Scalar-on-Shape Regression Models for Functional Data Analysis

Sayan Bhadra and Anuj Srivastava*
Department of Statistics, Florida State University

November 27, 2024

Abstract

Functional data contains two components: shape (or amplitude) and phase. This paper focuses on a branch of functional data analysis (FDA), namely Shape-Based FDA, that isolates and focuses on shapes of functions. Specifically, this paper focuses on Scalar-on-Shape (ScoSh) regression models that incorporate the shapes of predictor functions and discard their phases. This aspect sets ScoSh models apart from the traditional Scalar-on-Function (ScoF) regression models that incorporate full predictor functions. ScoSh is motivated by object data analysis, , e.g., for neuro-anatomical objects, where object morphologies are relevant and their parameterizations are arbitrary. ScoSh also differs from methods that arbitrarily pre-register data and uses it in subsequent analysis. In contrast, ScoSh models perform registration during regression, using the (non-parametric) Fisher-Rao inner product and nonlinear index functions to capture complex predictor-response relationships. This formulation results in novel concepts of *regression phase* and *regression mean* of functions. Regression phases are time-warpings of predictor functions that optimize prediction errors, and regression means are optimal regression coefficients. We demonstrate practical applications of the ScoSh model using extensive simulated and real-data examples, including predicting COVID outcomes when daily rate curves are predictors.

Keywords: shape regression, shape models, COVID data analysis, functional shapes, shape-based FDA, functional regression analysis.

1 Introduction and Literature Survey

Rapid advances in data collection and storage technologies have led to a surge in problems where the data objects are functions recorded over time and space. Functional datasets

in neuroimaging, biology, epidemiology, meteorology, and finance have fuelled a growing interest in Functional Data Analysis (FDA). FDA deals with statistical analysis, including clustering, summarizing, modeling, and testing functional data. Functional regression incorporates functional variables in regression models as predictors, responses, or both. Specifically, *Scalar-on-Function* (ScoF) regression occurs when the predictors are functions and responses are scalar (or vectors). This problem has widespread applications in many scientific domains, with several examples presented later in this paper. Scalar-on-function regression is a natural extension of the standard multivariate regression model for the FDA.

Our focus differs from traditional ScoF by emphasizing the *shapes* (amplitudes) of functions rather than the functions themselves. This focus is motivated, for example, by problems in neuroimaging where morphologies of anatomical objects are used to predict clinical measurements. Accordingly, we develop a regression model where *shapes of functions* are predictors for scalar responses. Mathematically, *shape* is a property that is invariant to certain transformations considered nuisances in shape analysis (Kendall et al. [1999], Dryden and Mardia [2016]). For scalar functions, $\{f_i\}$, shapes relate to the number and height of extremes (peaks and valleys), but the locations are considered nuisances. Changes in the locations of these points, represented by diffeomorphisms $\{\gamma_i\}$ and implemented using the transformation $\{f_i \mapsto f_i \circ \gamma_i\}$, are called phase changes and are ignored in shape analysis. Thus, the shapes of a function f_i and its diffeomorphic time warping $f_i \circ \gamma_i$ are deemed identical. Shape-based FDA (see Wu et al. [2024], Srivastava and Klassen [2016], Marron et al. [2014, 2015], Stoecker et al. [2023]) is gaining interest, especially when phase variability is less critical, such as in COVID-19 rate curves where peaks represent pandemic waves. Several methods (*e.g.*, Marron et al. [2014, 2015], Srivastava and Klassen [2016]) have been developed to separate shape from phase as a stand-alone tool in FDA. This paper

introduces a regression model that separates phases and amplitudes within the statistical model, not as a preprocessing step. This approach can enhance performance and interpretation by optimizing the phases when they are uninformative. The literature on shape regression is scarce, but extensive research exists in some related areas. We summarize these contributions next.

- **Scalar-on-Function (ScoF) Regression Models**: We start with the basic (parametric) functional linear model of (FLM) of Ramsay and Silverman [2005]:

$$y_i = \alpha + \langle \beta, f_i \rangle + \epsilon_i, \ i = 1, 2, \ldots, n, \tag{1}$$

where $f_i : [0, T] \to \mathbb{R}$ is the predictor (element of a function space $\mathcal{F}$) and $y_i \in \mathbb{R}$ is the response. Also, $\alpha \in \mathbb{R}$ is the offset, $\beta \in \mathcal{F}$ is the coefficient function, and $\epsilon_i \in \mathbb{R}$ are the measurement errors. Here $\langle \beta, f_i \rangle = \int \beta(t) f_i(t) \ dt$ denotes the $\mathbb{L}^2$ inner product. FLM assumes $i.i.d$ observation noise, $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. To estimate β, one commonly minimizes the term $\sum_{i=1}^{n} (y_i - \alpha - \langle \beta, f_i \rangle)^2$. Randolph et al. [2012] used principal components of the predictor functions as an orthonormal basis for β. To regularize β, one adds a penalty term $\lambda \mathcal{R}(\beta)$, where $\lambda > 0$ is a tuning parameter; see Marx and Eilers [1999], James et al. [2009], Reiss and Ogden [2007], Lee and Park [2012], Zhao et al. [2012].

Ait-Saïdi et al. [2008] introduced non-linearity to regression models by introducing a function $h : \mathbb{R} \to \mathbb{R}$ to result in the model $y_i = h(\langle \beta, f_i \rangle) + \epsilon_i$. They used a kernel to estimate h, whereas Eilers et al. [2009] alternatively optimize β and h with smoothness constraints. Several authors (James and Silverman [2005], Amato et al. [2006], Ferraty

Table 1: Listing of various models studied in this paper

Single-Index Scalar-on-Shape (SI-ScoSh)	$y_i = g(f_i(0)) + h\left(\sup_{\gamma_i \in \Gamma} \langle \beta, q_i \star \gamma_i \rangle\right) + \epsilon_i$
Single-Index Scalar-on-Function (SI-ScoF)($\mathbb{L}^2$)	$y_i = c_i + h\left(\langle \beta, f_i \rangle\right) + \epsilon_i$
Single-Index Scalar-on-Function (SI-ScoF)(FR)	$y_i = g(f_i(0)) + h\left(\langle \beta, q_i \rangle\right) + \epsilon_i$
Scalar-on-Shape (ScoSh):	$y_i = g(f_i(0)) + \sup_{\gamma_i \in \Gamma} \langle \beta, q_i \star \gamma_i \rangle + \epsilon_i$
Scalar-on-Function (ScoF)($\mathbb{L}^2$):	$y_i = g(f_i(0)) + \langle \beta, f_i \rangle + \epsilon_i$
Scalar-on-Function (ScoF)(FR):	$y_i = g(f_i(0)) + \langle \beta, q_i \rangle + \epsilon_i$

et al. [2013]) have studied multiple index models of the type:

$$y_i = \alpha_0 + \sum_{j=1}^{r} h_j\left(\langle \beta_j, f_i \rangle\right) + \epsilon_i \ , \tag{2}$$

for an arbitrary r. McLean et al. [2014] further generalized the model using a time-indexed set of functions: $y_i = \int_0^1 \langle \beta_{t,\cdot}, f_i(\cdot) \rangle \ dt + \epsilon_i$, (where β is now a bi-variate function). Amongst notable nonparametric approaches, Boj et al. [2010] introduced a weighted distance-based regression for functional predictors using semi-metrics on function spaces. Boj et al. [2016] introduced non-parametric link functions to generalize their earlier models.

- **Shape-on-Scalar (ShoSc) Regression Models**: There is extensive literature on the inverse problem, where the shapes of functions form responses, and predictors are Euclidean vectors, see Lin et al. [2017], Shi et al. [2009], Tsagkrasoulis and Montana [2018]. An example of this problem is when the scalar predictor is time, and the goal is to fit a time curve on a shape space given finite observations. This also relates to fitting *smoothing splines* on shapes. Intrinsic manifold valued regression models have been studied widely by Ghosal et al. [2023], Petersen and Müller [2019], whereas extrinsic models have been studied by Lin et al. [2019]. A wide literature on geodesic regression (Thomas Fletcher [2013], Shin and Oh [2022]) also belongs to this category.

