Computational Statistics and Data Analysis 151 (2020) 107017

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Regression models using shapes of functions as predictors

Kyungmin Ahn [a,*], J. Derek Tucker [b], Wei Wu [c], Anuj Srivastava [c]

[a] RIKEN Center for Biosystems Dynamics Research (BDR), Kobe, Japan
[b] Sandia National Laboratories, Albuquerque, NM, USA
[c] Florida State University, Tallahassee, FL, USA

ARTICLE INFO

Article history:
Received 2 September 2019
Received in revised form 19 May 2020
Accepted 23 May 2020
Available online 30 May 2020

Keywords:
Functional data analysis
Scalar-on-function regression
Functional single-index model
Function alignment
SRVF

ABSTRACT

Functional variables are often used as predictors in regression problems. A commonly used parametric approach, called *scalar-on-function regression*, uses the $\mathbb{L}^2$ inner product to map functional predictors into scalar responses. This method can perform poorly when predictor functions contain undesired phase variability, causing phases to have disproportionately large influence on the response variable. One past solution has been to perform phase–amplitude separation (as a pre-processing step) and then use only the amplitudes in the regression model. Here we propose a more integrated approach, termed *elastic functional regression model* (EFRM), where phase-separation is performed inside the regression model, rather than as a pre-processing step. This approach generalizes the notion of phase in functional data, and is based on the norm-preserving time warping of predictors. Due to its invariance properties, this representation provides robustness to predictor phase variability and results in improved predictions of the response variable over traditional models. We demonstrate this framework using a number of datasets involving gait signals, NMR data, and stock market prices.

1. Introduction

A fast growing subtopic in functional data analysis (FDA) (Ramsay and Silverman, 2005) is *regression* involving functional variables, either as predictors or responses or both. Morris (2015) categorizes regression problems involving functional data into three types: (1) functional predictor regression (scalar-on-function), (2) functional response regression (function-on-scalar) and (3) function-on-function regression. The functional predictor regression problem (or scalar-on-function) model was first studied by Ramsay and Dalzell (1991) and Cardot et al. (1999), and several other since then (Ahn et al., 2018; James, 2002; Reiss et al., 2017; Goldsmith and Scheipl, 2014; Fuchs et al., 2015; Ciarleglio and Ogden, 2016; Gertheiss et al., 2013; Cai and Hall, 2006). In this set up, predictors are scalar-valued functions on a fixed interval say $[0, T]$, call them $\{f_i \in \mathcal{F}\}$, elements of some pre-specified functional space $\mathcal{F}$, and responses are scalars $\{y_i \in \mathbb{R}\}$ (One can easily extend this framework to the case where responses are vector-valued). A simple and commonly-used model for this problem is the so-called *functional linear regression model* (FLM) given by:

$$y_i = \alpha + \langle \beta, f_i \rangle + \epsilon_i, \quad i = 1, \ldots, n \,, \tag{1}$$

where $\alpha \in \mathbb{R}$ is the intercept, $\beta \in \mathcal{F}$ is the regression-coefficient function, and $\epsilon_i \in \mathbb{R}$ is the observation noise. Also, $\langle \beta, f_i \rangle$ denotes the standard $\mathbb{L}^2$ inner product $\int_0^T f_i(t)\beta(t)\,dt$. (Notationally, we will use $\|\cdot\|$ to denote the $\mathbb{L}^2$ norm.) One assumes

* Corresponding author.
E-mail address: kyungmin.ahn@riken.jp (K. Ahn).

https://doi.org/10.1016/j.csda.2020.107017
0167-9473/

here that $\mathcal{F}$ has the $\mathbb{L}^2$ Hilbert structure to allow for this inner product between its elements. Similar to linear regression models with Euclidean variables, one can also estimate model parameters here by minimizing the sum of squared errors (SSE):

$$\{\hat{\alpha}, \hat{\beta}\} = \underset{\alpha\in\mathbb{R}, \beta\in\mathbb{L}^2}{\operatorname{argmin}} \left[\sum_{i=1}^{n} (y_i - \alpha - \langle \beta, f_i \rangle)^2 \right] . \tag{2}$$

However, since $\mathbb{L}^2$ is infinite dimensional, this problem is not sufficiently constrained to estimate $\hat{\beta}$ with a finite sample size n, and requires further restrictions. These constraints can come in form of a regularization term or a restriction of the solution space, or both. For restricting the solution space, one can use a complete orthonormal basis of $\mathcal{F}$, for representing β via its coefficients, and then truncate it to make the representation finite dimensional. A regularization is often imposed using a roughness measure on β, e.g. $\int \ddot{\beta}(t)^2 dt$ (For a function $f(t)$, we will use $\dot{f}(t)$ and $\ddot{f}(t)$ to denote its first and the second derivatives, respectively). The FLM model can easily be extended to a *generalized FLM* (Müller and Stadtmüller, 2005), where the conditional mean of the response given the predictors uses a known link function.

1.1. Basic issue: Predictor phase

While the use of functional data has grown in recent years, there has also been a growing awareness of a problem/issue that is specific to functions. Functional data often comes with a *phase variability*, *i.e.* a lack of registration between geometric features (peaks, valleys, etc.) across functions (Marron et al., 2015, 2014; Srivastava et al., 2011). Different observations can potentially represent different temporal rates of evolutions, introducing an intrinsic phase variability in the data. This situation arises, for example, in biological signals, growth curves, pandemic curves, and stock market data. In all these examples, functional measurements often lack temporal synchronizations across measurements.

In mathematical terms, the functional data is not $\{f_i\}$, as in the original model, but observed under random time warpings. Let Γ be the set of all time warping functions (formally defined later). In fact, depending on the context, three types of warpings are possible.

- **Value-preserving warping**: The most commonly-used mapping is $f_i \mapsto (f_i \circ \gamma_i)$. It is called *value-preserving warping* as it preserves the heights of the function f_i and only shifts them horizontally. It is often used in the alignment of peaks and valleys in functional data.
- **Area-preserving warping**: The mapping $f_i \mapsto (f_i \circ \gamma_i)\dot{\gamma}_i$ is called an *area-preserving warping* since it preserves the area under the curve f_i. It is often used when $\{f_i\}$ are probability density functions.
- **Norm-preserving warping**: Another warping results from the mapping $f_i \mapsto (f_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}$, called a *norm-preserving warping*, since it preserves the $\mathbb{L}^2$ norm of f_i. That is, $\|f_i\| = \|(f_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}\|$ for all $f_i \in \mathbb{L}^2$ and $\gamma \in \Gamma$.

Additional types of warpings may also be possible, depending on the need of the application. While in functional data alignment, one mainly uses the value-preserving warping of functions, we will keep our options more general in this paper. Since our goal is regression and prediction, not just functional alignment, we are free to incorporate any type of warping in the model, as needed. In the following, we will use $(f_i * \gamma_i)$ as a general notation for any of the above-mentioned warpings. The exact form will be made clear in specific contexts.

In FDA, it is often advantageous and sometimes imperative to take into account time warpings of functional data. Examples of such treatments in data analysis include (Marron et al., 2015, 2014; Srivastava et al., 2011) and in data modeling include Tucker et al. (2013). For instance, a common idea in FDA is to perform alignment of peaks and valleys across functions using the value-preserving warpings $(f_i * \gamma_i = f_i \circ \gamma_i)$ of their domains. These warpings $\{\gamma_i\}$ correspond to the *phase* components and the aligned functions $\{f_i \circ \gamma_i\}$ correspond to the *shape* or the *amplitude* components. To illustrate this, consider two data examples shown in Fig. 1. On the left, we see the *Tecator* data which shows absorbance curves for certain meat and has been used commonly in several FDA papers (Febrero-Bande and Oviedo de la Fuente, 2012; García-Portugués et al., 2014). These functions appear well registered and one can use them directly in a statistical model without any consideration of phase. The right side shows a different situation, involving the famous *Berkeley growth* data, where height changes of 69 male subjects are displayed in the middle panel. While these curves have a similar number of peaks and valleys, these features are not well aligned, due to differences in growth rates and the body clocks across subjects. Since this data contains a larger phase variability, the problem of phase–amplitude separation becomes important. The result of one such alignment algorithm (Srivastava et al., 2011) applied to the data is shown in the right panel. As the reader can see, the peaks and valleys in functions are now well aligned.

