (sparse and irregular setting). We show in examples that (elastic) methods designed for densely observed curves have limitations for such sparse settings. This problem is well known in functional data analysis ($d = 1$), where spline representations or other smoothing methods are frequently used to model sparsely and/or irregularly observed functions (e.g., Greven and Scheipl, 2017; Yao et al., 2005).

The main contributions of this paper thus are to carefully introduce spline functions to model elastic (Fréchet) mean curves in $\mathbb{R}^d$ on SRV or curve level, to show that the proposed model is identifiable via its spline coefficients modulo parameterization, and to discuss limitations of this identifiability. This extends approaches for functional data to curves in $\mathbb{R}^d$, $d \geq 2$ and to the elastic setting.

As part of the mean estimation, but also of interest in its own right, we also develop algorithms to align open and closed curves if at least one of them is piecewise linear, for instance, a sparsely observed curve treated as a polygon, and show local maximization properties of our algorithm for open curves. We show the usefulness of our methods for statistical analysis of irregularly or sparsely observed curves in two applications to a Parkinson spiral drawing test and to GPS movement tracks, involving mean computation, clustering, and classification of curves. Proofs of all formal statements are provided in Web Appendix B.

2 | ELASTIC ANALYSIS OF OBSERVED CURVES

In Section 2.1, we briefly review the SRV framework for analyzing curves modulo parameterization. Then, in Sections 2.2 and 2.3, we introduce our methods for elastic distance computation for irregularly or sparsely sampled curves, a building block for the spline-based Fréchet mean that we propose, and additionally of interest for distance-based analysis methods such as clustering or classification. In Sections 2.4 and 2.5, we introduce spline functions to model smooth or polygonal elastic mean curves and discuss identifiability of these modulo parameterization in Section 2.6. For all proposed methods, we focus on open curves for better readability and present adapted versions for closed curves in Web Appendix A.

2.1 | Square-root-velocity framework

Srivastava et al. (2010) show that for two absolutely continuous curves $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, the Fisher–Rao metric can be simplified to the L_2-distance between the corresponding SRV-curves, which can be minimized over the warping to obtain an elastic distance between the two curves.

Definition 1 Elastic distance; Srivastava et al., 2010. Let $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 : [0, 1] \to \mathbb{R}^d$ be absolutely continuous and $[\boldsymbol{\beta}_1]$ and $[\boldsymbol{\beta}_2]$ their respective equivalence classes modulo parameterization and translation. Then the elastic distance between $[\boldsymbol{\beta}_1]$ and $[\boldsymbol{\beta}_2]$ is

$$d([\boldsymbol{\beta}_1], [\boldsymbol{\beta}_2]) = \inf_{\gamma_1, \gamma_2 \in \Gamma} \|(\mathbf{q}_1 \circ \gamma_1) \cdot \sqrt{\dot{\gamma}_1} - (\mathbf{q}_2 \circ \gamma_2) \cdot \sqrt{\dot{\gamma}_2}\|_{L_2}, \quad (1)$$

with Γ being the set of boundary-preserving diffeomorphisms $\gamma : [0, 1] \to [0, 1]$, $\|\mathbf{q}\|_{L_2}^2 = \int_0^1 \|\mathbf{q}(t)\|^2 dt$ and SRV transformations $\mathbf{q}_1$ and $\mathbf{q}_2$ of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ defined via

$$\mathbf{q}_i(t) = \begin{cases} \dfrac{\dot{\boldsymbol{\beta}}_i(t)}{\sqrt{\|\dot{\boldsymbol{\beta}}_i(t)\|}} & \text{if } \dot{\boldsymbol{\beta}}_i(t) \neq 0 \\ 0 & \text{if } \dot{\boldsymbol{\beta}}_i(t) = 0 \end{cases} \quad \text{for } i = 1, 2.$$

Here, $(\mathbf{q}_i \circ \gamma_i) \cdot \sqrt{\dot{\gamma}_i}$ is the SRV transformation of the reparameterized curve $\boldsymbol{\beta}_i \circ \gamma_i$, $i = 1, 2$.

Srivastava and Klassen (2016) showed that it is sufficient to align one of the curves in (1),

$$d([\boldsymbol{\beta}_1], [\boldsymbol{\beta}_2]) = \inf_{\gamma \in \Gamma} \|\mathbf{q}_1 - (\mathbf{q}_2 \circ \gamma) \cdot \sqrt{\dot{\gamma}}\|_{L_2}. \quad (2)$$

Moreover, they pointed out that to obtain a proper quotient space structure on the space of absolutely continuous curves, we need to consider the closure of SRV-curves w.r.t. parameterization as equivalence classes. That is, for a curve $\boldsymbol{\beta}$ with SRV transformation $\mathbf{q}$, $[\boldsymbol{\beta}]$ consists of all curves whose SRV transformation is in the closure of $\{(\mathbf{q}_i \circ \gamma) \cdot \sqrt{\dot{\gamma}} | \gamma \in \Gamma\}$.

Note that any analysis based on this elastic distance will be modulo translation as a result of taking derivatives. If the position of the curve in space is of interest, it has to be analyzed separately. On the other hand, if curves are used to model shape objects, translation invariance is a desired property. In classic shape data analysis (Dryden and Mardia, 2016), the analysis should additionally be invariant under rotation and scaling, and parameterization invariance presents a further key aspect in functional shape analysis (Srivastava and Klassen, 2016). In this paper, we solely discuss parameterization invariance and

give examples of handwritten spirals and GPS tracks where this elastic analysis is suitable.

A solution to the variational problem in the distance (2) is usually approximated using a dynamic programming algorithm or gradient-based optimization (e.g., in Srivastava et al., 2010). Both approaches discretize the warping space Γ. The dynamic programming algorithm, for instance, assumes a discrete grid for the domain of the warping function. An extension by Bernal et al. (2016) allows for an unequal number of points on both curves and improves computation time. Lahiri et al. (2015) provide an algorithm to align two piecewise linear curves and show that an optimal warping exists if at least one curve is piecewise linear. Such an optimal warping also exists if both curves are continuously differentiable (Bruveris, 2016).

2.2 | Elastic distance for discretely observed curves

In practice, we observe curves in $\mathbb{R}^d$, $d \in \mathbb{N}$, not continuously but only discretely via evaluations of these curves on discrete (and potentially sparse and curve-specific) grids. An elastic analysis needs to explicitly address this point. We propose to treat a discretely observed curve β as a polygon parameterized with constant speed between the observed corners $\beta(s_0), \dots, \beta(s_m)$. This is illustrated in the toy examples (Figure 1) with observed points marked as black dots and the polygon connecting the observations indicated by gray lines. If, as in this example, no parameterization over [0,1] is given for the observed points, we will parameterize the polygon by arc length. Note that we address the case of sparsely observed curves here, whereas the problem of fragmented curves (i.e., curves with unobserved start or end points) generally cannot be handled by the proper distance defined in (1).