- **Scalar-on-Shape (ScoSh) Regression Models**: Ahn et al. [2020] first studied a ScoSh model but with a major limitation. Since y_is depend on the shapes of f_is, they must be invariant to phase changes in f_is. Thus, the response should remain unchanged if f_i is replaced by $f_i \circ \gamma$ in the model. In Eqns. 1 and 2, the $\mathbb{L}^2$ inner-product fails to provide this invariance because $\langle \beta, f_i \rangle \neq \langle \beta, f_i \circ \gamma_i \rangle$ generally. Even under identical dual transformation, we don't have equality, *i.e.*, $\langle \beta, f_i \rangle \neq \langle \beta \circ \gamma, f_i \circ \gamma \rangle$. This rules out using $\sup_\gamma \langle \beta, f_i \circ \gamma \rangle$ to remove phase variability, as it is degenerate and not phase-invariant (see Srivastava and Klassen [2016]). Ahn et al. [2020] replaced the $\mathbb{L}^2$ inner-product $\langle \beta, f_i \rangle$ in Eqns 1, 2 with $\sup_{\gamma_i} \langle \beta, (f_i \circ \gamma_i)\sqrt{\dot{\gamma_i}} \rangle$. Although this term has some stability to changes in γ_i, it does not achieve the desired invariance. Another approach is using a phase-invariant shape metric d_s in a nonparametric model, see Delicado [2024].

We modify past regression models using the Fisher-Rao Riemannian metric (FRM), termed d_{FR}, to create a new ScoSh model. d_{FR} is phase invariant in the sense that $d_{FR}(f_1, f_2) = d_{FR}(f_1 \circ \gamma, f_2 \circ \gamma)$ for all warpings γ. The use of Square-Root Velocity Function (SRVF), specified later, simplifies the computation of d_{FR}. Under SRVF, the Fisher-Rao inner product becomes the $\mathbb{L}^2$ inner product, $\langle f_1, f_2 \rangle_{FR} = \langle q_1, q_2 \rangle_{\mathbb{L}^2}$, where q_is are the SRVFs of f_is. This motivates an alternative term as a phase invariant inner product for the model. FRM's invariance complicates parameter estimation, as the phases are nuisance variables that need to be removed through optimization, affecting parameter estimation. Table 1 lists a summary of the regression models and their acronyms used in this paper. The main contributions of this paper are:

- It develops a new scalar-on-shape (ScoSh) regression model that uses the Fisher-Rao Riemannian metric to achieve invariance to the phase component of predictor functions. It solves the function registration (phase separation) inside the regression

model rather than treating it as a preprocessing step.

- It introduces a concept of *regression phase* and *regression mean* associated with functional data. While the past definitions of mean shape and phase of in FDA (Marron et al. [2014, 2015], Srivastava and Klassen [2016]) are based on optimal alignments of peaks and valleys, the regression phase and mean result from those optimal time warpings that help minimize prediction error of the response variable.

- It uses classical index models (single and multiple) for enveloping the Fisher-Rao inner products to introduce nonlinear relationships in the model.

- It performs exhaustive experimental evaluations of the proposed model using simulated data (with known ground truths) and real data with interpretable solutions. The modeling performances compete successfully with state-of-the-art methods.

- The ScoF models can differ depending on the inner product between β and f_i: the $\mathbb{L}^2$ and Fisher-Rao inner products. The $\mathbb{L}^2$ version is the commonly used FLM model, but we also include the Fisher-Rao version in the experiments for comparisons.

2 Proposed Method

The proposed scalar-on-shape (ScoSh) regression model requires the notion of shape in precise mathematical terms. First, we summarize the concept of *shapes* of scalar functions and their treatments. We then introduce the proposed ScoSh model and its properties. In the process, we also provide a novel concept of *Regression mean and phase*. We follow up with model estimation and a Bootstrap analysis of this estimator.

2.1 Background: Quantifying Shapes of Scalar Functions

Let $\mathcal{AC}$ be the set of all absolutely-continuous functions on $[0, 1]$ and $\mathcal{AC}_0$ be a subset of $\mathcal{AC}$ that satisfies $f(0) = 0$. Also, let Γ be the space of all boundary-preserving positive diffeomorphisms of the unit interval $[0, 1]$ to itself, $i.e.$, $\Gamma := \{\gamma : [0, 1] \to [0, 1] | \gamma(0) = 0, \gamma(1) = 1, \gamma$ is a diffeomorphism$\}$. Γ forms the time-warping group, and the action of Γ on $\mathcal{AC}_0$ is the mapping $\mathcal{AC}_0 \times \Gamma \to \mathcal{AC}_0$ given by $(f, \gamma) \triangleq f \circ \gamma$. The mapping $f \mapsto f \circ \gamma$ simply changes the **phase** of f but not its shape. Since the shape of f is deemed unchanged by the mapping $f \mapsto f \circ \gamma$, we define $f \sim g$ to be an equivalence relation on $\mathcal{AC}_0$, where $g = (f \circ \gamma)$ for some $\gamma \in \Gamma$. An equivalence class under this relation is given by: $[f] = \{f \circ \gamma | \gamma \in \Gamma\}$. Such an equivalence class uniquely represents a shape, and the set of all shapes is the quotient space $\mathcal{S} = \mathcal{AC}_0 / \Gamma = \{[f] | f \in \mathcal{AC}_0\}$.

To develop a regression model similar to Eqn 1 using elements of the shape space $\mathcal{S}$, we need an inner product on $\mathcal{S}$. As discussed in Srivastava and Klassen [2016], the classical $\mathbb{L}^2$ inner product is unsuitable for shape analysis. Instead, we use the Fisher-Rao Riemannian metric with the required invariance properties. This metric is complex, and one uses the Square-Root-Velocity-Function (SRVF) representation (Srivastava and Klassen [2016]) for simplification. The SRVF of a function $f \in \mathcal{AC}$ is defined to be $q = Q(f) \triangleq \text{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$. The mapping $Q : f \mapsto q$ is a bijection between $\mathcal{AC}_0$ and $\mathbb{L}^2$, with the inverse map given by $Q^{-1}(q)(t) \triangleq f(t) = \int_0^t q(s)|q(s)| \ ds$. Thus, the mapping $f \mapsto (f(0), q)$ is a bijection between the larger set $\mathcal{AC}$ and $\mathbb{R} \times \mathbb{L}^2$.

For any $f \in \mathcal{AC}_0$ and $\gamma \in \Gamma$, the SRVF of the composition $f \circ \gamma$ is given by $Q(f \circ \gamma) = (q \circ \gamma)\sqrt{\dot{\gamma}}$; We will denote it by $q \star \gamma$. For a shape class $[f] \subset \mathcal{AC}$, the corresponding subset in $\mathbb{L}^2$ given by: $[q] = \{(q \star \gamma) | \gamma \in \Gamma\}$. There are several advantages to using SRVFs in shape analysis of functions. One is that the Fisher-Rao inner product between any two functions

f_1 and f_2 is the $\mathbb{L}^2$ inner product between their SRVFs, $i.e.$, $\langle f_1, f_2 \rangle_{FR} = \langle q_1, q_2 \rangle_{\mathbb{L}^2}$ and the Fisher-Rao distance is $d_{FR}(f_1, f_2) = \|q_1 - q_2\|_{\mathbb{L}^2}$, where q_1, q_2 are the SRVFs of f_1, f_2, respectively. (From hereon, we will use $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ to denote the $\mathbb{L}^2$ inner product and norm.) With this identification, the invariance property of the Fisher-Rao metric can be stated as:

$$\langle q_1, q_2 \rangle = \langle q_1 \star \gamma, q_2 \star \gamma \rangle, \ \text{ or } \ \|q_1 - q_2\| = \|(q_1 \star \gamma) - (q_2 \star \gamma)\| \ . \tag{3}$$

This invariance property leads to a well-defined *shape metric* $d_S(q_1, q_2) \equiv \inf_{\gamma \in \Gamma} \|q_1 - (q_2 \star \gamma)\|$. Expanding the square of d_s, we get $\inf_{\gamma \in \Gamma} \left(\|q_1\|^2 + \|q_2\|^2 - 2\langle q_1, q_2 \star \gamma \rangle \right) = \|q_1\|^2 + \|q_2\|^2 - \sup_{\gamma \in \Gamma} 2\langle q_1, q_2 \star \gamma \rangle$. This shows that if the norms of q_1, q_2 are constant, then d_S is negatively proportional to the quantity: $\sup_{\gamma \in \Gamma} \langle q_1, q_2 \star \gamma \rangle$. This last term motivates the phase-invariant inner product in the proposed model.