In general regression models both components of predictors – phase and shape – are useful. However, there are situations where only one of them, most notably, the shape, is of interest in predicting a response variable. This situation arises, for instance, in cases where the response depends primarily on the numbers and heights of the modes in the predictor functions, and the *locations* of modes and anti-modes are not influential and are considered nuisance. To motivate this further, using the human growth data, imagine a certain response variable, say the gender of the subject, that depends primarily on shapes of these curves and not on the locations of growth spurts. Thus, shape-based functional regression becomes a useful tool in this context. Motivated by such problems, we shall develop a regression model where only the shape (or amplitude) of a function is used in the model and its phase is removed from the consideration.

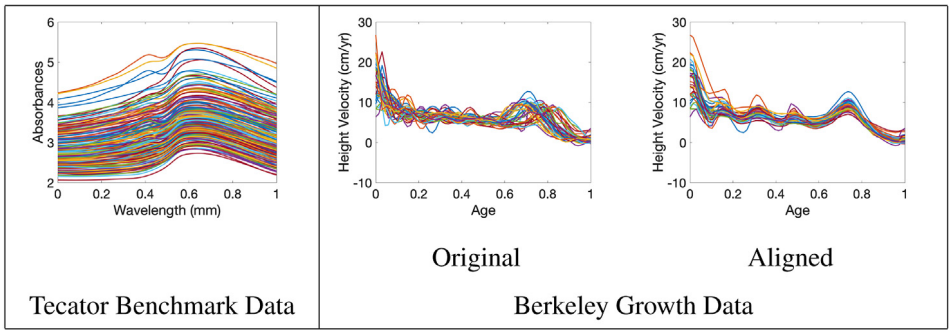


Fig. 1. Example of functional data with and without phase variability.

The phase variability in functional predictors, even if small, can have a disproportionately large influence on statistical analysis. One consequence of phase variability is the inflation of variance in the predictor itself, *i.e.* the variance of $\{(f_i * \gamma_i)\}$ can be much bigger than that of $\{f_i\}$, rendering any ensuing variance-based analysis ineffective. Another consequence is the change in the regression model itself. Under the value-preserving warping, using the Taylors' expansion, we get

$$f_i(\gamma_i(t)) = f_i(\gamma_{id}(t)) + \dot{f}_i(t)(\gamma_i(t) - \gamma_{id}(t)) + \quad \text{higher order terms} ,$$

with $\gamma_{id}(t) = t$. Dropping the higher-order terms and replacing f_i by $f_i \circ \gamma_i$ in Eq. (1), we get

$$E[y_i|\beta, f_i] = \alpha + \langle \beta, f_i \rangle + \left\langle \beta, \dot{f}_i \cdot (\gamma_i - \gamma_{id}) \right\rangle .$$

The conditional mean gets changed, up to the first order, by an amount captured by the third term on the right side. Depending on the value of $\{\dot{f}_i\}$, this change can be significant, adversely affecting the prediction performance. Although this derivation involves value-preserving time warping, a similar analysis can be repeated for other group actions also, with similar conclusions. Sometimes phase variability is due to simple linear or affine shifts, and can be handled trivially, but in general phases are nonlinear functions and require more comprehensive mathematical tools.

We further illustrate the issue of phase variability using a simulated example. Specifically, we quantify deterioration in prediction performance as the amount of random warpings in the predictor functions is increased. The results are presented in Fig. 2. The left panel shows a set of predictors $\{f_i\}$ used in these experiments. For a fixed β and $\alpha = 0$, we simulate responses y_is using Eq. (1). Then, we use this data $\{(f_i, y_i), i = 1, 2, \ldots, 100\}$ to estimate the model parameters, including $\hat{\beta}$, using Eq. (2) (using a finite number of basis elements to represent β). Next, we use this estimated $\hat{\beta}$ to predict responses y_i^{test} for new predictors f_i^{test}. However, we *contaminate* the test predictors in one of two ways: (i) value preserving $f_i^{test} \mapsto (f_i^{test} \circ \gamma_i)$, and (ii) area preserving $f_i^{test} \mapsto (f_i^{test} \circ \gamma_i)\sqrt{\dot{\gamma}_i}$. Ignoring this contamination and using a standard predictor, we obtain predictions and quantify prediction performance using the coefficient of determination R^2. Specifically, we study changes in R^2 as the amount of contamination (warping noise) increases. The warping functions used in this experiment are given by $\gamma_i(t) = t + \alpha_i t(1-t)$, where $\alpha_i \sim U(-a, a)$; the larger the value of a, the larger is the warping noise. The bottom row shows examples of warping functions for different values of a. The middle and the last panels in the top row show plots of R^2 versus a (averaged over 200 runs) for the two types of contaminations. In both cases we observe a superlinear decay in the performance as a increases. These experiments underline the fact that even a small amount of phase variability in predictors, either value-preserving or norm-preserving, can lead to a significant deterioration in the prediction performance. Thus, one needs to account for this variability inside the model itself in an intrinsic way.

We reiterate that phase is nuisance in some but not all situations. One should not always expect the shapes of predictor functions to be predominant. Phase components may also carry important information about the responses and one should not always ignore them. However, in some cases, as illustrated through examples presented later in this paper, shapes are the primary predictors and one wants regression models that can exploit this knowledge.

1.2. Potential solutions

This leads us to an important question: *What kind of regression models allow inclusion of only the shape or amplitude of the predictor functions and deemphasize their phases*? In general, there are some parametric and nonparametric choices available.

1. **Pre-Aligned Functional Linear Model (PAFLM)**: One obvious solution is to simply remove the phase variability in the given functions $\{f_i\}$ using one of several pre-existing functional alignment algorithms (see e.g. Ramsay and Li, 1998; Liu and Müller, 2004; Srivastava et al., 2011; Tucker et al., 2013). Then, one can use the aligned functions, or amplitudes, for predicting the response variable using previously-mentioned FLM. The alignment algorithms

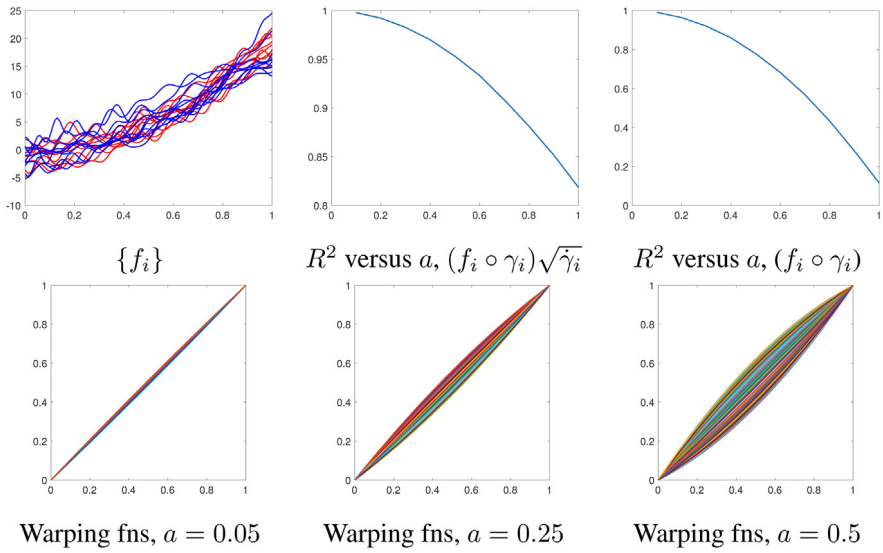


Fig. 2. Experiments show superlinear decrease in R^2 prediction measure as the amount of phase variability is increased in predictor functions.

are typically based on matching the given $\{f_i\}$ one-by-one to a template function which, in turn, is constructed iteratively using the means of the aligned functions. The limitation of this approach, in a regression setting, is that this alignment is performed independent of the response variable. In other words, the values $\{y_i\}$ do not play any role in the alignment.