If β is such a polygon, the problem of finding an optimal reparameterized curve $\beta \circ \gamma$ to another arbitrary curve can be simplified (similarly as in Lahiri et al., 2015). We show that instead of solving the minimization problem (2) over the space Γ of warping functions, we only need to solve a maximization problem over a subset of $\mathbb{R}^{m-1}$ w.r.t. the new parameterizations $t_1 = \gamma^{-1}(s_1), \dots, t_{m-1} = \gamma^{-1}(s_{m-1})$ at the observed corners.

Lemma 1. *Let β be a polygon in $\mathbb{R}^d$ with constant speed parameterization between its corners $\beta(s_0), \dots, \beta(s_m)$. For its piecewise constant SRV transformation q, denote $q|_{[s_j, s_{j+1}]} = q_j \in \mathbb{R}^d$ for all $j = 0, \dots, m-1$. Let $\tilde{\beta}$ be an absolutely continuous curve with SRV transformation p, $\|p\|_\infty < \infty$. Then calculating the optimal γ in (2) to obtain the elastic distance $d([\beta], [\tilde{\beta}])$ is equivalent to the following problem:*

$$\text{Maximize} \quad \Phi(t) = \sum_{j=0}^{m-1} \sqrt{(s_{j+1} - s_j) \int_{t_j}^{t_{j+1}} \langle p(t), q_j \rangle_+^2 \, dt}$$

(3)

$$\text{w.r.t.} \quad t = (t_1, \dots, t_{m-1}), \quad 0 = t_0 \le t_1 \le \cdots \le t_m = 1,$$

where $\langle \cdot, \cdot \rangle_+$ denotes the positive part of the scalar product in $\mathbb{R}^d$. For a maximizer t of (3), there is a $\gamma : [0, 1] \to [0, 1]$ with $\gamma(t_j) = s_j$ for all $j = 1, \dots, m-1$ that minimizes (2).

The proof includes an explicit construction of the minimizing warping function $\gamma \in \bar{\Gamma}$ (or a minimizing sequence of warping functions), where $\bar{\Gamma}$ is the set of absolutely continuous curves $\gamma : [0, 1] \to [0, 1]$, onto and with $\dot{\gamma} \ge 0$ almost everywhere. The statement for Γ follows as Γ is dense in $\bar{\Gamma}$ and the warping action of $\bar{\Gamma}$ continuous (Bruveris, 2016). Thus, the warping problem can be simplified if one of the SRV-curves is piecewise constant, independent of the form of the second SRV-curve p. If p is at least continuous, for example, the SRV-curve of a model-based smooth mean curve like the fish mean in Figure 1 on the top right, the loss function in (3) is differentiable. We propose to tackle the remaining maximization problem with a gradient descent algorithm that can handle linear constrains (for instance, method BFGS in constrOptim from R-package "stats;" R Core Team, 2020) and provide a derivation of the gradient in Web Appendix B.

2.3 | Elastic distance for two piecewise linear curves

We present an algorithm that can be used to find an optimal warping function, and therefore, compute the elastic distance, when both curves are piecewise linear. This is relevant either because we model one of the curves as a linear spline (mean) (see Subsection 2.4), as we do for the tree shapes in Figure 1, or because we want to compute the elastic distance between two observed curves, for example, two different discretely observed fish or trees. The latter allows any distance-based analysis of the data such as clustering or classification.

To obtain an optimal warping for a curve with piecewise constant SRV transformation q to a curve with SRV transformation p, we first note that the maximization in one t_j direction of the objective function in (3) only depends on the current values of t_{j-1} and t_{j+1} for any p. If p is also a piecewise constant SRV-curve, we can even derive a closed-form solution of the maximization problem in (3) w.r.t. each $t_j \in [t_{j-1}, t_{j+1}]$ (cf. Web Appendix B). Hence, we propose a coordinate wise maximization procedure in Algorithm 1, iterating updates of odd and even indices.

Algorithm 1: Elastic distance for two open polygons

Input: piecewise constant SRV-curves $\boldsymbol{p}, \boldsymbol{q}$; convergence tolerance $\epsilon > 0$;

starting values $0 \leqslant t_1^{(0)} \leqslant \cdots \leqslant t_{m-1}^{(0)} \leqslant 1$ `// e.g. relative arc length`

for $k \in \mathbb{N}$ **do**

    for $j = 1, \ldots, m-1$ **do**

       if $j - k$ *even* **then**

          $t_j^{(k)} = \operatorname{argmax}_{t_j \in \left[t_{j-1}^{(k-1)}, t_{j+1}^{(k-1)} \right]} \Phi \big|_{\{t_{j'} = t_{j'}^{(k-1)}, j' \neq j\}}$

       else if $j - k$ *odd* **then**

          $t_j^{(k)} = t_j^{(k-1)}$

    if $\| \boldsymbol{t}^{(k)} - \boldsymbol{t}^{(k-2)} \| < \epsilon$ *and* $\| \boldsymbol{t}^{(k-1)} - \boldsymbol{t}^{(k-3)} \| < \epsilon$ **then**

       return $\boldsymbol{t}^{(k)} = (t_1^{(k)}, \ldots, t_{m-1}^{(k)})$

The warping problem for two (open) piecewise linear curves has been previously discussed by Lahiri et al. (2015). They propose a precise matching algorithm, which produces a globally optimal reparameterization of $\boldsymbol{q}$, but is arguably demanding to implement. Our algorithm can be seen as an alternative, which is much more straightforward to understand and to extend to the closed case (cf. Web Appendix A) not explicitly addressed by Lahiri et al. (2015). We provide an implementation in the R-package "elasdics." Although our algorithm does not guarantee finding a globally optimal solution, we observe convincing results in simulations (Section 3) and can prove local maximization in the following sense:

Theorem 1. *Every accumulation point of the sequence* $(\boldsymbol{t}^{(k)})_{k \in \mathbb{N}} = (t_1^{(k)}, \ldots, t_{m-1}^{(k)})_{k \in \mathbb{N}}$ *resulting from Algorithm 1 is a local maximizer of Φ in (3).*

To prove this theorem, we first establish that the directional derivatives exist and are nonpositive for all coordinate directions. Then we show that this carries over to all directional derivatives using local concavity of the objective function.

If the sequence $(\boldsymbol{t}^{(k)})_{k \in \mathbb{N}}$ has more than one accumulation point, they all give the same value $\Phi(\boldsymbol{t})$. They then correspond to different reparameterizations of the second curve, but give the same distance between the two curves. This can happen as the warping problem does not guarantee unique solutions (see Web Appendix C for an example). In practice, one can pick any maximizing $\boldsymbol{t}$ to obtain a locally optimal warping function. As we cannot guarantee this $\boldsymbol{t}$ to also be a global maximizer, we propose using varying starting points to find a global maximum.