2.2   Proposed Scalar-on-Shape (ScoSh) Regression Model

To focus on shapes of $\{f_i\}$, we need invariance to the phase of $\{f_i\}$, $i.e.$, replacing any f_i with $f_i \circ \gamma_i$ should not change the response y_i. To achieve this, we use $\sup_{\gamma \in \Gamma} \langle \beta, q_i \star \gamma \rangle$ as a surrogate for $\langle \beta, f_i \rangle$ in Eqn. 1. The invariance of the Fisher-Rao inner product and the group structure of Γ results in the property: $\sup_{\gamma \in \Gamma} \langle \beta, q_i \star \gamma \rangle = \sup_{\gamma \in \Gamma} \langle \beta, (q_i \star \gamma_0) \star \gamma \rangle$, for any $\gamma_0 \in \Gamma$. Thus, this expression is truly invariant to the phase of f_i and depends only on its shape. To add flexibility to the model, we introduce two functions: (1) an index function $h : \mathbb{R} \to \mathbb{R}$, and (2) an offset function $g : \mathbb{R} \to \mathbb{R}$. We will assume that $h, g \in \mathcal{C}(\mathbb{R}, \mathbb{R})$. The overall model can now be stated as:

$$y_i = g(f_i(0)) + h \left(\sup_{\gamma_i \in \Gamma} \langle \beta, q_i \star \gamma_i \rangle \right) + \epsilon_i \ , \ i = 1, \ldots, n. \tag{4}$$

8

Here $q_i, \beta \in \mathbb{L}^2$, $\gamma_i \in \Gamma$, $g, h \in \mathcal{C}$, and $\epsilon_i \in \mathbb{R}$ are *i.i.d.* from $N(0, \sigma^2)$. We will call this the Single-Index Scalar-on-Shape (SI-ScoSh) model. The parameters of this model are $\{\beta, h, g, \sigma^2\} \in \mathbb{L}^2 \times \mathcal{C} \times \mathcal{C} \times \mathbb{R}_+$. As a special case, we will also study when $h(x) = x$ and will call it the ScoSh model (without the SI prefix). Next, we discuss important properties of this model and impose conditions on the parameters to enforce identifiability.

1. **Fisher-Rao vs. $\mathbb{L}^2$ Inner Product**: One might ask why not use $\sup_{\gamma_i \in \Gamma} \langle \beta, f_i \circ \gamma_i \rangle$ instead of $\sup_{\gamma_i \in \Gamma} \langle \beta, q_i \star \gamma_i \rangle$ in the model? The reason is that the former is degenerate and loses information about f_i. Mathematically, the issue is $\langle f_1, f_2 \rangle \neq \langle f_1 \circ \gamma, f_2 \circ \gamma \rangle$. In contrast, the invariance property of SRVFs in Eqn. 3 is essential for this model.

2. **Properties of the Supremum Term**: The term $\sup_{\gamma \in \Gamma} \langle \beta, q_i \star \gamma \rangle$ is not linear in q_i, due to the presence of the sup operation. Also, this term is non-negative, which limits its direct use in the regression model. However, using the index function h allows for negative values of y_is.

3. **Identifiability of β**: Note that β is defined only up to its equivalence class $[\beta]$ since, $\sup_{\gamma_i \in \Gamma} \langle \beta, q_i \star \gamma_i \rangle = \sup_{\gamma_i \in \Gamma} \langle \beta \star \gamma_0, q_i \star \gamma_i \rangle$, for any $\gamma_0 \in \Gamma$. To ensure uniqueness, we restrict ourselves to a specific element of this class, as follows: We impose an additional *centering* condition on β through the phases $\{\widehat{\gamma}_i\}$. We require that $\frac{1}{n} \sum_{i=1}^{n} \widehat{\gamma}_i = \gamma_{id}$ (note that $\gamma_{id}(t) = t$), where $\widehat{\gamma}_i = \text{argmax}_{\gamma_i \in \Gamma} \langle \beta, q_i \star \gamma_i \rangle$. Once all the $\widehat{\gamma}_i$s are computed, we can simply use their average $\bar{\gamma} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\gamma}_i$ to center any estimate of β. In a standard FLM model (Eqn. 1), the search for β can be restricted to the span of $\{f_i\}$ since any component of β lying in the orthogonal of the span is lost after the inner product. This simplification does not hold in the proposed model. Even when $h(x) = x$, β is an element of a much larger space: $\text{span}\{[q_i], i = 1, \ldots, n\} = \{\sum_{i=1}^{n} a_i(q_i \star \gamma_i) \mid \gamma_1, \ldots, \gamma_n \in \Gamma, \ a_i \in \mathbb{R}\}$.

4. **Identifiability of** h: Another degree of freedom is associated with the scale of the argument of h. Since $h(\sup_{\gamma_i \in \Gamma} \langle \beta, q_i \star \gamma_i \rangle) = h(\frac{1}{a} \sup_{\gamma_i \in \Gamma} \langle a\beta, q_i \star \gamma_i \rangle)$, for any $a \in \mathbb{R}_+$, this adds an ambiguity to the definition. One can remove it by imposing a constraint such as $\int h(t)\, dt = 1$, or if using a polynomial form, fixing a coefficient of h.

5. **Identifiability of** g: We can resolve any ambiguity in g by setting $g(0) = 0$.

With these constraints, the model is fully specified, and the parameters are well-defined.

2.3 Model Parameter Estimation

Next, we study the problem of estimating model parameters from the observed data $\{(f_i, y_i) \in \mathcal{AC} \times \mathbb{R},\ i = 1, 2, \ldots, n\}$. We pre-compute the SRVFs $\{q_i\} \in \mathbb{L}^2$ of the predictor functions $\{f_i\}$. Then, given the observations $\{(y_i, q_i, f_i(0)) \in \mathbb{R} \times \mathbb{L}^2, i = 1, \ldots, n\}$, the inference problem is to estimate the quantities h, g, β, and σ^2 from the data. To simplify estimation, we will express $\beta \in \mathbb{L}^2$ using a truncated orthogonal basis $\mathcal{B} = \{b_j, j = 1, \ldots, J\}$ according to: $\beta(t) = \sum_{j=1}^{J} c_j b_j(t)$. $\mathcal{B}$ can be either a predefined basis, *e.g.*, the Fourier basis, or can be extracted from the training dataset through functional PCA. Then, the maximum-likelihood estimates of h, g and $\mathbf{c} = \{c_j\}$ are given by:

$$(\widehat{\mathbf{c}}, \widehat{h}, \widehat{g}) = \operatorname*{argmin}_{\mathbf{c} \in \mathbb{R}^J, h \in \mathcal{C}(\mathbb{R}, \mathbb{R}), g \in \mathcal{C}(\mathbb{R}, \mathbb{R})} H(\mathbf{c}, g, h), \quad \text{where} \tag{5}$$

$$H(\mathbf{c}, g, h) \triangleq \left[\sum_{i=1}^{n} \left\{ y_i - g(f_i(0)) - h\left(\sup_{\gamma_i \in \Gamma} \left\langle \sum_{j=1}^{J} c_j b_j, (q_i \star \gamma_i) \right\rangle \right) \right\}^2 \right].$$

One can impose a roughness penalty on β to control its smoothness, if needed.

Iterative Parmeter Estimation: To minimize H with respect to g, h, and β, we use a coordinate-descent approach, optimizing one parameter at a time while fixing the others. Estimating σ^2 from the residual variance is straightforward and not discussed. Algorithm

2 summarizes these steps with finer details about the estimation process presented in the

---

Algorithm 1 Estimation of β keeping h and g fixed

1: **Input** $\widehat{h}, \widehat{g}$., matrix of SRVF's $q = \{\tilde{q_1}, \cdots, \tilde{q_n}\}$, basis functions $b_1(t), \cdots, b_J(t)$.

2: **Initialize** $\mathbf{c} \in \mathbb{R}^J$. Compute initial $\widehat{\beta}(t) = \sum_{j=1}^{J} c_j$.

3: For each observation i, **find** the optimum time warping function : $\gamma_i' = arg \sup_{\gamma_i \in \Gamma} \left\langle \widehat{\beta}, q_i \star \gamma_i \right\rangle$

   using the Dynamic Programming algorithm(Srivastava and Klassen [2016]).

4: **Update** the SRVF's registering them to $\widehat{\beta}$: $q_i' = q_i \star \gamma_i'$.

5: Using an optimization method, (such as `fminunc` or `simulannealbnd` in MATLAB)

   minimize the cost function (5): $\widehat{\mathbf{c}} = arg \min_{c \in \mathbb{R}^J} H(\mathbf{c}, \widehat{h}, \widehat{g})$.

6: **Update** $\widehat{\beta}(t) = \sum_j \widehat{c}_j \cdot b_j(t)$ and $q_i = q_i' \ \forall i$.

7: **If** $H\left(\widehat{\mathbf{c}}, \widehat{h}, \widehat{g}\right)$ is large, return to step 3, else go to step 8.

8: To remove the extra degree of freedom in β, **compute** $\bar{\gamma} = \frac{1}{n} \sum_i \gamma_i'$.