2. **Joint Modeling & Alignment Under Value-Preserving Warping Using the $\mathbb{L}^2$ Inner-Product**: Another possibility is to remove the phase within FLM by introducing an extra step. For instance, when using the contaminated predictors $\{\tilde{f}_i = f_i \circ \gamma_i\}$, under the value-preserving warping, one can try to solve for the unknown warpings by adding optimization over γ_is, as follows. We can modify the model in Eq. (1) to become:

$$y_i = \alpha + \sup_{\gamma_i \in \Gamma} \left(\int_0^T \tilde{f}_i(\gamma_i(t))\beta(t)\,dt \right) + \epsilon_i, \quad i = 1, \ldots, n\,. \tag{3}$$

This additional optimization over Γ is supposed to nullify the original contamination in f_is. However, this approach as specified has a major shortcoming. As described in several places, see e.g. Marron et al. (2014) and Srivastava and Klassen (2016), the optimization over γ_i under the $\mathbb{L}^2$ inner product is actually degenerate, due to a phenomenon called the *pinching effect*. Some authors minimize pinching by restricting the set of warpings in Eq. (3) in a predetermined manner. This restriction is unnatural as it is impossible to predict the optimal set of warpings needed to align future data.

3. **Nonparametric Regression Model**: Nonparametric models for functional regression are gaining popularity since they do not require any predetermined model and are purely data driven. Developed and studied by Ferraty and Vieu (2006), a nonparametric model for functional regression is given by: $y_i = G(f_i) + \epsilon_i$. Here $G : \mathcal{F} \to \mathbb{R}$ is an unknown smooth map, estimated by the functional Nadaraya–Watson (NW) estimator (Nadaraya, 1964). For the given data $\{(f_i, y_i),\ i = 1, 2, \ldots, n\}$, the estimator is given by:

$$\hat{G}(f) = \frac{\sum_{i=1}^n y_i K(d(f_i, f)/b)}{\sum_{i=1}^n K(d(f_i, f)/b)}, \tag{4}$$

where K is the standard Gaussian kernel, b is the bandwidth parameter, and d is a distance on the predictor (function) space. Naturally, the choice of distance d is critically important in such kernel estimators. If we use the standard $\mathbb{L}^2$ norm in $\mathcal{F}$ for d, then the prediction will remain dependent on the phase of the predictors. Instead, if we choose a distance that compares *shapes of the predictors* and ignores their phases, *i.e.* d is a proper shape metric, then the model becomes invariant to phase.

1.3. Proposed approach

There is possibility of a different parametric approach that stems from modifying the main term in FLM (Eq. (1)) directly, and making it invariant to the phase. This approach is motivated by the use of invariant metrics, such as the Fisher–Rao metric and the elastic Riemannian metric in FDA (Srivastava et al., 2011; Srivastava and Klassen, 2016). In fact, depending on the chosen warping, this elastic FDA framework gives several ideas although only a couple of them

are discussed here. This framework is based on replacing the $\mathbb{L}^2$ inner product and the $\mathbb{L}^2$ distance in FDA by invariant Riemannian metrics and invariant distances between functions. The invariant quantities provide better mathematical and numerical properties, and indeed lead to a superior performance in FDA. The challenge in using these invariant metrics comes from their complicated expressions, but that is overcome using square root velocity function (SRVF) representation (Srivastava et al. (2011)). The SRVF of a function f is defined by: $q(t) = \text{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$. One works with the SRVFs q_is instead of the predictors f_is and the Fisher–Rao metric simplifies to the standard $\mathbb{L}^2$ metric under this change of variables. This framework motivates at least two ways of fixing the pinching problem in Eq. (3):

1. **Use SRVF Representation and Value-Preserving Warping**: The first idea is to compute SRVFs of the given predictors, and then simply replace the term $\sup_{\gamma_i} \langle f_i \circ \gamma_i, \beta \rangle$ in Eq. (3) by the term: $\sup_{\gamma_i} \langle (q_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}, \beta \rangle$. This is motivated by the fact that under Fisher–Rao invariant metric, the inner product between functions is exactly equal to the $\mathbb{L}^2$ inner product of their SRVFs.
   The corresponding time warpings of SRVFs, q_is, are given by $(q_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}$, and are norm preserving. That is, $\|q_i\| = \|(q_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}\|$ for all $q_i \in \mathbb{L}^2$ and $\gamma_i \in \Gamma$, and thus pinching is no longer possible. More importantly, the model is now completely independent of the phase components of the predictors f_is.
2. **Use Original Functions and Norm-Preserving Warping**: The other option is to incorporate the norm-preserving transformations of the functions themselves $(f_i \mapsto (f_i \circ \gamma_i)\sqrt{\dot{\gamma}_i})$ in the model, without resorting to SRVFs. As noted earlier, this warping changes both the locations and the heights of peaks and valleys in function, but preserves its $\mathbb{L}^2$ norm. In this case we replace the problematic $\mathbb{L}^2$ inner-product term in Eq. (3) by the term $\sup_{\gamma_i} \langle (f_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}, \beta \rangle$. This option is especially suitable when f_is are noisy and an SRVF transformation may further enhance this noise. By working with f_is, one inherits all the nice properties of the Fisher–Rao framework and avoids enhancing the noise. However, this warping is different from the value-preserving warping $f \circ \gamma_i$ used in traditional functional alignment. Thus, γ_is here can be called *phase* in a broader sense but not in a classical sense. In the end, the regression model is invariant to the phase of the predictors, except the phase is now defined using the mapping $f_i \mapsto (f_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}$.

Each of these models help remove the phase variability, avoid the pinching effect, and improve prediction performance. Ultimately, the choice of a model depends on the nature of the data and the goals of the application. The response variables in both these models are invariant to the respective warpings of the predictor functions. In this paper, we will develop the second approach and will call this the *elastic functional regression* (EFRM) model.

The rest of this paper is as follows. In Section 2, we develop the proposed elastic functional regression model and discuss estimation of model parameters. We demonstrate this model using some simulated data and real data, and compare its performance against some current ideas in Section 3. Lastly, Section 4 ends the paper with some concluding remarks.

2. Elastic functional regression model (EFRM)

In this section, we layout a regression model for *scalar-on-function* problem with the property that the response variable is invariant to the phase component of the predictor. This framework is based on ideas used previously for alignment of functional data, or phase–amplitude separation, using invariant metrics and the SRVF representation of functions. We start by briefly introducing those concepts and refer the reader to Srivastava et al. (2011) for additional details.

2.1. Model specification

As mentioned earlier, the use of $\mathbb{L}^2$ inner-product or $\mathbb{L}^2$ norm for alignment of functions leads to a well-known problem called the *pinching effect*. While some papers avoid this problem using a combination of external penalties and search space reductions, a more comprehensive solution comes from using an elastic Riemannian metric with appropriate invariance properties. This metric, called the *Fisher–Rao metric* for functions, avoids the pinching effect without any external constraints and results in superior alignment results. Let f be a real-valued function on the interval $[0, 1]$ (with appropriate smoothness) and let $\mathcal{F}$ denote the set of all such functions. Let Γ be the set of all boundary preserving diffeomorphisms of the unit interval $[0, 1]$, *i.e.* $\Gamma = \{\gamma : [0, 1] \to [0, 1] \mid \gamma(0) = 0, \gamma(1) = 1, \gamma \text{ is a diffeomorphism}\}$. For the purpose of alignment, one represents a function f using its square-root velocity function (SRVF): $q(t) = \text{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$. One of the advantages of using SRVF is that under the transformation $f \mapsto q$, the complicated Fisher–Rao Riemannian metric and the Fisher–Rao distance map into much simpler expressions ($\mathbb{L}^2$ inner product and $\mathbb{L}^2$ norm, respectively). If we warp a function f by a time warping γ, *i.e.*, map $f \mapsto (f \circ \gamma)$, then its SRVF changes by $q \mapsto (q \circ \gamma)\sqrt{\dot{\gamma}}$. The latter quantity will be denoted by $(q * \gamma)$. The invariance property of the Fisher–Rao metric implies that for any $q_1, q_2 \in \mathbb{L}^2$ and $\gamma \in \Gamma$, we have: $\|(q_1 * \gamma) - (q_2 * \gamma)\| = \|q_1 - q_2\|$. In other words, the action of Γ on $\mathbb{L}^2$ is by isometries. A special case of this equation is that $\|(q * \gamma)\| = \|q\|$ for all q and γ. Thus, this action preserves the $\mathbb{L}^2$ norm of the SRVF and, therefore, avoids any pinching effect.