Our algorithm computes the elastic distance between two piecewise linear and continuous curves. These curves form a subspace in the space of absolutely continuous curves and are called splines of degree 1. For modeling smooth (differentiable) curves, for example, for a mean function, a spline space of a higher degree may be more suitable.

2.4 | Modeling spline curves or spline SRV-curves

As common in functional data analysis (Ramsay and Silverman, 2005), we like to model curves or means for samples of curves as splines. This is in particular beneficial for sparsely observed curves, which cannot be evaluated at arbitrary points. Moreover, splines impose parsimonious models for smooth curves, which can help to avoid overfitting the observed curves given limited information.

Definition 2 (Spline curves). We call $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)^T :$ $[0,1] \rightarrow \mathbb{R}^d$ with $d \in \mathbb{N}$ a d-dimensional spline curve of degree $l \in \mathbb{N}_0$ if all its components $\xi_1, \ldots, \xi_d : [0,1] \rightarrow \mathbb{R}$ are spline curves of degree l with a common knot set $0 = \kappa_0 < \kappa_1 < \cdots < \kappa_{K-1} < \kappa_K = 1$ for some $K \geq 2$. That means that $\xi_1, \ldots, \xi_d$ are piecewise polynomial of degree l between the knots $\kappa_0, \ldots, \kappa_K$, as well as continuous and $(l-1)$-times continuously differentiable on the whole domain [0,1] for $l \geq 1$. Denote by $S_{K; \kappa_0, \ldots, \kappa_K}^l$ the set of all such spline curves.

We can either model the curve $\boldsymbol{\beta}$ as a d-dimensional spline curve, or its SRV transformation $\boldsymbol{p}$ (see Figure 2). If $\boldsymbol{\beta}$ is a spline of degree $l \geq 2$, the corresponding SRV-curve $\boldsymbol{p}$
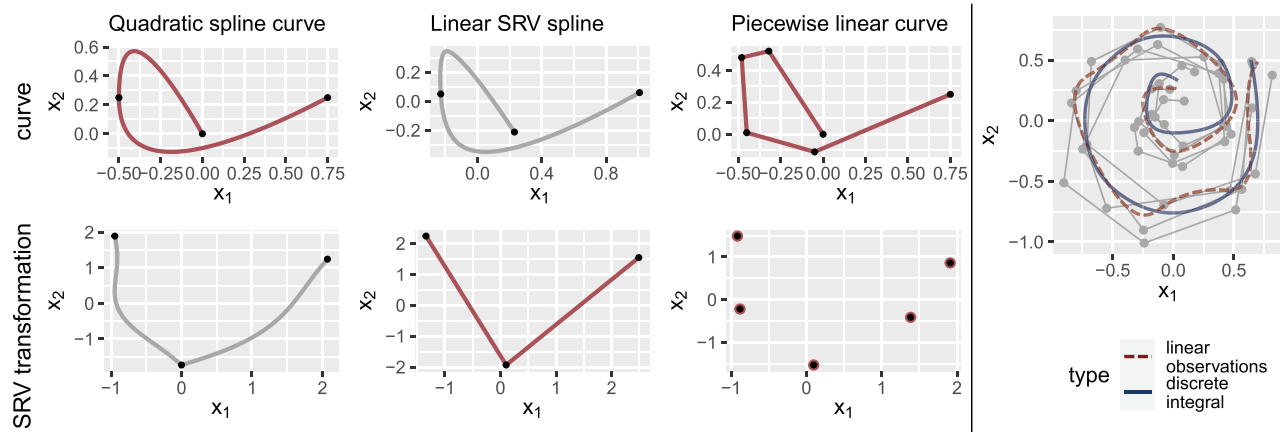
FIGURE 2 Left: Two-dimensional curves and corresponding SRV transformations. Spline curves are plotted as red curves with their values at knots marked as black dots; other curves are gray. Note that the SRV-curve in the sixth panel is piecewise constant in t and t is not visible in the image. Right: Smooth means (with 11 knots each) for four spiral curves based on linear splines on SRV level. The dashed mean curve is based on assuming piecewise linear observations for the integral approximations and the solid mean curve is based on the integral approximation using the mean value theorem

will not be a spline curve. The same holds true for curve $\boldsymbol{\beta}$ if $\boldsymbol{p}$ is a spline of degree $l \geq 1$. Only if $\boldsymbol{\beta}$ is piecewise linear ($l = 1$), then both $\boldsymbol{\beta}$ and its piecewise constant SRV transformation are splines. However, if we use linear spline curves, we need a large number of knots to obtain similarly smooth curves as using linear splines on SRV level, and thus, expect less parsimonious models.

To use these spline curves or spline SRV-curves as model spaces modulo warping, we need to ensure model identifiability, that is, that each equivalence class contains at most one spline curve. The unique spline representative then allows to identify and interpret the equivalence class of a curve modulo warping via its spline basis coefficients. We will see in Subsection 2.6 that this is true for quadratic or cubic splines on curve level and for linear spline SRV-curves (under mild conditions). Linear spline curves are identifiable under additional assumptions.

Therefore, we can use the space of cubic, quadratic, or linear spline curves as a model space for smooth curves. However, using quadratic or cubic splines on the curve level would not imply a vector space structure on the SRV level, where the distance is computed. We therefore propose to consider linear spline (and thus continuous) SRV-curves to model smooth curves. If $\boldsymbol{p}$ is a continuous SRV transformation of $\boldsymbol{\beta}$, the backtransform $\boldsymbol{\beta}(t) = \boldsymbol{\beta}(0) + \int_0^t \boldsymbol{p}(s)\|\boldsymbol{p}(s)\|ds$ is differentiable, as the norm $\|\cdot\|$ is also continuous. Alternatively, constant spline SRV-curves can be used to model less regular, polygonal mean curves. We thus work with a linear or constant spline model on SRV level in the following.

2.5 | Elastic means for samples of curves

As the space of curves modulo parameterization and translation does not form a Euclidean space, standard statistical techniques for describing probability distributions cannot be applied directly. In particular, we cannot define the expected value as an integral or the mean as a weighted average, which would require a linear structure of the space. To generalize the mean as a notion of location to arbitrary metric spaces, Fréchet (1948) proposed to use its property of being the minimizer of the expected squared distances.

Definition 3 Fréchet mean; Fréchet, 1948. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{X}$ a metric space with distance function d, equipped with the Borel-σ-Algebra. For a random variable $X : \Omega \to \mathcal{X}$, we call every element in $\operatorname{arginf}_{A \in \mathcal{X}} \boldsymbol{E}_P(d(X, A)^2)$ an expected element of X. For a set of observations $x_1, \ldots, x_n \in \mathcal{X}$, we define the Fréchet mean as an element in $\operatorname{arginf}_{A \in \mathcal{X}} \sum_{i=1}^n d(x_i, A)^2$.