9: Obtain the estimate $\widehat{\beta} = \widehat{\beta} \circ \bar{\gamma}^{-1}$.

Algorithm 2 Elastic shape regression model

1: **Initialise** $\widehat{h}(x) = h_0(x)$, $\widehat{g}(x) = 0$.

2: Given $\widehat{h}, \widehat{g}$, **estimate** $\widehat{\beta}$ using Algorithm 1.

3: Once obtained $\widehat{\beta}$, create $y_i' = y_i - \widehat{g}(f_i(0))$ and **estimate** $\widehat{h}$ using

   - **Define** estimated inner product as $\widehat{y_i} = \sup_{\gamma_i \in \Gamma} \langle \widehat{\beta}, q_i \star \gamma_i \rangle$.

   - **Fit** a polynomial or a non-parametric curve $\widehat{h}$ between the responses y_i''s and the estimated inner products $\widehat{y_i}$'s.

4: Remove the scaling degree of freedom from our estimate (by fixing the highest coefficient of h to 1 and adjusting the other coefficients of h and all of β accordingly).

5: With $y_i'' = y_i - \widehat{h}(\widehat{y_i})$ **calculate** $\widehat{g}$ using

11

- **Fit** a quadratic polynomial $\widehat{g}$ on the y''s. (As explained in Appendix 6.1, we restrict our search for optimal g to a quadratic polynomial).

6: Iterate steps 3 to 5 until H converges.

7: If $H\left(\widehat{c}, \widehat{h}, \widehat{g}\right)$ is small, then stop; else return to step 2.

2.4 Estimator Analysis Using Bootstrap Sampling

The estimators of β, h, and g haev been defined using a joint optimization problem (Eqn. 5) involving multiple parameters and nuisance variables. Ideally, one would like the distributions of estimated quantities for bias and consistency analysis. Several asymptotic distributions of β and h have been derived for FLM and related models (*e.g.*, Li et al. [2010], Morris [2015]). However, estimating regression parameters in the shape context is much more difficult. The cost function, which includes a supremum over the nuisance variables $\{\gamma_i\}$, is nonlinear and complex. Γ is an infinite-dimensional, nonlinear manifold, adding to the complexity. Additionally, Eqn. 4 has a potentially nonlinear index function h, complicating prediction error analysis. Du et al. [2015] developed a theory for regression modeling and analysis in shape matching, but their context differs from our functional data setting.

Lacking analytical distributions, we take a computational approach and rely on bootstrap sampling. Bootstrapping allows us to examine estimator properties (*e.g.*, variance) by sampling with replacement and approximating the distribution of estimators $(\widehat{\beta}, \widehat{h}, \widehat{g})$. We will empirically analyze these estimators by generating numerous bootstrap replicates.

To illustrate this approach, we conducted an experiment with parameters: $h(x) = x^2 - x$, $g(x) = x$, and β as shown in Fig. 1 top-left, and data simulated from Eqn. 4 (simulation details are provided later in Section 3). To evaluate estimator performance using Bootstrap, we generated 100 randomizations of train-test sets, performed estimation using Algorithm 2, and evaluated performance. From the bootstrap replicates, we computed 95% confidence
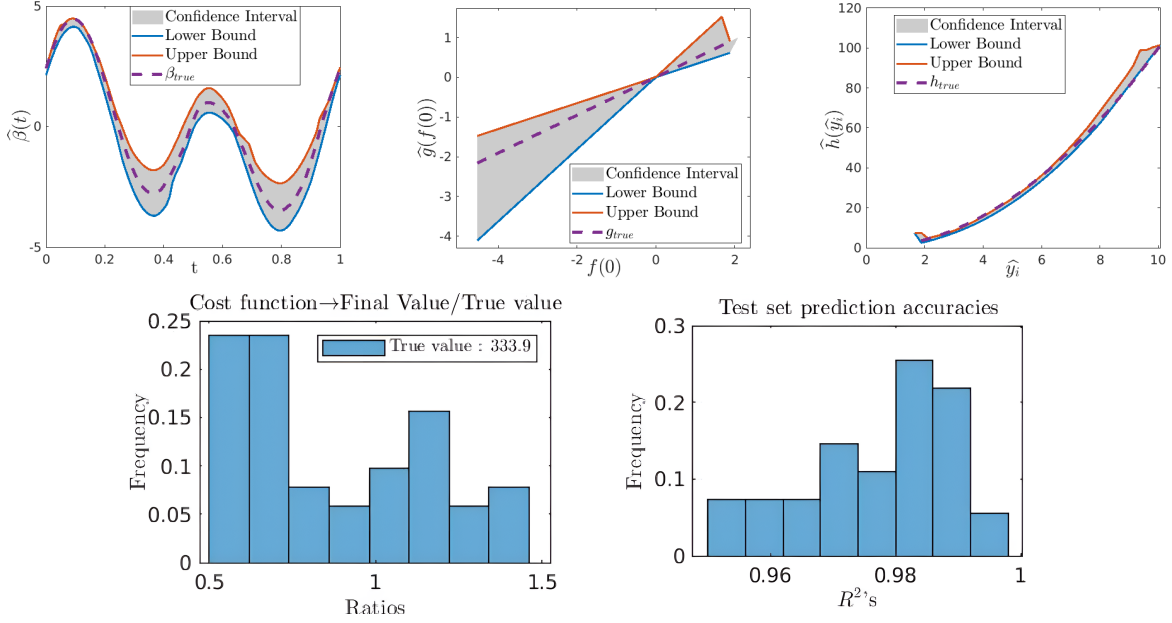
Figure 1: Top: The three panels plot 95% bootstrap confidence intervals of the estimated quantities $\widehat{\beta}$, $\widehat{g}$, and $\widehat{h}$, and their ground truth values. Bottom: The left panel shows a histogram of the ratio of H_{final} and H_{true} for bootstrap samples, and the right panel plots a histogram of the R^2 values of these samples.

intervals and compared them to the true values. Fig. 1 shows $\widehat{\beta}$, $\widehat{g}$, and $\widehat{h}$ from left to right. The gray regions depict the 95% confidence intervals, with red and blue curves denoting the bounds and dotted curves representing true values. These plots show that the true values of β, h, and g lie within the confidence intervals, validating our numerical approach.

The bottom-left shows a histogram (from 100 bootstrap samples) of the ratio $\frac{H_{final}}{H_{true}}$, where H_{final} is the converged value of H and H_{true} is the value of H for ground truth parameters. We can see that the final H values converge to within $0.5 - 1.5$ times the true value of the cost function. This underscores the good convergence properties of our gradient approach. The bottom-right histogram shows R^2 values (prediction accuracy) on test data for each of the 100 model fits, highlighting the excellent prediction performance of the estimated model.

2.5   Regression Phase and Regression Mean

Our estimation of model parameters involves aligning predictor SRVFs $\{q_i\}$ to the coefficient β using time warpings γ_i during estimation. This perspective allows us to define the phase components of f_is in a different way than the traditional phase-amplitude separation.

Classical Phase-Amplitude Separation: In the past work (Tucker et al. [2013], Marron et al. [2014, 2015], Srivastava and Klassen [2016], Zhang et al. [2018]), the *phases* of functions have been defined as the time-warpings required to align their peaks and valleys. Mathematically, the phase for a function f_i (with SRVF q_i) is defined as $\widehat{\gamma}_i = \arg\sup_{\gamma \in \Gamma} \langle \mu, q_i \star \gamma \rangle$, where $\mu \in \mathbb{L}^2$ is the Karcher or the Fréchet mean of the given functions and is defined using:

$$
\begin{aligned}
\mu &= \arg\inf_{q \in \mathbb{L}^2} \sum_{i=1}^{n} \left(\inf_{\gamma_i \in \Gamma} \| q - q_i \star \gamma_i \|^2 \right) \\
&= \arg\inf_{q \in \mathbb{L}^2} \sum_{i=1}^{n} \left(\|q\|^2 + \|q_i\|^2 - 2 \sup_{\gamma_i \in \Gamma} \langle q, q_i \star \gamma_i \rangle \right).
\end{aligned}
\tag{6}
$$

Note that $\{\widehat{\gamma}_i\}$ are defined through the optimization in Eqn 6. The left panel of Fig. 2 shows a cartoon example of this idea, where SRVFs $\{q_i\}$ are warped into $\{(q_i \star \gamma_i)\}$ to align with the current estimate of the shape average μ.