This framework motivates several solutions for avoiding the pinching problem associated with the inner-product term in Eq. (3). While one can work with the SRVFs of the given predictor functions, they are prone to noise in the original data due to the involvement of a time derivative in the definition of SRVF. In case the original data is noisy, this noise

gets enhanced by taking a derivative. As a workaround to this problem, we treat the given predictor functions to be in the SRVF space already. That is, we assume the action of warping γ_i on an f_is is given by $(f_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}$ and not $f_i \circ \gamma_i$. With this action, we have that $\|(f_i * \gamma_i)\| = \|(f_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}\| = \|f_i\|$.

Based on this argument, the inner-product term in Eq. (3) can be replaced by the term: $\sup_{\gamma_i \in \Gamma} \langle \beta, (f_i * \gamma_i) \rangle$. This is a scalar quantity and represents a modified linear relationship between the predictor and the response. One can impose a single-index model on top of this construction to generalize this model. Such single-index models have been used commonly in conjunction with FLMs, see e.g. Stoker (1986), Ait-Saïdi et al. (2008), Reiss et al. (2017), Eilers and Marx (1996) and Jiang and Wang (2011). For any $h : \mathbb{R} \to \mathbb{R}$, a smooth function, define the EFRM model as:

$$y_i = h\Big(\sup_{\gamma_i \in \Gamma} \langle \beta, (f_i * \gamma_i) \rangle\Big) + \epsilon_i, \, i = 1, \ldots, n \tag{5}$$

To complete model specification, we assume ϵ_is to be *i.i.d* zero-mean, Gaussian random variables. This model has the following properties.

1. **Nonlinear Relationships**: There are two sources of nonlinearity in the relationship between f_i and y_i. Although the inner product $\langle \beta, f_i \rangle$ is linear in f_i, the supremum over Γ makes the term $\sup_{\gamma_i \in \Gamma} \langle \beta, (f_i * \gamma_i) \rangle$ nonlinear. Furthermore, the inclusion of h allows EFRM to makes relationship firmly nonlinear.
2. **Invariance to Phase**: For a fixed model description (β, h), the mean of response y_i is invariant to the phase of f_i due to the fact that $\sup_{\gamma_i} \langle \beta, (f_i * \gamma_i) \rangle = \sup_{\gamma_i} \langle \beta, ((f_i * \gamma_0) * \gamma_i) \rangle$, for all $\gamma_0 \in \Gamma$. Even though the mean of y_i is invariant to the phase, we note that the estimated values of β and h (covered in the next section) can depend on the phase of f_i.
3. **Identifiability of** β: In view of the equality mentioned in the previous item, the regression coefficient β is not fully specified. This is because if $\hat{\beta}$ is an estimator of β, then so is $\hat{\beta} \circ \gamma_0$ for any $\gamma_0 \in \Gamma$. To avoid this ambiguity, we impose an additional constraint on the model that all the maximizers $\{\hat{\gamma}_i = \arg\sup_{\gamma_i} \langle \beta, (f_i * \gamma_i) \rangle\}$ together satisfy the condition that $\frac{1}{n} \sum_{i=1}^{n} \hat{\gamma}_i = \gamma_{id}$.
4. **Difference from GFLM**: The single-index model used here is quite similar to a generalized FLM (GFLM), but with an important difference. In a single-index model, the index function h is unknown and needs to be estimated from the data itself, while in generalized model h is assumed known. One can easily switch from EFRM to GFLM, if needed, by using a known h.

2.2. Parameter estimation

Next we consider the problem of estimating EFRM parameters using MLE. The unknown parameters are: the index function h and the coefficient of regression β. We take an iterative approach, laid out in Eilers et al. (2009), where one updates estimates of h or β while keeping the other fixed. Thus, we first focus on techniques for estimating these quantities separately.

Estimation of β keeping h fixed. : Given a set of observations $\{(f_i, y_i)\}$, the goal here is to solve for MLE of β, while keeping h fixed. In order to reduce the search space to a finite-dimensional set, we will assume that $\beta \in \{\sum_{j=1}^{J} c_j b_j | c_j \in \mathbb{R}\}$ for a fixed orthonormal basis $\mathcal{B} = \{b_j, j = 1, 2, \ldots\}$ of $\mathbb{L}^2([0, 1], \mathbb{R})$. The estimation problem is now given by:

$$\hat{c} = \underset{c \in \mathbb{R}^J}{\operatorname{argmin}} \;\; H(c), \quad \text{where} \;\; H : \mathbb{R}^J \to \mathbb{R}, \;\; \text{given by}$$

$$H(c) = \left(\sum_{i=1}^{n} (y_i - h(\sup_{\gamma_i \in \Gamma} \left\langle \sum_{j=1}^{J} c_j b_j, (f_i * \gamma_i) \right\rangle))^2 \right) .$$

We use a MATLAB function `fminunc`, which in turn uses the quasi-Newton method, to solve the minimization problem. Contained within this problem are a set of optimizations over γ_is. For a fixed c, this optimization is performed using the dynamic programming algorithm (DPA) for each $i = 1, 2, \ldots, n$. This set of calls to DPA are inside the definition of H and are performed for each candidate value of c. Thus, any update of c requires recomputing the optimal warping functions, resulting in an iterative process. Finally, once c (or β) is estimated, we can impose the condition for specification of β, *i.e.* $\frac{1}{n} \sum_{i=1}^{n} \hat{\gamma}_i = \gamma_{id}$ as follows. For this, we use the current $\hat{\gamma}_i$s to compute their average $\bar{\gamma} = \frac{1}{n} \sum_{i=1}^{n} \hat{\gamma}_i$ and replace β by $\beta \circ \bar{\gamma}$. The full process for estimating β is summarized in Algorithm 1.

To analyze this estimator, one has to study the choice of J relative to the sample size n, and develop an asymptotic theory. Since this analysis is very similar to existing papers involving functional predictors (Li et al., 2010; Morris, 2015), we simply refer to that literature for asymptotic analysis.

Estimation of h keeping β fixed. Next we consider the problem of estimating the index function h given the data and the current estimate of β. The reason for introducing this single-index model is to capture nonlinear relationship between the predicted responses and observed responses. While there are many potential nonparametric estimators for h, we keep the model simple by restricting to lower-order polynomials. We allow h to be only linear, quadratic, and cubic: $h(x) = ax + b$, $h(x) = ax^2 + bx + c$, and $h(x) = ax^3 + bx^2 + cx + d$, etc.

Algorithm 1 Estimation of β keeping h fixed

1: Initialization Step. Choose an initial $c \in \mathbb{R}^J$ and compute $\hat{\beta}(t) = \sum_{j=1}^{J} c_j b_j(t)$.
2: Use an optimization method (such as `fminunc` in MATLAB) to find $\hat{c}$ that minimizes the cost function H.

- To define H, use the current $\hat{c}$ (and $\hat{\beta}$) to perform the following for each $i = 1, 2, \ldots, n$,
  - Solve for $\hat{\gamma}_i = \text{argmin}_{\gamma \in \Gamma} \|\hat{\beta} - (f_i * \gamma_i)\|^2$, using the Dynamic Programming algorithm (DPA).
  - Compute the aligned functions $\tilde{f}_i \leftarrow (f_i * \gamma_i) \equiv (f_i \circ \hat{\gamma}_i)\sqrt{\dot{\hat{\gamma}}_i}$.

3: Update $\hat{\beta}(t) = \sum_{j=1}^{J} \hat{c}_j b_j(t)$. If the $|H(\hat{c})|$ is large, then return to step 2.
4: Compute $\bar{\gamma} = \frac{1}{n}\sum_{i=1}^{n} \hat{\gamma}_i$ and replace β by $\beta \circ \bar{\gamma}$.

For estimating h, we first predict responses according to: $\hat{y}_i = \sup_{\gamma_i \in \Gamma} \left\langle \hat{\beta}, (f_i * \gamma_i) \right\rangle$, and then we fit a polynomial function h between the predicted responses $\hat{y}_i$ and the observed responses y_i using the least squares error criterion. The full parameter estimation procedure is presented in Algorithm 2.

Algorithm 2 Elastic Functional Regression Model

1: Initialize h as the identity function ($h(x) = x$).
2: Given h, use Algorithm 1 to estimate $\hat{\beta}$.
3: For a given $\hat{\beta}$, update h using the least squares criterion.
4: If $|H(\hat{c})|$ is small, then stop. Else, return to step 2.