Thus, Fréchet means are empirical versions of expected elements and neither of them need to exist or be unique. For a uniform distribution on the sphere, for example, every point on the sphere is a valid Fréchet mean. This nonuniqueness can occur for the elastic distance as well, see the example given in Web Appendix C. Nevertheless, Ziezold (1977) showed a set version of the law of large numbers for the Fréchet mean, which means that for independently and identically distributed random variables

$X_1, \ldots, X_n : \Omega \to \mathcal{X}$, the set of Fréchet means converges to the set of the expected elements.

As discussed in the previous subsection, we propose to use linear or constant splines on SRV level as model spaces for the Fréchet mean. For a set of curves with SRV transformations $q_1, \ldots, q_n$ and for a given degree $l \in \{0, 1\}$ and a given set of knots $\kappa_0, \ldots, \kappa_K$, we thus define

$$\bar{p} \in \operatorname*{arginf}_{p \in S^l_{K; \kappa_0, \ldots, \kappa_K}} \sum_{i=1}^n \inf_{\gamma_i} \left\| p - (q_i \circ \gamma_i)\sqrt{\dot{\gamma_i}} \right\|_{L_2}^2 \quad (4)$$

as the SRV transformation of the spline Fréchet mean (i.e., SRV transformation of the Fréchet mean restricted to the spline SRV space) w.r.t. the elastic distance (2). The corresponding restricted Fréchet mean $\bar{\beta}$ is thus either a polygon or a smooth curve. Similarly to the proposal of Srivastava and Klassen (2016) for densely observed curves, we tackle the minimization problem (4) with an iterative approach in Algorithm 2, alternating between

$\left\| \bar{p} - (q_i \circ \gamma_i)\sqrt{\dot{\gamma_i}} \right\|_{L_2}^2$ in the sum need to be approximated, because the curves β_i are only observed on a finite grid $0 = s_{i,0} \leq s_{i,1} \leq \cdots \leq s_{i,m_i} = 1$, and the SRV-curves $q_1, \ldots, q_n$ are thus unobserved. One option is to assume that the SRVs q_i of the observed curves are piecewise constant as in the warping step. As $\bar{p}$ is piecewise linear, $(q_i \circ \gamma_i)\sqrt{\dot{\gamma_i}}$ also is (see proof of Lemma 1 in Online Appendix B), which leads to a closed-form solution of the integral. Alternatively, we derive an approximation of the integrals in the L_2 fitting step of Algorithm 2 using the mean value theorem and the monotonicity of the warping in Web Appendix B.5. Both approaches lead to a (weighted) least-squares problem for the spline coefficients of $\bar{p}$. (An adapted algorithm for closed curves in Web Appendix A uses an additional penalty for openness with increasing weight.) We compare them using an example in Figure 2 on the right, where the second approach here leads to a better fit of the estimated spiral shape (and is used in the following).

Algorithm 2: Elastic spline Fréchet mean

Input: SRV transformations q_i of discretely observed curves β_i, $i = 1, \ldots, n$;

initial mean $\bar{p}_{new} = \operatorname{arginf}_{\bar{p}} \sum_{i=1}^n \|\bar{p} - q_i\|_{L_2}^2$; convergence tolerance $\epsilon > 0$

while $\|\bar{p}_{old} - \bar{p}_{new}\| > \epsilon$ **do**

$\quad \bar{p}_{old} = \bar{p}_{new}$;

$\quad \gamma_i = \operatorname{arginf}_\gamma \left\| \bar{p}_{old} - (q_i \circ \gamma)\sqrt{\dot{\gamma}} \right\|_{L_2}^2, \quad \forall i = 1, \ldots, n$;   // warping step

$\quad \bar{p}_{new} = \operatorname{arginf}_{\bar{p}} \sum_{i=1}^n \left\| \bar{p} - (q_i \circ \gamma_i)\sqrt{\dot{\gamma_i}} \right\|_{L_2}^2$ // L_2 spline fit via (weighted)

$\quad$ least-squares

return $\bar{p} = \bar{p}_{new}$

---

fitting the mean and optimizing the warping for each of the observations, but now using our warping approach for sparse curves and modeling the mean with a constant or linear spline. If we were to model the Fréchet mean in a spline space on curve level instead of SRV level, the mean fitting step would be a minimization problem in a nonlinear space, hence more challenging. That is why we refrain from using splines on curve level, although we show that quadratic and cubic splines are identifiable via their coefficients as well (Theorem 2).

For the warping step, we update the optimal warpings γ_i of the observed curves β_i, $i = 1, \ldots n$ via interpreting them as observed polygons with piecewise constant SRV transformations q_i, $i = 1, \ldots n$, as in Lemma 1. We tackle the remaining maximization problem (3) using a gradient descent algorithm as discussed before if $\bar{p}$ is piecewise linear and Algorithm 1 if $\bar{p}$ is piecewise constant. In the L_2 spline fitting step, the integrals

2.6 | Identifiability of spline curves

We model curves or means for samples of curves using basis representations. If we study equivalence classes of curves modulo reparameterization, we have to ensure unique spline representatives in each class, meaning that elements of the quotient space are identifiable via their basis coefficients. To see why this is not self-evident, consider as a simple counterexample in $\mathbb{R}^1$ the space of quadratic polynomials $P : [0, 1] \to \mathbb{R}$, a subspace of the quadratic spline space. Note that $\gamma_a(x) = ax^2 + (1 - a)x$ defines a feasible warping function for all $a \in]0, 1[$, because γ_a is differentiable with $\gamma'_a(x) \geq 0$ and $\gamma_a(0) = 0$, $\gamma_a(1) = 1$. Hence, all quadratic polynomials of the form $P(x) = p_1 \gamma_a(x) + p_0$ with $p_0, p_1 \in \mathbb{R}$ are elements of the same equivalence class, although they have varying basis coefficients $a p_1$, $(1 - a)p_1$ and p_0 for $a \in]0, 1[$ w.r.t. the monomial basis expansion. This counterexample shows in

particular that one-dimensional spline functions do not have unique representatives in the space of functions modulo reparameterization. Moreover, every 1d function is in the orbit of a linear spline with at least as many knots as the function has local extrema. As identifiability is essential in any modeling approach, it is fortunate that in contrast to $d = 1$, we can show that in $\mathbb{R}^d$ with $d \geq 2$, nearly all quadratic or cubic spline curves have unique basis representations.