Regression-Based Function Aligment: Similarly, we define optimal time-warping in the ScoSh model using $\widehat{\gamma}_i = arg\max_{\gamma \in \Gamma} \left\langle \widehat{\beta}, q_i \star \gamma \right\rangle$, where the estimator of β is:

$$
\begin{aligned}
\widehat{\beta} &= \arg\inf_{\beta \in \mathbb{L}^2} \sum_{i=1}^{n} \left(y_i - g(f_i(0)) - h(\sup_{\gamma_i \in \Gamma} \langle \beta, q_i \star \gamma_i \rangle) \right)^2 \\
&= \arg\inf_{\beta \in \mathbb{L}^2} \sum_{i=1}^{n} \left(y_i - \sup_{\gamma_i \in \Gamma} \langle \beta, q_i \star \gamma_i \rangle \right)^2, \quad \text{assuming } h(x) = x, \quad g = 0.
\end{aligned}
\tag{7}
$$

Comparing Eqns. 6 and 7, we see the parallels between μ and $\widehat{\beta}$. In Eqn. 6, one seeks

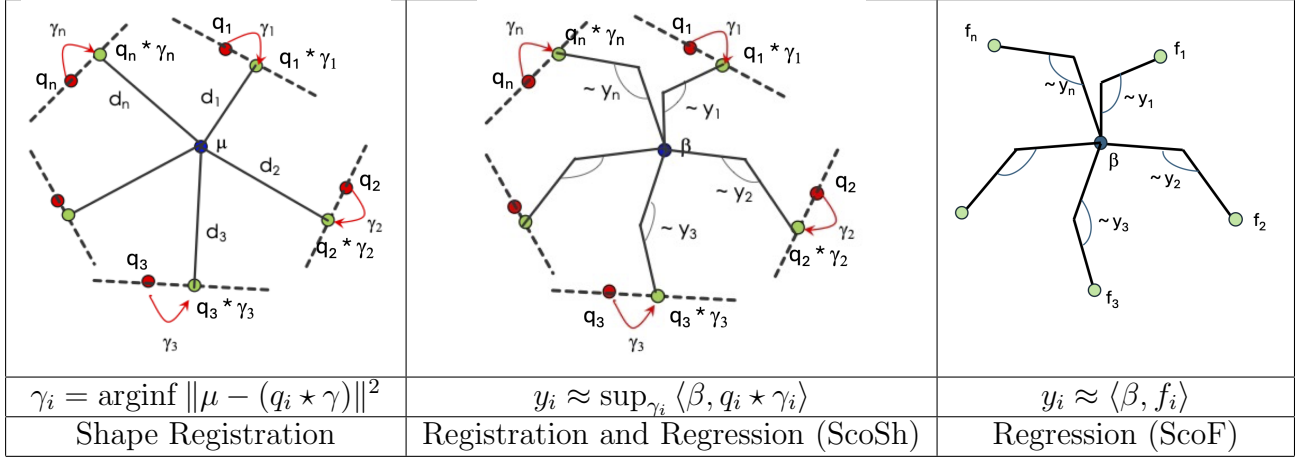| $\gamma_i = \arg\inf \lVert \mu - (q_i \star \gamma) \rVert^2$	$y_i \approx \sup_{\gamma_i} \langle \beta, q_i \star \gamma_i \rangle$	$y_i \approx \langle \beta, f_i \rangle$
Shape Registration	Registration and Regression (ScoSh)	Regression (ScoF)

Figure 2: Left: Alignment and shape averaging of functions using SRVFs; Middle: Alignment of SRVFs to regression coefficient β to approximate responses y_i through inner products; Right: Approximation of responses y_i through inner products without any alignment.

a μ that is closest to all $q_i \star \widehat{\gamma}_i$, and in the process making $2\sup_{\gamma_i \in \Gamma} \langle q, q_i \star \gamma_i \rangle$ as close to $\lVert \mu \rVert^2 + \lVert q_i \rVert^2$ as possible. Similarly, in Eqn. 7, the optimal $\widehat{\beta}$ makes $\sup_{\gamma_i \in \Gamma} \langle \beta, q_i \star \gamma_i \rangle$ as close to y_i as possible (assuming $h(x) = x$, $g = 0$). This motivates naming $\widehat{\beta}$ as the *regression mean* of the shapes of $\{f_i\}$ *w.r.t* responses $\{y_i\}$. The middle panel of Fig. 2 shows a cartoon illustration of this idea. The right depicts a ScoF or FLM model where one approximates responses $\{y_i\}$ using the inner products between $\{f_i\}$ and β without any alignment.
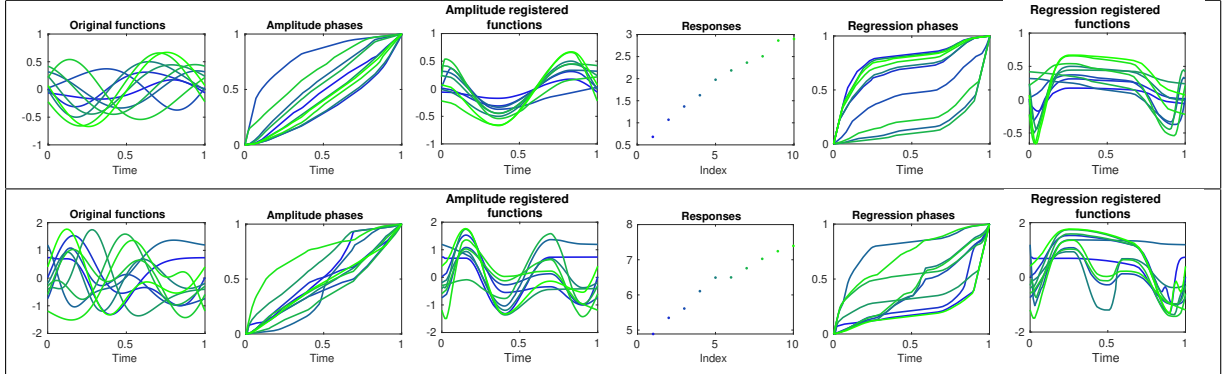

Figure 3: Two examples contrasting amplitude-phase and regression phase. In each row, the leftmost panel shows the original $\{f_i\}$, the second shows amplitude phases, and the third shows the aligned functions. The fourth panel shows response data $\{y_i\}$, the fifth shows regression phases, and the last shows regression registered functions.

Next, we present two simple examples in Fig. 3 to illustrate and compare regression means and amplitude means. Each row shows a different example. The simulation setup for these examples is the same as in Section 3. Here f_i's are constructed using simple Fourier basis, h and g are both lower-order polynomials, and true β is made of five Fourier bases. The traditional phase-amplitude separation seeks to align peaks and valleys in $\{f_i\}$, while the regression-based separation tries to match the inner product of β and $(q_i \star \gamma_i)$ with y_i. The results naturally show significant differences in the phases of the two approaches.

3 Experimental Results: Simulated Data

In this section, we simulate several datasets and use them to evaluate the proposed as well as some current models.

Simulation Setup: In this experiment, we generate $f_i^0(t) = c_{i,1}\sqrt{2}\sin(2\pi t) + c_{i,2}\sqrt{2}\cos(2\pi t)$, where $c_{i,1}, c_{i,2} \sim \mathcal{N}(0, 1^2)$. To create predictors with arbitrary phases, we perturb each of these f^0's by random γ_i's : $f_i(t) = f_i^0 \circ \gamma_i$, where $\gamma_i(t) = t + \alpha \cdot t(T - t)$ $\{t \in [0, T]$, $\alpha \in U(-1, 1)\}$. We calculate the corresponding SRVF's (q_i's) of each of these f_i's using $q_i = \text{sign}(\dot{f}_i(t))\sqrt{|\dot{f}_i(t)|}$. To define coefficient vector β, we use first J elements of the Fourier basis $\{b_j\} = \{\sqrt{2}\cos(2\pi jx), \sqrt{2}\sin(2\pi jx), j = 1, 2\ldots, J/2\}$ and some fixed coefficients $\tilde{c}_0 = \{1, \cdots, 1\}$. Also, we use low-order polynomials for h (listed in the experiments) and a fixed $g(x) = x^2 - 1$. Then, we calculate responses y_i's by adding $\epsilon_i \sim \mathcal{N}(0, 0.01^2)$ as per Eqn. 4. For a sample size of $n = 100$, we use 80% of the dataset for training and the rest for testing in a five-fold cross-validation. For each random split, we use Algorithm 2 to estimate the model parameters.