2.3. Prediction of response under the elastic regression model

One of the goals of EFRM is to predict values of the response variable for the future predictor observations. Here we describe the prediction process under EFRM. As the model suggests, this prediction is based on alignment of predictors to the coefficient $\hat{\beta} = \sum_{j=1}^{J} \hat{c}_j b_j$ using DPA. For a given predictor $f^{(test)}$, the predicted value of y is:

$$\hat{y}^{(test)} = \hat{h}\left(\sup_{\gamma_i \in \Gamma} \left\langle \sum_{j=1}^{J} \hat{c}_j b_j, (f^{(test)} * \gamma_i) \right\rangle\right). \tag{6}$$

We will use this predictor to evaluate prediction performance of EFRM, relative to current models, using both simulated data and real data.

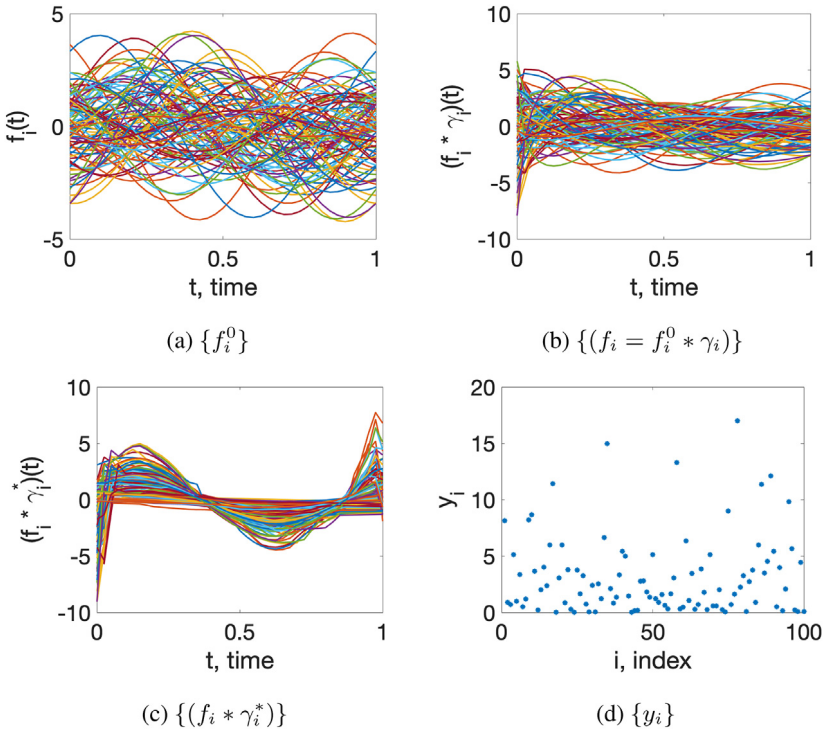
3. Experimental illustration

We will compare EFRM with four natural alternatives. Either these models are commonly used in the literature or they are simple modifications of the current models for handling the phase variability in the predictors. These models are: Functional Linear Model (FLM); Pre-Aligned Functional Linear Model (PAFLM); Nonparametric regression model (NP) using a Gaussian kernel function and two different choices of d. We briefly summarize and introduce these models.

Functional linear model (FLM). FLM has already been introduced in Eq. (1). As stated earlier, it does not specifically account for the presence of phase variability in the predictor data and is vulnerable to that nuisance variability.

Pre-aligned functional linear model (PAFLM). As the name suggests, PAFLM is where one pre-aligns the predictor functions (using a phase–amplitude separation algorithm) and then performs standard FLM. To clarify further, one performs phase–amplitude separation and then discards the phase component. In the results presented here, we use the *"Complete Alignment Algorithm"* presented in Srivastava et al. (2011). This alignment is suboptimal from the perspective of regression, since the response variable is not used in the alignment.

Nonparametric kernel approach. As mentioned earlier, one can use the Nadaraya–Watson estimator (of the kind given in Eq. (4)) for predicting y for a new predictor function f. The only quantity left unspecified in that equation is the metric structure on $\mathcal{F}$. In the following we choose the distance to be either the $\mathbb{L}^2$ norm or a weight shape distance. The weighted shape distance uses a pre-alignment of predictor functions and is defined as follows. Let the predictors $\{f_i\}$ be pre-aligned (as discussed above) resulting the phases $\{\hat{\gamma}_i\}$ and amplitude $\{f_i * \hat{\gamma}_i\}$. Then, define the distance $d(f, f_i) = \lambda d_a(f, f_i) + (1 - \lambda) d_p(f, f_i)$, where $\lambda \in [0, 1]$ is a proportion parameter. Here d_a denotes the amplitude distance:



(a) $\{f_i^0\}$ (b) $\{(f_i = f_i^0 * \gamma_i)\}$

(c) $\{(f_i * \gamma_i^*)\}$ (d) $\{y_i\}$

Fig. 3. Simulated data 1. (a) shows the original functions, $\{f_i^0\}$, (b) shows them after random warpings, $\{f_i\}$, (c) shows predictors after optimizations over γ_i in the generative model in Eq. (5), $\{f_i * \gamma_i^*\}$, and (d) displays ordered response variables, $\{y_i\}$, from that model.

$d_a(f, f_i) = \|f - (f_i * \hat{\gamma}_i)\|$ and d_p denotes the phase distance: $d_p(f, f_i) = \|\sqrt{\dot{\hat{\gamma}}_i} - \sqrt{\dot{\gamma}_{id}}\|$. The optimal value of the bandwidth $\hat{b}$ can be obtained via cross-validation:

$$\hat{b} = \underset{b \in \mathbb{R}_+}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - G_{(-i)}(f_i))^2, \quad \text{with} \quad G_{(-i)}(f) = \frac{\sum_{j=1, j\neq i}^{n} y_j K((d(f_j, f))/b)}{\sum_{j=1, j\neq i}^{n} K((d(f_j, f))/b)}$$

For the joint estimation of λ and b, we first compute the optimal bandwidth $\hat{b}$ for each $\lambda \in [0, 1]$. Then, we choose the optimal $\hat{\lambda}$ which gives the lowest cross-validation error.

Next, we present experimental results from these and EFRM on a number of data sets.

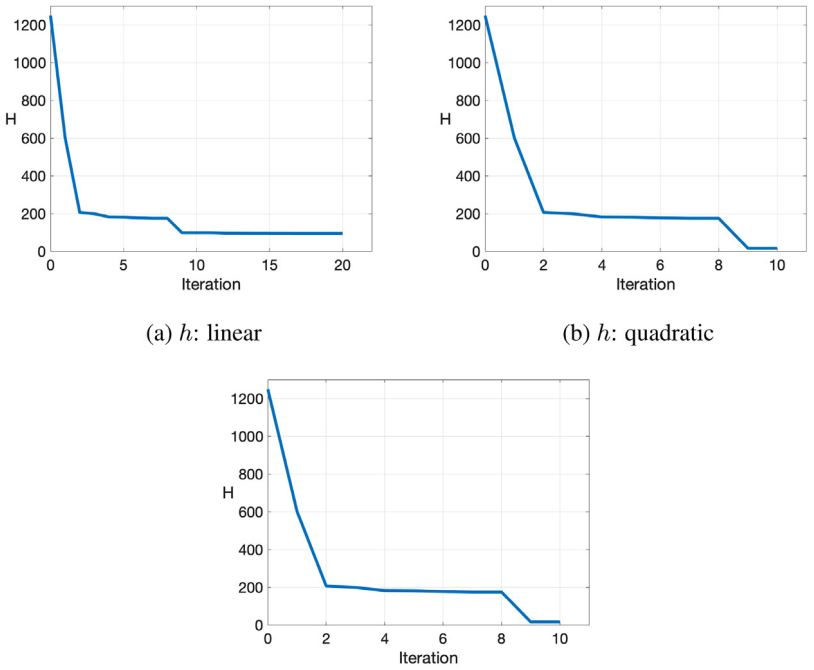
3.1. Simulation study

In the studies presented in this section, we perform a five-fold cross-validation and compute the mean and standard deviation of root mean square error (RMSE) for predicting the response variable. We use this RMSE for comparing performances of different regression models.