Theorem 2. _Let $d \geq 2$ and $Q, P : [0, 1] \to \mathbb{R}^d$ be quadratic or cubic spline curves, where Q has a nonlinear image between each of its knots. Moreover, let $\gamma : [0, 1] \to [0, 1]$ be monotonically increasing and onto. Then $P = Q \circ \gamma \Rightarrow \gamma = id$._

Thus, nearly all equivalence classes modulo reparameterization contains at most one spline curve. Hence we can identify these curves modulo warping via their spline basis coefficients. The only exception are splines with linear image, which occur if and only if the splines in each coordinate direction are multiples of each other modulo translation. Note that we do not make any assumptions on the knots here, in particular the knots could be different for Q and P. That means there is almost always a unique representative modulo warping in $\bigcup_{K,\kappa_0,...,\kappa_K} \mathcal{S}^l_{K;\kappa_0,...,\kappa_K}$ for given $l = 2, 3$, that is, in the union of all spline spaces with varying (also varying number of) knots. Considering only quadratic or cubic splines is crucial, as this statement is not true for nonprime spline degrees. We show a counterexample for splines of degree four in Web Appendix C. The result for cubic spline curves also implies uniqueness of representatives for linear spline SRV-curves, another useful result for identifiable modeling of elastic curves.

Corollary 1. _Let $\beta_1, \beta_2 : [0, 1] \to \mathbb{R}^d$ with SRV functions q_1 and q_2, respectively. If q_1 and q_2 are nowhere constant linear splines and $q_2(t) = q_1(\gamma(t))\sqrt{\dot{\gamma}(t)}$, then $q_1 = q_2$._

In summary, the space of linear SRV spline curves seems particularly suitable to model smooth elastic curves as they are identifiable, that is, there is a unique representation in this space, and the corresponding curves are differentiable, which leads to visually smooth curves. In our toy example, we used linear spline SRV-curves to model the smooth fish mean (Figure 1, top right).

Remark 1 (Linear spline curves). Linear spline curves or equivalently piecewise constant SRV-curves are identifiable via their spline basis coefficients modulo warping, if we consider one spline space $\mathcal{S}^1_{K;\kappa_0,...,\kappa_K}$ but not the union of several such spaces, and assume that the curve is not differentiable at all of its knots (i.e., no knot is superfluous). For an illustration, see Web Appendix C.

Hence, with this weaker identifiability result, piecewise constant SRV-curves are a suitable model space as well, with curves modeled as polygons. This is more appropriate for mean curves that are assumed to have sharp corners, like the trees in Figure 1.

As we use these spline spaces for estimation of smooth or polygonal curves, we need the following result on continuity of the embedding. It allows us to interpret estimated coefficients—for instance, compare the coefficients of two estimated group means to investigate local differences—as it ensures convergence of the spline coefficients if we construct a converging sequence of curves. For instance, we aim to construct such a sequence for the elastic mean in Algorithm 2. We show that this continuity property holds whenever the model space Ξ is a (subset of a) finite-dimensional spline space of the following form. Note that, for simplicity, we do not consider unions of spline spaces here.

Definition 4. Let Ξ be one of the following for given fixed $K \geq 2$, $0 = \kappa_0 < \cdots < \kappa_K = 1$: (i) a subset of $\mathcal{S}^l_{K;\kappa_0,...,\kappa_K}$, $l = 2, 3$, which consists of identifiable splines as described in Theorem 2, additionally centered (i.e., with integral zero) to account for translation; (ii) a set of identifiable curves with linear spline SRV-curves in $\mathcal{S}^1_{K;\kappa_0,...,\kappa_K}$ from Corollary 1; or (iii) the set of curves with piecewise constant SRV-curves in $\mathcal{S}^0_{K;\kappa_0,...,\kappa_K}$ from Remark 1.

Lemma 2 (Topological embedding). _Let $f : (\Xi, \|\cdot\|) \to (\mathcal{A}, d)$ be the embedding of the spline coefficients defining the functions in Ξ, equipped with the usual Euclidean distance $\|\cdot\|$, into the space $\mathcal{A}$ of absolutely continuous curves w.r.t. the elastic distance d. Then f is a topological embedding, that is, f is a homeomorphism on its image._

Thus, the distance of spline coefficients and the elastic distance of curves modulo translation are topologically equivalent on suitable spline spaces. Consequently, a sequence of curves converges w.r.t. the spline coefficients if, and only if, it converges w.r.t. the elastic distance. Overall, we see that any spline model Ξ in Definition 4 yields an identifiable model for the Fréchet mean of observed curves, with the possibility to interpret spline coefficients. This also holds for converging series of estimators which we aim to construct in our algorithms.

3 | SIMULATION

We test our methods, which we made available for public use in the R-package "elasdics," on simulated data. A first simulation focuses on the special case of equal numbers of observed points on the curves, where we can

compare our methods to an existing implementation of the SRV framework in the R package "fdasrvf" (Tucker, 2020) based on Srivastava et al. (2010). Results presented in Web Appendix D show that Algorithm 1 (and its variant for closed curves) produce clearly better alignment for sparsely and irregularly sampled curves. The corresponding average elastic distance is smaller for our method in all cases, for example, a reduction of 25% and 26% on average for 30 observed points per curve in the open and closed setting, respectively. As expected, this difference decreases if 90 points of the closed butterfly shapes are selected (1% reduction on average), as in this case, the points are nearly observed on a regular, fairly dense grid, which is the setting "fdasrvf" is designed for. This simulation also shows that a highly unbalanced distribution of observed points on the curves causes difficulties for the mean computation in "fdasrvf" as well, which is not the case for our methods.

Here we mainly discuss the second simulation, focusing on the convergence and the identifiability of the newly proposed spline means and their associated coefficients. As we vary the number of points per curve, there is no competitor to compare our methods with. For a given template curve β with known B-spline coefficients $\vartheta_1, \ldots, \vartheta_B$, we generate a sample of observed curves $\beta_1, \ldots, \beta_n$ by independently sampling the coefficients $\vartheta_{i,b} \sim \mathcal{N}(\vartheta_b, \sigma^2)$ for all $i = 1, \ldots, n$, $b = 1, \ldots, B$. If the template curve is closed, we additionally close the sampled curves via minimizing a penalty function penalizing openness in gradient direction. The penalty is given in Web Appendix A for estimating a closed mean. The points $t_{i,1}, \ldots, t_{i,m_i-1}$ on which β_i is observed are sampled uniformly on $[0,1]$, where the number of observed points m_i is sampled uniformly either from $\{10, \ldots, 15\}$ (very sparse and unbalanced) or $\{30, \ldots, 50\}$ (less sparse but unbalanced).