Model Comparisons: Next, we compare performance of the ScoSh model with three other models (refer to Table 1 for model acronyms and specifications): (1) SI-ScoF(FR),
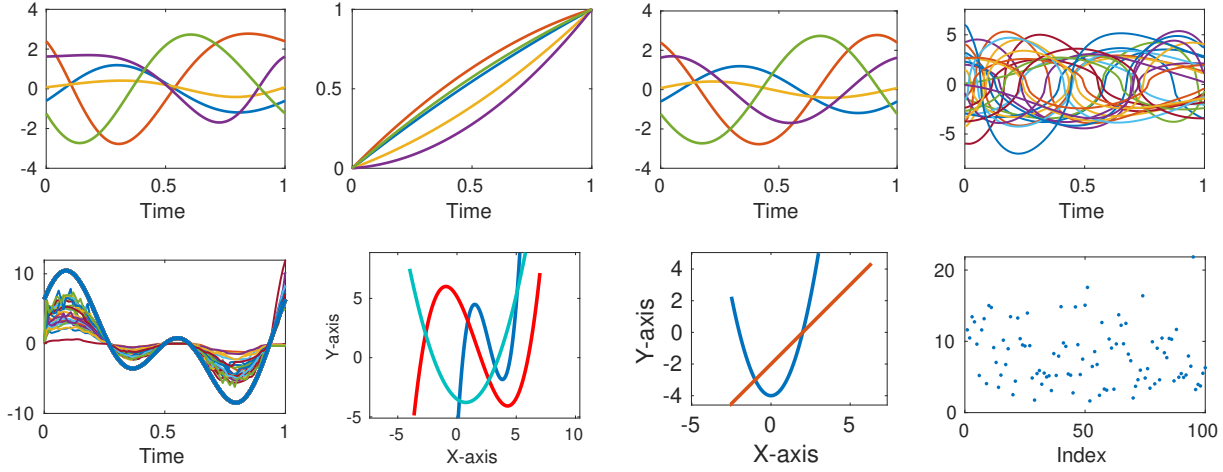
Figure 4: Simulating experimental data. Top: From left to right, some initial predictor functions $\{f_i^0\}$ formed using a Fourier basis, random time-warpings $\{\gamma_i\}$, time-warped functions $\{f_i = f_i^0 \circ \gamma_i\}$, and their SRVFs $\{q_i\}$. Bottom: SRVF's after registering with a β (blue line), some index functions h, some polynomial offset functions g, and the responses $\{y_i\}$ generated after adding random noise.

which uses the functions without alignment, (2) ScoSh, which uses SRVFs with alignment but sets h as identity, and (3) ScoF(FR), resembling the classical FLM but using the Fisher-Rao inner product. During estimation, SI-ScoSh iteratively optimizes over β, h, and g while registering functions. SI-ScoF(FR) optimizes over β, h, and g without registration. ScoSh includes registration and optimization over β, g. ScoF(FR) estimates β, g.

3.1   Evaluating Response Prediction

We sequentially generate data from one of these stated models, apply all the models to that data, and quantify model performances using five-fold validation. The original model is naturally expected to perform the best, but comparing the performances of others is also informative. We quantify prediction performance using the $R^2 \left(= 1 - \frac{\sum_i (y_i - \widehat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \right)$ statistic ($\bar{y}$ is the mean of the y_i's and $\widehat{y}_i$ is the predicted value of y_i's). In the tables, columns represent different polynomial choices for true h and numbers of basis functions (J) for true β, while rows correspond to different fitted models. The entries in cells are the means of

R^2 values over five-fold replications, with standard deviations in parenthesis. Additional tables can be found in the Supplementary Material.

1. **Data from SI-ScoSh model**: The left part of Table 2 shows results for data generated from the SI-ScoSh model; this model has a nonlinear predictor-response relationship and non-informative predictor phases. The first two rows show SI-ScoSh results for different estimators of h, both providing very high R^2 values. ScoSh model, *i.e.*, $h(x) = x$, performs increasingly worse as the true h becomes more complex. SI-ScoF(FR) captures some predictor-response relation but is inferior to ScoSh models. ScoF(FR) performs much worse, indicating the need to remove nuisance phase variability for effective performance. Note that a negative R^2 means that predicted values are worse than the fixed guess $\bar{y}$.

Table 2: Test (R^2) prediction performance comparison for data generated from SI-ScoSh (left three columns) and SI-ScoF(FR)(right three columns). linear: $h_{true}(x) = 3x - 2$, quadratic: $h_{true}(x) = x^2 - 3x + 2$, cubic: $h_{true}(x) = (x - 0.5)(x - 3)(x - 4.5)$. $\beta_{true}[J = 4] = \sum_{i=1,3} 2b_i + \sum_{i=2,4} \sqrt{2}b_i$ and $\beta_{true}[J = 6] = \sum_{i=1,3,5} 2b_i + \sum_{i=2,4,6} \sqrt{2}b_i$.

		SI-ScoSh			SI-ScoF(FR)	
h_{true}	linear	quadratic	cubic	linear	quadratic	cubic
$J(of\ \beta_{true})$	4	6	4	4	6	4
SI-ScoSh: Poly	0.96(0.02)	0.98(0.01)	0.98(0.01)	0.92(0.05)	0.89(0.06)	0.87(0.05)
SI-ScoSh: SVM	0.97(0.01)	0.98(0.01)	0.97(0.01)	0.94(0.02)	0.90(0.04)	0.89(0.04)
SI-ScoF(FR)	0.72(0.09)	0.48(0.26)	0.23(0.30)	0.99(0.01)	0.99(0.01)	0.99(0.01)
ScoSh	0.94(0.02)	0.72(0.10)	0.50(0.21)	< 0	< 0	< 0
ScoF(FR)	< 0	< 0	< 0	< 0	< 0	< 0

2. **Data from SI-ScoF model**: The right part of Table 2 shows prediction performance for data from the SI-ScoF(FR) model – a nonlinear index function h and an informative phase component. As expected, SI-ScoF(FR) performs best, with SI-ScoSh also doing well. ScoSh performs poorly, indicating the importance of the index function. Also, optimizing over $\{\gamma_i\}$ loses informative phase components and reduces performances. Prediction performance decreases from left to right as the complexity of h increases.

3. **Data from ScoSh model**: Table 3 shows prediction performances for data from the

ScoSh model, with $h(x) = x$ and non-informative phase components. Both ScoSh and SI-ScoSh give accurate predictions, with SI-ScoSh being a generalization of ScoSh. SI-ScoF(FR), despite keeping nuisance phases, performs decently as the index function helps compensate for the mismatch. The ScoF(FR) model, which keeps the nuisance phases but does not use an index function, performs poorly.

Table 3: Test (R^2) prediction performance comparison for data generated from ScoSh.

$J(of\ \beta)$	4	6
SI-ScoSh: Poly	0.98(0.01)	0.99(0.01)
SI-ScoSh: SVM	0.97(0.01)	0.98(0.01)
SI-ScoF(FR)	0.83(0.04)	0.68(0.2)
ScoSh	0.98(0.01)	0.96(0.03)
ScoF(FR)	< 0	< 0

Table 4: Test (R^2) prediction performance comparison for data generated from ScoF(FR)

$J(of\ \beta)$	4	6
SI-ScoSh: Poly	0.91(0.05)	0.92(0.05)
SI-ScoSh: SVM	0.94(0.03)	0.92(0.03)
SI-ScoF(FR)	0.99(0.01)	0.99(0.01)
ScoSh	< 0	< 0
ScoF(FR)	0.99(0.01)	0.99(0.01)

4. **Data from ScoF model**: Table 4 shows results on data generated from the ScoF(FR) model. Both SI-ScoF(FR)and ScoF(FR) show perfect R^2's. The proposed model, SI-ScoSh, also shows near-perfect prediction. ScoSh fails to capture the predictor-response relationship when phases are not nuisances.

From these experiments, we conclude that treating (predictor) phases as informative, when the data is generated using arbitrary phases, reduces the performance substantially. Conversely, ignoring the phases when they contain relevant information also impairs performance. Interestingly, the index function h can compensate to some extent for phase mistreatment, making indexed models perform better than non-indexed ones. However, this compensation is limited to simpler h_{true} and β_{true}; as they get more complex in shape, the index function struggles to compensate for phase mistreatment.

3.2 Evaluating Parameter Estimation

This section systematically evaluates estimation performances for different model parameters using simulated data.