3.1.1. Simulated data 1

In the first experiment, we simulate $n = 100$ observations using the model stated in Eq. (5). For the predictors, we use a Fourier basis and random coefficients to form the functions, $f_i^0(t) = c_{i,1}\sqrt{2}sin(2\pi t) + c_{i,2}\sqrt{2}cos(2\pi t)$ with $c_{i,1}, c_{i,2} \sim N(0, 1^2)$. Then we perturb them using random $\{\gamma_i\}$ to obtain the predictors $\{f_i = (f_i^0 * \gamma_i)\}$. We also simulate the coefficient function β using the same Fourier basis but with a fixed coefficient vector $c_0 = [1, 1]$. We plug these quantities in the model, use a quadratic polynomial for h, and add independent observation noise, $\epsilon_i \sim N(0, 0.01^2)$, to obtain responses $\{y_i\}$. This process is illustrated in Fig. 3. We use a random 80–20 split for training and testing, respectively.

Model estimation. Using the training data, we estimate the model parameters h and β, as described in Algorithm 2. In order to evaluate this algorithm, we use three different bases for estimating β during training: (1) Fourier basis with only two elements, (2) Fourier basis with four elements, and (3) B-spline basis with four elements. The reason for using different bases for estimation is to study the effects of basis on the model performance. We also try three different polynomials: linear, quadratic, and cubic, as h during estimation.



(a) h: linear (b) h: quadratic

(c) h: cubic

Fig. 4. The evolution of cost H for each choice of the index function, h, and using Fourier basis with two elements for β.

Table 1
The average of $\text{RSE}_{\mathbb{L}^2}$ (Root Squared Error) of $\hat{\beta}$ and $\hat{h}$ for different choices of parameter sets on simulated data 1.

Basis	Fourier2		Fourier4		Bspline4	
Parameter	β	h	β	h	β	h
h: Linear	2.326	1.372	**2.726**	1.077	9.607	1.075
h: Quadratic	**2.268**	0.284	2.862	0.247	9.914	0.278
h: Cubic	2.288	**0.283**	2.777	**0.231**	**8.803**	**0.276**

Fig. 4 shows the evolution of cost function H during optimization in Algorithm 2 for each of index functions: linear, quadratic, and cubic, in Figs. 4a, 4b, and 4c, respectively. These experiments use a Fourier basis with two elements to estimate β. These plots show that the cost H goes down in all cases and the optimization algorithm provides at least local solutions reliably. The optimized values are found to be the best for the quadratic and cubic $\hat{h}$, which makes sense since a quadratic h was used to simulate the data.

It is also important to quantify estimation performance for model parameters β and h. In order to quantify these errors, we calculate the Root Squared Error $\text{RSE}_{\mathbb{L}^2} = \sqrt{\int [a(t) - \hat{a}(t)]^2 \, dt}$, where $a(t)$ is a functional parameter and $\hat{a}(t)$ is its estimate (for $a = \beta,\ h$). We then compute the averages of $\text{RSE}_{\mathbb{L}^2}$ over a five-fold cross-validation. The estimation errors for β and h for this simulation experiment are presented in Table 1. Overall, the choice of a cubic h does well in the estimation. If we compare these RSEs with prediction performances in Table 2, we see that a better estimation of β and h provides a better predictor of the response variable, which is natural.

Prediction performance. To evaluate prediction performance, we use the model parameters estimated using the training step for predicting the response variable for the test data. This prediction follows the procedure laid out in Eq. (6). The predicted responses are then compared with the true responses to quantity the prediction error. We perform five-fold cross-validation to evaluate this error more precisely. Then we compute the average and the standard deviation of RMSE ($\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$) from five different folds and use these quantities to compare different models.

The results for average five-fold RMSEs and corresponding standard deviations are shown in Table 2. As these results show, EFRM is able to provide a better prediction performance than the competing models despite using very simple tools. The predictions from PAFLM are less accurate since this method pre-aligns functional predictors without considering response variables $\{y_i\}$. The nonparametric regression model using the $\mathbb{L}^2$ norm shows some improvement in prediction, when compared to FLM and PAFLM, since it is not restricted to linear relationships between the response and functional

Table 2
The average and the standard deviation (in parentheses) of RMSEs for three model-based methods on simulated test data. The true values are Fourier2 basis and a quadratic h.

Parametric			
Basis	Fourier2	Fourier4	Bspline4
h: Linear	1.140 (0.130)	1.109 (0.257)	1.604 (0.270)
h: Quadratic	0.527 (0.308)	0.599 (0.213)	1.509 (0.412)
h: Cubic	**0.520 (0.299)**	**0.564 (0.179)**	**1.477 (0.406)**
FLM	2.765 (0.458)	2.855 (0.440)	2.858 (0.399)
PAFLM	5.021 (4.415)	5.741 (5.383)	5.084 (4.703)
Nonparametric			
NP-$\mathbb{L}^2$		1.652 (0.275)	
NP-shape		1.960 (0.368)	

Table 3
The average and the standard deviation (in parentheses) of the five RMSE's for three model-based methods on simulated test data.

	Model	RMSE
Parametric	h: Linear	5.984 (2.670)
	h: Quadratic	4.548 (1.703)
	h: Cubic	**4.379 (1.876)**
	FLM	7.698 (1.746)
	PAFLM	36.540 (9.932)
Nonparametric	NP-$\mathbb{L}^2$	8.969 (1.691)
	NP-shape	10.030 (1.424)

predictors. However, this model also fail to account for the phase variations and the predictions are found to be less accurate than EFRM.

3.1.2. Simulated data 2

In the second experiment, we again simulate $n = 100$ observations using the model stated in Eq. (5), but this time we use a B-spline basis with 20 elements and random coefficients to form the predictor functions. As earlier, we simulate the coefficient function β using the same basis and a fixed coefficient vector. Then we plug these quantities in the model, use a quadratic polynomial function h, and add independent observation noise, $\epsilon_i \sim N(0, 0.01^2)$, to obtain the responses $\{y_i\}$. Skipping further details, we focus directly on prediction performance (using the same B-spline basis with 20 elements).

Prediction performance. The prediction results are shown in Table 3. Despite increased complexity of predictors, resulting from a larger basis set, EFRM still performs better relative to the competing methods.

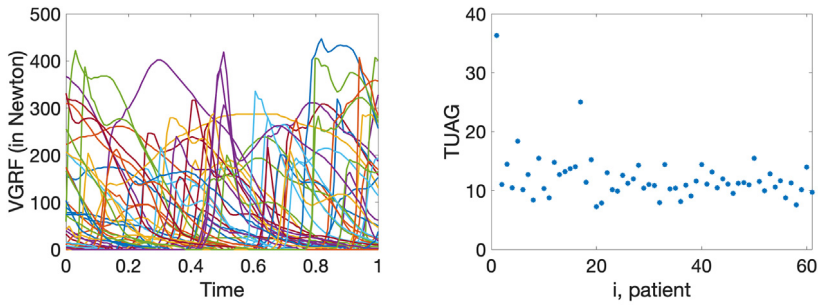
A part of the success of EFRM can be attributed to the fact that the data was indeed simulated from that model itself. Therefore, it is natural that this model does better than others. However, these experiments also point to the robustness of the response variable to random phase variability in the functional predictors. Technically, the response is invariant to this phase variability. Additionally, the model benefits from optimization over Γ alongside the estimation of β and h. In this way, the model chooses phases in a way that helps maximize prediction performance.

3.2. Application to real data

Next, we apply EFRM to three real data examples. There are several important application areas where functional variables form predictors. Examples include biology, human anatomy, biochemistry, finance, epidemiology, and so on. We take three representative examples from human biometrics, chemistry and stock market. The goal in each case is to use shapes of functional predictors in predicting scalar response variables.

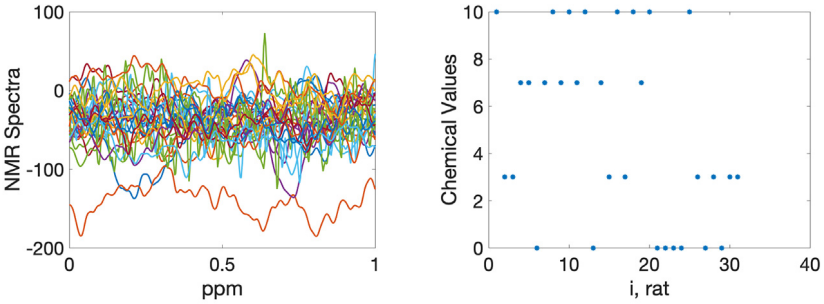
Description of the data.