Examples for curves sampled with standard deviation $\sigma = 4$ from a heart-shaped template curve, modeled as linear spline on SRV level with 10 equally spaced inner knots, are displayed in Figure 3. Two further examples for open curves are given in Web Appendix C. The samples in the very sparse setting are hardly recognizable as heart shapes (Figure 3, right). However, the elastic mean curve over $n = 5$ observations, estimated using the true knot set and linear SRV splines to allow a comparison of estimated and true coefficients, represents the original heart surprisingly well even in this challenging setting. We repeated this simulation 100 times each for varying numbers of observations $n \in \{5, 20\}$ and observed points per curve m_i (Figure 3, left). For $m_i \in \{10, \ldots, 15\}$ observations per curve, we generally obtain a heart-shaped mean, which seems smaller and shows less pronounced features than

the template. Increasing the number of observed curves from $n = 5$ to $n = 20$ decreases the variance of the mean curve, but a certain bias due to undersampling the curves remains. Likewise, the variance of the spline mean coefficients is smaller for $n = 20$ than for $n = 5$, but their distribution is still not centered at the coefficients of the template (indicated as black dots in Figure 3).

If we increase the number of points on each curve to $m_i \in \{30, \ldots, 50\}$, the estimated means w.r.t. the elastic distance adapt closer to the template. Moreover, the variance of the estimated spline coefficients decreases as well as their distance to the template. The reduction of variance indicates convergence of the spline coefficients for $n \to \infty$, although we do not expect them to precisely converge to the coefficients of the template in this simulation setup, not even if $m_i \to \infty$ for all $i = 1, \ldots, n$. This is because we draw the sample curves $\beta_1, \ldots, \beta_n$ such that β is the mean w.r.t. the L_2 distance on SRV level, but this does in general not imply that β is the mean w.r.t. the elastic distance. Nevertheless, we expect this difference to be small, as the coefficients in the rightmost boxplot are close to the black dots that indicate the template's coefficients. In addition, their low variance for $n = 20$ confirms our theoretical results on identifiability of spline coefficients in our model (Corollary 1) and continuity of the embedding (Lemma 2).

As expected, the run time of our elastic mean algorithm grows with the number of observed curves as well as with the number of observed points per curve. On a standard Windows PC, we report run times of 19 s ($n = 5$) and 30 s ($n = 20$) on average for one mean in the very sparse setting. In the less sparse setting, $m_i \in \{30, \ldots, 50\}$, the run times increase to 22 and 88 s for $n = 5$ and $n = 20$, respectively.

So far, we have discussed the convergence of correctly specified spline means, as in this case, convergence of elastic means corresponds to convergence of the corresponding spline coefficients (Lemma 2). As correct specification is questionable in practice, we demonstrate the behavior of our methods in the case of model misspecification (varying spline degree and number of knots) in a further simulation given in Web Appendix D. We observe that both smooth and polygonal means reproduce the original template well and that results are not very sensitive to the number of knots, given that it is sufficiently large. Generally, the elastic distance to the template decreases for an increasing number of knots. Distances to the template are smaller for the smooth than for the polygonal model means for a fixed number of knots, and decrease to a lower level, indicating more parsimonious models and less undersampling bias for truly smooth means when using linear SRV-curve models.
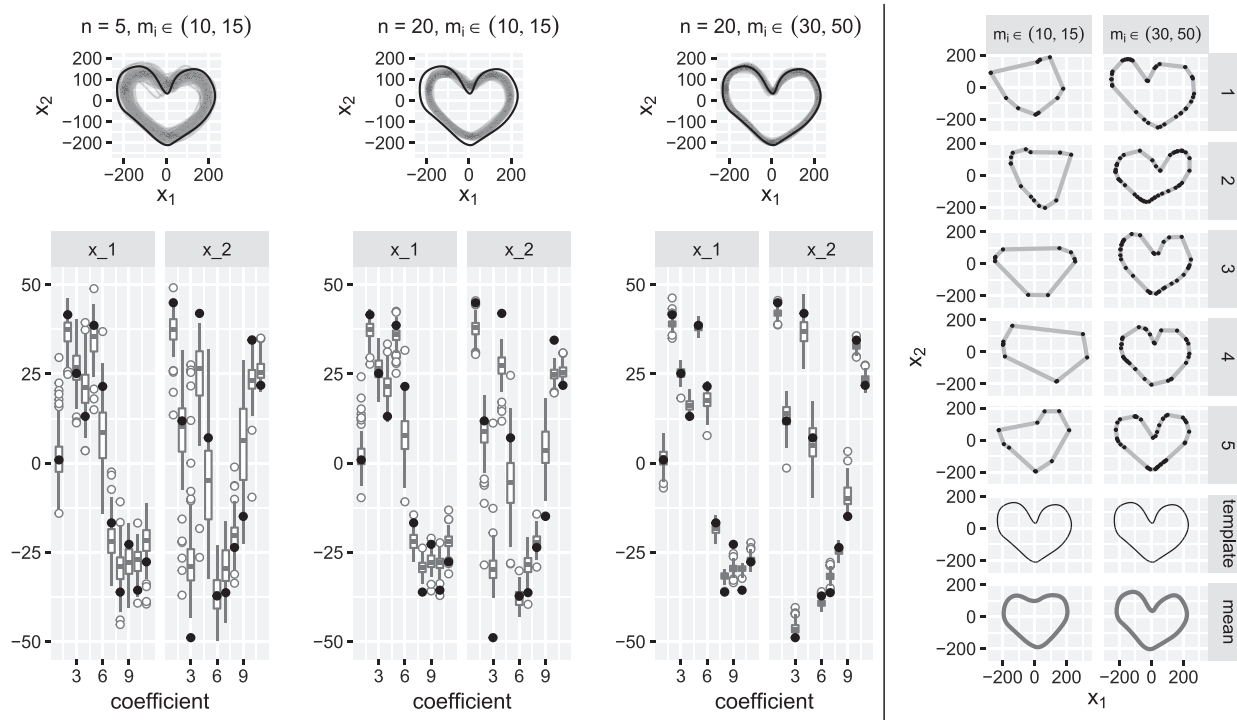
FIGURE 3 Top left: Smooth means (in gray) computed for a set of n simulated curves drawn from the heart-shaped template curve (in black) via sampling its B-spline coefficients from a normal distribution with standard deviation $\sigma = 4$ and $m_i, i = 1, \ldots, n$ points observed per curve. The means are computed using linear SRV splines and the same knot set as the template (10 equally spaced inner knots). Bottom left: Corresponding distribution of spline mean coefficients (in gray) and template coefficients (in black). Right: Simulated data $i = 1, \ldots 5$ with observed values marked as black dots and corresponding smooth elastic means over $n = 5$ observations in gray

4 | APPLICATIONS ON REAL DATA

As our main goal is to develop statistical (elastic) analysis methods for discretely observed data curves, we demonstrate their practicality on two datasets.