1. **Estimation of Index Function** h: In this experiment, we study how the varying degree of the index function h affects the estimation performance of the SI-ScoSh model. We generate data from a quadratic or cubic h_{true} and allow different degrees $(1-4)$ during estimating of h. The pictorial results are shown in Fig. 5 while error summaries are presented in Table 5. The left two panels of Fig. 5 show estimated h for different h_{true}. One can see that higher-order polynomials improve estimation.
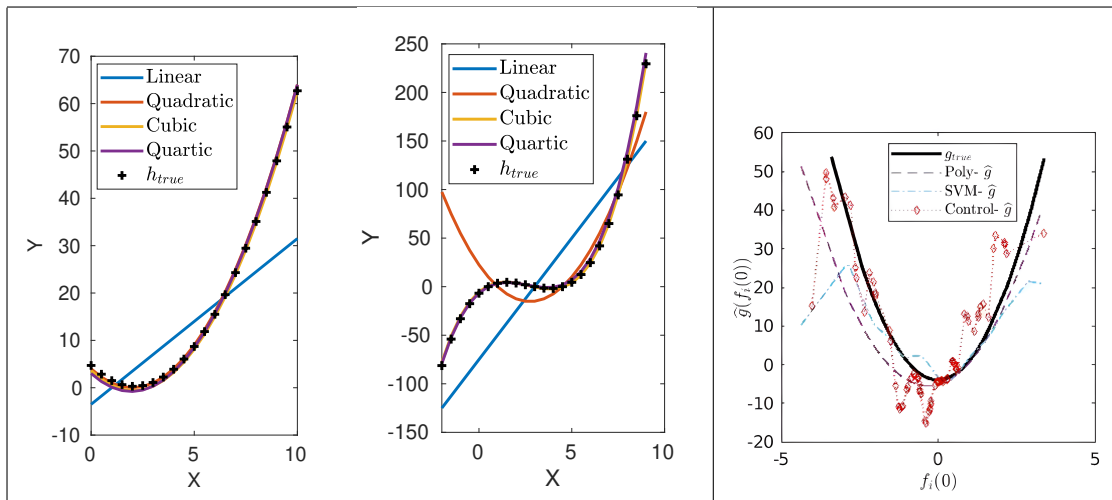

Figure 5: The estimate $\widehat{h}$ for $h_{true} = x^2 - 4x + 4$ (left) and $h_{true} = (x - \frac{1}{2})(x - 3)(x - 4.5)$ (middle) under the SI-ScoSh model. The right panel shows estimates of $\widehat{g}$ with a quadratic g_{true} for SI-ScoSh and SI-ScoF(FR) models.

2. **Estimation Regression Coefficient** β: Here we study estimation of β using different basis sets of $\mathbb{L}^2$ space. We construct β_{true} from 4 or 6 Fourier basis elements, and we estimate it under the SI-ScoSh model for different J values. As seen in the left and middle panels of Fig. 6, increasing J beyond the true degrees doesn't improve estimation of $\widehat{\beta}$ any further. This trend is mirrored in the predictive R^2's, presented in Table 6. For $J < J_{true}$, $\widehat{\beta}$'s have worse prediction performance compared to $J \geq J_{true}$ as they fail to capture the

Table 5: Prediction performance comparison for different complexities of $\widehat{h}$ with a quadratic (top) and cubic (bottom) h_{true} having $\beta_{true} = \sum_{i=1}^{4} \sqrt{2}b_i(t)$ and $g_{true}(x) = x^2 - 1$.

h_{true}	Pred. Performance	SI-ScoSh : Maximum allowed degree of $\widehat{h}$				SI-ScoF(FR)
quadratic	Test R^2	linear	quadratic	cubic	quartic	
	Mean(SD)	0.87(0.05)	0.98(0.01)	0.99(0.01)	0.98(0.01)	0.66(0.27)
	RMSE $(\widehat{h} - h_{true})$	4.15	0.13	0.19	0.20	9.1
cubic	Test R^2	linear	quadratic	cubic	quartic	
	Mean(SD)	0.67(0.11)	0.74(0.22)	0.95(0.04)	0.96(0.03)	0.51(0.2)
	RMSE $(\widehat{h} - h_{true})$	10.83	6.0	2.3	3.5	24.8

shape of the β_{true}. Fig. 6 and Table 6 show that further increasing the number of basis elements for β does not necessarily improve performance.
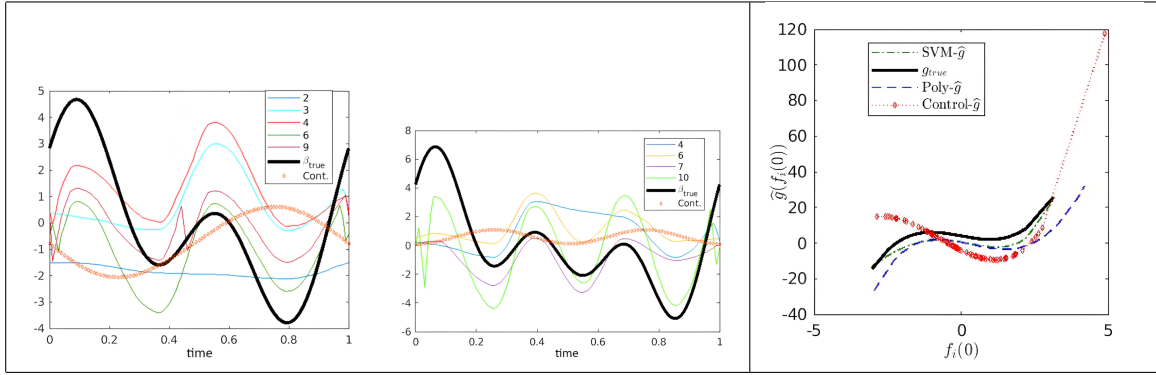

Figure 6: $\widehat{\beta}$'s when β_{true} has $J = 4$ (left) and $J = 6$ (middle). The numbers beside the colored lines in the legend show J used in estimating β. The orange diamonds show the estimated β for the SI-ScoF(FR) model. The right panel shows estimates of $\widehat{g}$ with a cubic g_{true} for SI-ScoSh and SI-ScoF(FR) models.

Note that we use the shape metric d_s, rather than RMSE, for evaluating $\widehat{\beta}$. As discussed in Section 2.2, the shape of β is more relevant in ScoSh model than β itself.

3. **Estimation Error for** g: Here, data is generated with a fixed h (a quadratic) and β composed of four Fourier basis elements, but we set g_{true} to be either quadratic or cubic. Then, we estimate g under the SI-ScoSh model and see how well we recover the structure of g_{true} under different models. The results are shown in the rightmost panels of Figs. 5 and 6. Both the relative RMSE between the true and estimated g and the prediction performances (see Table 7) establish the superiority of the SI-ScoSh model over the SI-ScoF(FR) model.

Table 6: Prediction performance comparison for different complexities of $\widehat{\beta}$ with $\beta_{true} = \sum_{i=1,3} 2b_i + \sum_{i=2,4} \sqrt{2}b_i$ (top) and $\beta_{true} = \sum_{i=1,3,5} 2b_i + \sum_{i=2,4,6} \sqrt{2}b_i$ (bottom). The used $h_{true}(x) = x^2 - 3x + 2$ and $g(x) = x^2 - 1$

β_{true}	Pred. Performance	SI-ScoSh : Number of basis functions for $\widehat{\beta}$					SI-ScoF(FR)
J=4	Test R^2	J=2	J=3	J=4	J=6	J=9	
	Mean(SD)	0.75(.07)	0.93(.03)	0.99(.01)	0.98(.01)	0.99(.01)	0.62(.14)
	RMSE ($\widehat{\beta} - \beta_{true}$)	3.6	2.8	2.4	3.8	3.2	5.6
J=6	Test R^2	J=4		J=6	J=7	J=10	
	Mean(SD)	0.85(0.08)		0.96(0.02)	0.99(0.01)	0.99(0.01)	0.4(0.45)
	RMSE ($\widehat{\beta} - \beta_{true}$)	4.9		4.0	3.8	4.5	7.2

Table 7: Prediction performance comparison among different models for different g_{true}'s

$h_{true}(x) = x^2 - 3x + 2$, $\beta_{true} = \sum_{i=1,3} 2b_i(t) + \sum_{i=2,4} \sqrt{2}b_i(t)$	$g_{true}(x)$	$x^3 - 3x + 4$		$5x^2 - 4$	
	Prediction	R^2	$\frac{\|\widehat{g} - g_{true}\|}{\|g_{true}\|}$	R^2	$\frac{\|\widehat{g} - g_{true}\|}{\|g_{true}\|}$
	SI-ScoSh (Poly)	0.99	0.33	0.98	0.36
	SI-ScoSh (SVM)	0.98	0.28	0.98	0.29
	SI-ScoF(FR)	0.54	0.59	0.12	0.97

3.3 Evaluating Model Invariance to Random Phases

The main goal of this paper is to design a regression model that is invariant to phase variability in predictor functions. While the proposed ScoSh and SI-ScoSh models satisfy this requirement theoretically, we also evaluate this property empirically. Specifically, we design response variables y_is that are by definition invariant to phase shifts in f_i. In other words, the responses are dependent exclusively on the shape of the corresponding predictor. We choose two cases: (1) $y_i = (\max(f_i(t)) - \min(f_i(t)) + \epsilon_i$, (2) $y_i = \int_0^1 |\dot{f}_i(t)| \, dt + \epsilon_i$. Here $\epsilon_i \sim \mathcal{N}(0, 0.5)$, and the predictors $\{f_i\}$ are generated as in Section 3. Then, we apply the proposed model to the noisy and time-warped data and study the results.