1. **Gait in Parkinson's Disease Data**: First, we use Gait data collected for diagnosing Parkinson's disease, taken from the well-known *Physionet* (Goldberger et al., 2000) database. The database contains *Vertical Ground Reaction Force* (VGRF) records of subjects as they walk at their usual, self-selected pace for approximately two minutes on level ground. A total of eight sensors are placed underneath each foot for measuring forces (in Newtons) as functions of time. The outputs of these 16 sensors (left: 8 and right: 8) are digitized and recorded at 100 samples per second. From the original data, we extract very short segments (the first $1-100$ time points from total 12119) for simplicity and efficiency of computation. Based on demographic information, each patient has his/her own *Timed Up And Go* (TUAG) test which is a simple test used to assess a patient's mobility and requires both static and dynamic balance.



(a) VGRF, $\{f_i\}$ (b) TUAG, $\{y_i\}$

Fig. 5. Gait in Parkinson's disease data.



(a) 31 NMR spectra, $\{f_i\}$ (b) chemical values, $\{y_i\}$

Fig. 6. Metabonomic 1H-NMR data.

We consider VGRF records as predictors and TUAG as scalar responses, with each subject forming an independent observation.

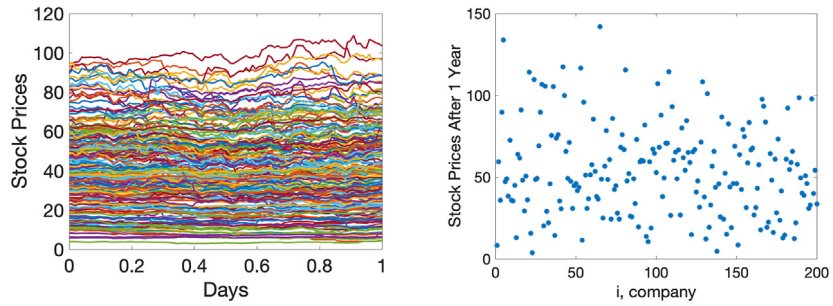
There are three groups of patients in Gait in Parkinson's disease data. We focus on two groups named *"Ga"* and *"Si"* (Frenkel-Toledo et al., 2005b,a; Yogev et al., 2005) in the dataset to ensure the same demographic information among the participants. This results in a total of 61 functions or curves for the analysis.

Fig. 5 plots segments of VGRF for each of the 61 patients in the left panel and TUAG values in the right panel. In the experiments presented later, we randomly select 40 as training and the rest 21 as test.

2. **Metabonomic 1H-NMR Data**: Metabonomic 1H-NMR (Nuclear Magnetic Resonance) data (Winning et al., 2009) originates from 1H NMR analysts of urine from thirty-two rats, fed a diet containing an onion by-product. The aim is to evaluate the *in vivo* metabolome following the intake of onion by-products. The data set contains 31 NMR spectra in the region between $(0 - 3000)$ ppm as predictors and some reference chemical values as responses.
   Since we have 31 total observations, we randomly select 21 curves as the training set and rest 10 curves as the test set. Similar to the Gait in Parkinson's disease data, we extract the first 300 time points from 29001 time points for efficient computation and statistical analysis. Fig. 6 displays the plots of NMR spectra of 31 rats (left panel) and the chemical values which are considered as response variable (right panel).
3. **Historical Stock Data**: *QuantQuote* posts large amounts of free historical stock data on their website for free download. There are total of 200 companies and each company has total 3926 stock entries during the interval 1/2/1998 to 8/9/2013. For each company, we collected stock prices from 3/20/2012 to 8/9/2012 to form functional predictors. Thus, there are 100 daily time points over the selected interval forming predictor functions. We take the stock prices on 8/9/2013, which is exactly one year after the end of predictor interval, as the scalar response variable. Our goal is to predict one-year future stock price for each company based on historical stock prices.
   Fig. 7 shows an example of this stock data. The 200 functional predictors are shown in Fig. 7a and scalar response variables are shown in Fig. 7b. We use first 140 curves to fit the model and remaining 60 curves as test.

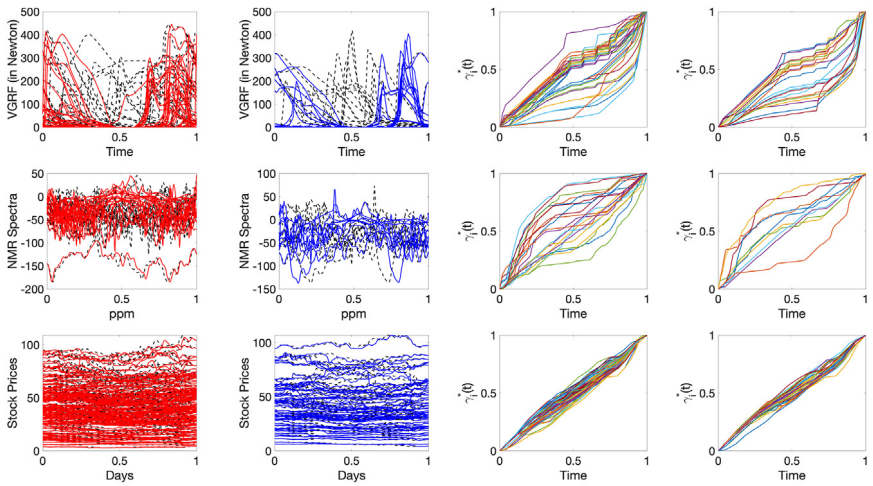
Analysis of real data. For representing the coefficient function β, we use a B-spline basis with 20 elements and estimate parameters using Algorithm 2.

Fig. 8 shows "aligned" functional predictors during training and testing. Each row corresponds to a real data set — gait in Parkinson's disease (first row), metabonomic 1H-NMR (second row), and historical stock market (third row). The



(a) Past Stock Prices, $\{f_i\}$ (b) One-year future value $\{y_i\}$

Fig. 7. Stock data.



(a) Training (b) Test (c) Training (d) Test

Fig. 8. $\{f_i\}$ vs. warped $\{f_i\}$ and $\{\gamma_i^*\}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

original functions are drawn in black dashed curves and the warped functions are overlaid using the red/blue solid colors. Figs. 8a and 8b show the curves for the training data and the test data, respectively. The corresponding optimal warpings for training and test are shown in Figs. 8c and 8d, respectively. We remind the readers that the predictors have been warped using the norm-preserving action during the optimization step. They appear more aligned than before but are not as aligned as one would get from a pure alignment procedure. This alignment results in an increased ability of the model to predict the response variable. Thus, this warping is more to help regress the responses y_is to the predictors f_is, rather than to align peaks and valleys in f_is.

Prediction results. Table 4 presents prediction RMSE for different models studied in this experiment. It shows that EFRM model outperforms other models on all three datasets. In the case of 1H-NMR data, EFRM using a cubic index function does the best, while in other cases lower order polynomials perform better. This could be because the response variable in NMR example is categorical with four values and one needs a cubic polynomial to fit these response levels. Predictions from the kernel regression model are close second to EFRM.

4. Concluding remarks

The development of functional regression models that can handle phase variability in functional predictors is a challenging problem in FDA. We have proposed a new elastic approach that uses the shapes of functions, rather than the full functions, as predictors in regression models. The notion of shape is based on a norm-preserving warping of the

Table 4
Prediction RMSE for predicted response variable under each model.

Model	Gait	1H-NMR	Stock
h : Linear	2.741	4.849	**9.007**
h : Quadratic	**2.466**	4.106	9.130
h : Cubic	2.594	**4.025**	9.227
FLM	7.483	213.490	10.405
PAFLM	19.158	202.941	11.086
NP-$\mathbb{L}^2$	6.559	4.251	9.795
NP-shape	2.625	4.251	9.540

Table 5
RMSEs of using SRVF representation and value-preserving warping.

Model	h: Linear	h: Quadratic	h: Cubic
RMSE	18.032	18.139	18.110

predictors and handles the nuisance phase variability by optimizing the $\mathbb{L}^2$ inner product over the warping group inside the model. We compare the prediction RMSE of the model with several existing methods, to demonstrate effectiveness of this technique in both simulated data and real data.