4.1 | Classifying spiral curve drawings for detecting Parkinson's disease

(Isenkul et al., 2014) provide a dataset of spiral curve drawings by Parkinson patients and healthy controls in a so-called Archimedes spiral-drawing test, which is a common, noninvasive tool for diagnosing patients with Parkinson's disease. The data have been obtained in two different settings: In the "static spiral test," the participants had to follow a template on a digital tablet; in the "dynamic test," the template curve appeared and disappeared in certain time intervals. We propose an intuitive classifier mimicking a doctor's decision of the form: Classify as "Parkinson" if the distance of the drawn curve to the template curve exceeds a threshold for one or for both of the settings. As the template curve has not been recorded, we use the elastic mean (see Subsection 2.5) of all curves

from the static spiral test with piecewise constant splines and 201 knots on SRV level, instead. Then we compute the elastic distance of each observed spiral curve to the template using Algorithm 1. We report a leave-one-curve-out cross-validated accuracy of 72.5% for the static, 90.0% for the dynamic setting, and 92.5% for the classifier based on both, which indicates good separation in particular for the dynamic spiral test.

A detailed description of our analysis and a comparison to the methods implemented in the "fdasrvf" package can be found in Web Appendix E. Our methods lead to better classification accuracy in this application and the mean calculation proves to be faster.

4.2 | Clustering and modeling smooth means of GPS-tracks

The second dataset is an example of increasingly common human movement data and comprises GPS waypoints tracked on Tempelhof Field, a former airfield (up to 2008) in Berlin, which is now used as a recreation area. The dataset consists of 55 paths with 15–45 waypoints each, recorded by members of our working group using their
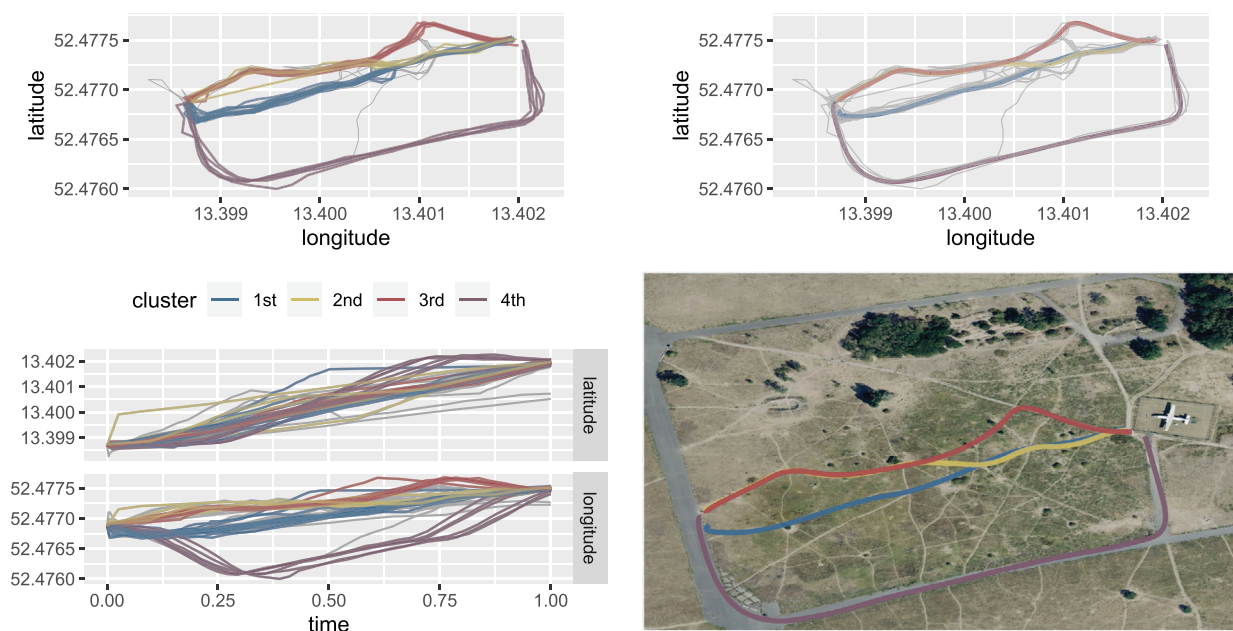
FIGURE 4 Top left: The observed trajectories with elements of the four largest clusters indicated by color. Bottom left: Longitude and latitude for the trajectories (with the four largest clusters indicated by the same colors) over relative time. Top right: Smooth means modeled as linear SRV-curves with 10 inner knots for the four largest clusters and centered at the mean center of the observed paths per cluster to account for translation. Bottom right: Cluster means plotted on Microsoft Bing Map accessed via the R package "OpenStreetMap" (Fellows, 2019)

mobile phones for tracking. Due to the variety of mobile devices used, the number of points per curve differs considerably, resulting in irregularly and quite sparsely observed data. We are solely interested in analyzing the paths (Figure 4, bottom right) the participants walked on, not the trajectories over time. Separately looking at longitude and latitude over time suggests that the individuals had quite different walking patterns and did not move with constant speed. This implies that standard (nonelastic) functional data analysis is not suitable here.

Clustering and smooth mean estimation allow us to recover the paths that the individuals walked on. In a further step, these could be used to identify new paths on Tempelhof field not yet included in existing maps. In a first step, the tracks are clustered using average linkage based on the elastic distance and the elbow criterion for stopping. Here we apply Algorithm 1 to approximate the pairwise distance between the sparsely observed open tracks. In a second step, we compute a smooth elastic Fréchet mean for each of the four largest clusters using Algorithm 2 and linear splines on SRV level with 10 inner knots. The clustering result displayed in Figure 4, top row, is visually satisfying. Looking at longitude and latitude separately clearly indicates that clustering based on the L_2 distance would not work well.

The smooth mean curves for each of the four largest clusters (Figure 4, top right) seem to describe the observed tracks well, despite the dimension reduction (24 spline

coefficients compared to 30–90 observations per curve) and also match the actual paths visible in the satellite image (Figure 4, bottom right) provide by Microsoft Bing and made available for R in the package "OpenStreetMap" (Fellows, 2019).

5 | DISCUSSION

Although our approach addresses the discrete and often sparse nature of observed curves explicitly, the interpretation as polygons with observed values at the corners underestimates the curvature of the real unobserved curves. This leads to a kind of shrinkage bias for the estimated elastic mean for sparsely observed curves. Although this bias toward curves with smaller curvature decreases with increasing observations per curve, it would be of interest to develop correction methods for (very) sparse settings in future work.

We have shown that the SRV splines modulo parameterization used for modeling the elastic mean is in general identifiable via their coefficients and we have confirmed this result in simulations. Although we did not explicitly address the choice of the optimal number of knots for such splines, a further simulation has shown that the estimation of the mean curve is not sensitive to the specific spline degree and choice of knots, given the number of knots is sufficiently large. As the union of any spline

space with fixed degree but varying knots is dense in the space of absolutely continuous curves w.r.t. the elastic distance, using an increasing number of knots would ensure that the mean curve can be arbitrarily well approximated. For a finite dataset, this would lead to overfitting the curves though, which may be addressed via penalized estimation, although the interpretation of coefficients and convergence properties would need to be studied in this setting.

Another appealing direction for further research is to include our methods for sparsely and irregularly sampled curves in existing approaches for functional shape analysis. Here the curves have to be aligned w.r.t. scaling and/or rotation in addition to the alignment w.r.t. parameterization and translation. As this is usually done iteratively, it seems promising to combine this with the iterative warping and mean fitting steps in our methods. Furthermore, elastic mean estimation for irregularly and/or sparsely sampled curves can be seen as a first step toward elastic regression models for such data. That means our methods might be useful building blocks for modeling curves or shapes depending on covariates using splines.

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are available in the Supporting Information of this article, with the exception of the data analyzed in the Parkinson's spirals application, which are available from https://www.researchgate.net/publication/291814924.

ORCID

Lisa Steyer https://orcid.org/0000-0002-6987-1520
Almond Stöcker https://orcid.org/0000-0001-9160-2397
Sonja Greven https://orcid.org/0000-0003-0495-850X

REFERENCES

Backenroth, D., Goldsmith, J., Harran, M.D., Cortes, J.C., Krakauer, J.W. & Kitago, T. (2018) Modeling motor learning using heteroscedastic functional principal components analysis. *Journal of the American Statistical Association*, 113(523), 1003–1015.

Bernal, J., Dogan, G. & Hagwood, C.R. (2016) Fast dynamic programming for elastic registration of curves. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 111–118).

Bruveris, M. (2016) Optimal reparametrizations in the square root velocity framework. *SIAM Journal on Mathematical Analysis*, 48(6), 4335–4354.

Dryden, I. & Mardia, K. (2016) *Statistical shape analysis: with applications in R*. Wiley Series in Probability and Statistics. Chichester: Wiley.

Fellows, I. (2019) *OpenStreetMap: Access to Open Street Map Raster Images*. R package version 0.3.4.

Fréchet, M. (1948) Les éléments aléatoires de nature quelconque dans un espace distancié. In: *Annales de l'institut Henri Poincaré*, volume 10 (pp. 215–310).

Greven, S. & Scheipl, F. (2017) A general framework for functional regression modelling. *Statistical Modelling*, 17(1–2), 1–35.

Isenkul, M., Sakar, B., Kursun, O. et al. (2014) Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease. In: *The 2nd International Conference on e-Health and Telemedicine (ICEHTM-2014)*, volume 5 (pp. 171–175).

Joshi, S.H., Narr, K., Phillips, O., Nuechterlein, K., Asarnow, R., Toga, A. & Woods, R. (2013) Statistical shape analysis of the corpus callosum in schizophrenia. *NeuroImage*, 64, 547–559.

Lahiri, S., Robinson, D. & Klassen, E. (2015) Precise matching of PL curves in R^N in the square root velocity framework. *Geometry, Imaging and Computing*, 2, 133–186.

Lu, Y., Herbei, R. & Kurtek, S. (2017) Bayesian registration of functions with a Gaussian process prior. *Journal of Computational and Graphical Statistics*, 26(4), 894–904.

Marron, J.S., Ramsay, J.O., Sangalli, L.M. & Srivastava, A. (2015) Functional data analysis of amplitude and phase variation. *Statistical Science*, 30(4), 468–484.

Matuk, J., Bharath, K., Chkrebtii, O. & Kurtek, S. (2021) Bayesian framework for simultaneous registration and estimation of noisy, sparse, and fragmented functional data. *Journal of the American Statistical Association*, 1–17, in press.

R Core Team. (2020) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramsay, J. & Silverman, B. (2005) *Functional data analysis*. Springer Series in Statistics. New York, NY: Springer.

Ramsay, J.O. & Li, X. (1998) Curve registration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2), 351–363.

Sakoe, H. & Chiba, S. (1978) Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26, 159–165.

Srivastava, A. & Klassen, E. (2016) *Functional and shape data analysis*. Springer Series in Statistics. New York: Springer.

Srivastava, A., Klassen, E., Joshi, S. & Jermyn, I. (2010) Shape analysis of elastic curves in Euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7), 1415–1428.

Steyer, L. (2021) *elasdics: elastic analysis of sparse, dense and irregular curves*. R package version 0.2.0.

Tucker, J.D. (2020) *fdasrvf: elastic functional data analysis*. R package version 1.9.7.

Yao, F., Müller, H.G. & Wang, J.L. (2005) Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470), 577–590.

Ziezold, H. (1977) On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In: *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians* (pp. 591–602). Springer.

SUPPORTING INFORMATION

Web Appendices A, B and C referenced in Section 2 and Web Appendix D referenced in Sections 3 and 4 are available with this paper at the Biometrics website on Wiley Online Library. All developed methods are implemented in the R-package elasdics (Steyer, 2021) available on CRAN and the code to reproduce the findings of this paper is available in the Supporting Information of this article.

Figure 1: First three iterations of the algorithm for closed mean curves on a toy dataset

Figure 2: Left: Two piecewise linear curves in gray with Frechet mean curves in red and blue

Figure 3: Three constant SRV splines (right) with corresponding linear spline curves (middle)

Figure 4: Comparison of the optimal alignment produced by our method CWO and the one computed with DP

Figure 5: Elastic means for irregularly sampled curves

Figure 6: Example simulated data in gray with observed values marked as black dots and corresponding smooth elastic means over n = 5 observations in blue

Figure 7: Top: Smooth means (in blue) computed for a set of n curves drawn from the open template curve (in red) via sampling its B-spline coefficients from a normal distribution with standard deviation $\sigma = 0.3$ and mi, i=1, . . . , n points observed per curve

Figure 8: Top: Smooth means (in blue) computed for a set of n curves drawn from the open template curve (in red) via sampling its B-spline coefficients from a normal distribution with standard deviation $\sigma=0.4$ and mi, i=1,..., n points observed per curve

Figure 9: Left: Smooth mean based on linear splines on SRV level with varying number of knots and therefore coefficients computed on a sample of 20 curves with mi ϵ30, 50 points per curve

Figure 10: Left: Spiral curves drawn by either a healthy control group or by patients with Parkinson's disease in two different settings

Figure 11: Left: Distance of the curves drawn by the participants to the mean spiral curve for both settings

Figure 12: Optimal warping in both settings separated by the actual status and the predicted status using the classifiers based on only the corresponding distance each and leave-one-out cross-validation

Table 1: Classification accuracy in the dynamic setting with a varying fraction of points per curve

Table 2: Comparison of the classification accuracy in the dynamic setting with a varying number of points per curve

Table 3: Run-times for the mean computation of the spiral data in seconds

Figure 13: Left: Comparison of means for the spirals in the static setting with 100 observations per curve

Data S1