As discussed in Section 1.3, there is another model that can potentially eliminate the effects of phase variability in the predictor functional data. This model involves SRVFs $\{q_i\}$ of the predictors and uses the term $\sup_{\gamma_i} \langle (q_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}, \beta \rangle$ as the argument of the index function h. However, we have not pursued this model because, despite theoretical advantages, the practical performances of this model are sometimes low. As an example, we study the prediction problem using the same stock market data as in item 3 of Section 3.2. The prediction RMSE for this model is listed in Table 5, and is found to be worse than the results shown in Table 4. We conjecture that it is because the noise in predictor data gets enhanced when computing SRVFs (due to the presence of a time derivative in SRVF expression). Thus, we prefer the second option mentioned in Section 1.3 for EFRM.

Availability

MATLAB programs are available on Github: https://github.com/fdastat/elastic-regression

Acknowledgments

This research was supported in part by the National Science Foundation (NSF), USA grants NSF-1621787 and NSF-1617397 to AS. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

Ahn, K., Tucker, J.D., Wu, W., Srivastava, A., 2018. Elastic handling of predictor phase in functional regression models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 324–331. http://dx.doi.org/10.1109/cvprw.2018.00072.

Ait-Saïdi, A., Ferraty, F., Kassa, R., Vieu, P., 2008. Cross-validated estimations in the single-functional index model. Statistics 42 (6), 475–494. http://dx.doi.org/10.1080/02331880801980377.

Cai, T.T., Hall, P., 2006. Prediction in functional linear regression. Ann. Statist. 34 (5), 2159–2179. http://dx.doi.org/10.1214/009053606000000830.

Cardot, H., Ferraty, F., Sarda, P., 1999. Functional linear model. Statist. Probab. Lett. 45 (1), 11–22. http://dx.doi.org/10.1016/s0167-7152(99)00036-x.

Ciarleglio, A., Ogden, R.T., 2016. Wavelet-based scalar-on-function finite mixture regression models. Comput. Statist. Data Anal. 93, 86–96. http://dx.doi.org/10.1016/j.csda.2014.11.017.

Eilers, P.H.C., Li, B., Marx, B.D., 2009. Multivariate calibration with single-index signal regression. Chemometr. Intell. Lab. Syst. 96 (2), 196–202. http://dx.doi.org/10.1016/j.chemolab.2009.02.001.

Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with b-splines and penalties. Statist. Sci. 89–102. http://dx.doi.org/10.1214/ss/1038425655.

Febrero-Bande, M., Oviedo de la Fuente, M., 2012. Statistical computing in functional data analysis: the R package fda. usc. J. Stat. Softw. 51 (4), 1–28. http://dx.doi.org/10.18637/jss.v051.i04.

Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis: Theory and Practice. Springer Science & Business Media, http://dx.doi.org/10.1007/0-387-36620-2.

Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., Hausdorff, J.M., 2005a. Effect of gait speed on gait rhythmicity in Parkinson's disease: variability of stride time and swing time respond differently. J. NeuroEng. Rehabil. 2 (1), 23. http://dx.doi.org/10.1186/1743-0003-2-23.

Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., Hausdorff, J.M., 2005b. Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease. Mov. Disord. 20 (9), 1109–1114. http://dx.doi.org/10.1002/mds.20507.

Fuchs, K., Scheipl, F., Greven, S., 2015. Penalized scalar-on-functions regression with interaction term. Comput. Statist. Data Anal. 81, 38–51. http://dx.doi.org/10.1016/j.csda.2014.07.001.
García-Portugués, E., González-Manteiga, W., Febrero-Bande, M., 2014. A goodness-of-fit test for the functional linear model with scalar response. J. Comput. Graph. Statist. 23 (3), 761–778. http://dx.doi.org/10.1080/10618600.2013.812519.
Gertheiss, J., Goldsmith, J., Crainiceanu, C., Greven, S., 2013. Longitudinal scalar-on-functions regression with application to tractography data. Biostatistics 14 (3), 447–461. http://dx.doi.org/10.1093/biostatistics/kxs051.
Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C., Stanley, H.E., 2000. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. Circulation 101 (23), e215–e220. http://dx.doi.org/10.1161/01.cir.101.23.e215.
Goldsmith, J., Scheipl, F., 2014. Estimator selection and combination in scalar-on-function regression. Comput. Statist. Data Anal. 70, 362–372. http://dx.doi.org/10.1016/j.csda.2013.10.009.
James, G.M., 2002. Generalized linear models with functional predictors. J. R. Stat. Soc. Ser. B Stat. Methodol. 64 (3), 411–432. http://dx.doi.org/10.1111/1467-9868.00342.
Jiang, C.R., Wang, J.L., 2011. Functional single index models for longitudinal data. Ann. Statist. 39 (1), 362–388. http://dx.doi.org/10.1214/10-AOS845.
Li, Y., Wang, N., Carroll, R.J., 2010. Generalized functional linear models with semiparametric single-index interactions. J. Amer. Statist. Assoc. 105 (490), 621–633. http://dx.doi.org/10.1198/jasa.2010.tm09313.
Liu, X., Müller, H.G., 2004. Functional convex averaging and synchronization for time-warped random curves. J. Amer. Statist. Assoc. 99, 687–699. http://dx.doi.org/10.1198/016214504000000999.
Marron, J.S., Ramsay, J.O., Sangalli, L.M., Srivastava, A., 2014. Statistics of time warpings and phase variations. Electron. J. Stat. 8 (2), 1697–1702. http://dx.doi.org/10.1214/14-ejs901.
Marron, J.S., Ramsay, J.O., Sangalli, L.M., Srivastava, A., 2015. Functional data analysis of amplitude and phase variation. Statist. Sci. 30 (4), 468–484. http://dx.doi.org/10.1214/15-sts524.
Morris, J.S., 2015. Functional regression. Annu. Rev. Stat. Appl. 2, 321–359. http://dx.doi.org/10.1146/annurev-statistics-010814-020413.
Müller, H.G., Stadtmüller, U., 2005. Generalized functional linear models. Ann. Statist. 774–805. http://dx.doi.org/10.1214/009053604000001156.
Nadaraya, E.A., 1964. On estimating regression. Theory Probab. Appl. 9 (1), 141–142. http://dx.doi.org/10.1137/1109020.
Ramsay, J.O., Dalzell, C.J., 1991. Some tools for functional data analysis. J. R. Stat. Soc. Ser. B Stat. Methodol. 53 (3), 539–572. http://dx.doi.org/10.1111/j.2517-6161.1991.tb01844.x.
Ramsay, J.O., Li, X., 1998. Curve registration. J. R. Stat. Soc. Ser. B Stat. Methodol. 60, 351–363. http://dx.doi.org/10.1111/1467-9868.00129.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, second ed. Springer, http://dx.doi.org/10.1007/b98888.
Reiss, P.T., Goldsmith, J., Shang, H.L., Ogden, R.T., 2017. Methods for scalar-on-function regression. Internat. Statist. Rev. 85 (2), 228–249. http://dx.doi.org/10.1111/insr.12163.
Srivastava, A., Klassen, E., 2016. Functional and Shape Data Analysis. Springer, http://dx.doi.org/10.1007/978-1-4939-4020-2.
Srivastava, A., Wu, W., Kurtek, S., Klassen, E., Marron, J.S., 2011. Registration of functional data using Fisher-Rao metric. arXiv:1103.3817.
Stoker, T.M., 1986. Consistent estimation of scaled coefficients. Econometrica 1461–1481. http://dx.doi.org/10.2307/1914309.
Tucker, J.D., Wu, W., Srivastava, A., 2013. Generative models for functional data using phase and amplitude separation. Comput. Statist. Data Anal. 61, 50–66. http://dx.doi.org/10.1016/j.csda.2012.12.001.
Winning, H., Roldán-Marín, E., Dragsted, L.O., Viereck, N., Poulsen, M., Sánchez-Moreno, C., Cano, M.P., Engelsen, S., 2009. An exploratory NMR nutri-metabonomic investigation reveals dimethyl sulfone as a dietary biomarker for onion intake. Analyst 134 (11), 2344–2351. http://dx.doi.org/10.1039/b918259d.
Yogev, G., Giladi, N., Peretz, C., Springer, S.l., Simon, E.S., Hausdorff, J.M., 2005. Dual tasking, gait rhythmicity, and Parkinson's disease: which aspects of gait are attention demanding?. Eur. J. Neurosci. 22 (5), 1248–1256. http://dx.doi.org/10.1111/j.1460-9568.2005.04298.x.