Fig. 7 presents results from these experiments. The two rows show results for two data cases. We train the models with a training set and evaluate them on a separate test set. Finally, we compare the prediction performances of SI-ScoSh and SI-ScoF(FR) on test sets. SI-ScoSh achieves $R^2 = 0.98$, while SI-ScoF(FR) has $R^2 < 0.1$. SI-ScoF(FR)'s lack of
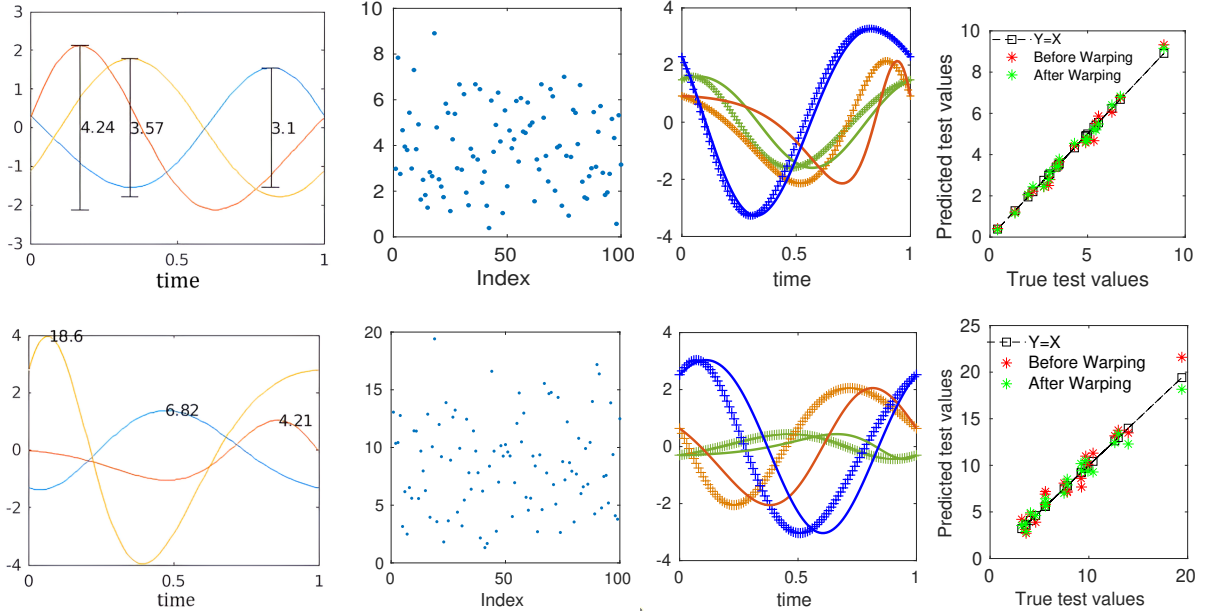
Figure 7: Examples of f_is (left), noisy measurements of y_is (second), the original test set in solid lines and their perturbed version in hashed lines (third panel), and SI-ScoSh predicted $\widehat{y}_i$ plotted versus true y_is (last panel). Top: Responses y_i are the max amplitude of predictors f_i. Bottom: Responses y_i are lengths of predictors f_i.

optimization over γ_is results in inferior performance. The high performance of the ScoSh model underscores its invariance to random phases of predictor functions.

4 Experimental Results: Real Data

In this section, we investigate the use of proposed ScoSh models on several real datasets. In each case, the functions are given without any prior registration, and we investigate the effectiveness of regressing scalar responses on the shapes of predictors. The detailed prediction performances of the different models are provided in a table format in the Supplementary Material.

1. **Spanish Weather Data**: This data contains daily summaries of geographical data of 73 Spanish weather stations selected from 1980-2009. Although this dataset contains other variables measured at each weather station, we focus only on the temperatures.

We form a predictor function f_i for each station with 365 average temperature values as follows. Each value is the average temperature recorded on a day (*e.g.*, February 3rd) for all years from 1980 to 1993. The corresponding scalar response y_i is the mean of temperatures for all days between 1994 and 2009 at that station. This data is shown in the top row of Fig. 8. The goal is to use past temperature patterns for each station to predict future average temperatures.
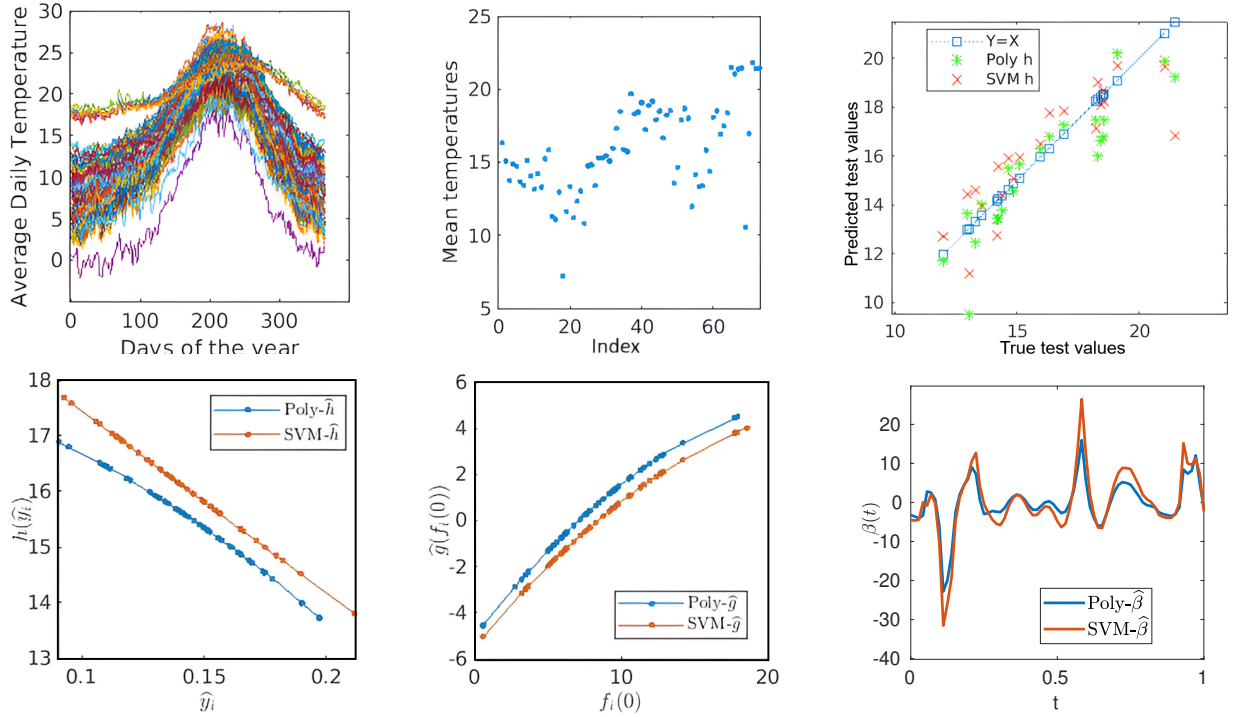

Figure 8: Spanish weather results – Top: Predictor functions $\{f_i\}$ (left), the responses $\{y_i\}$ (middle), and model predictions $\{\widehat{y}_i\}$ against true test values $\{y_i\}$ (right). Bottom: the estimated parameters under SI-ScoSh model - $\widehat{h}$ (left), $\widehat{g}$ (middle), and $\widehat{\beta}$ (right).

We apply the proposed and the competing models to this dataset to evaluate their prediction performances. We use two versions of SI-ScoSh. For estimating index function h with a parametric curve, we obtain $R^2 = 0.92$ on the test set, and for the non-parametric method (SVM) with different kernels (Polynomial/Rbf), we get $R^2 = 0.89$. SI-ScoSh performs best among all models, while, in contrast, SI-ScoF(FR) gives a prediction performance of merely $R^2 = 0.58$ and SI-ScoF$(\mathbb{L}^2)$ an $R^2 = 0.45$. Simpler indexed models

fail to capture these relationships – R^2's are less than 0.1 for ScoF ($\mathbb{L}^2$ & FR) and ScoSh.
The parameter estimates of the SI-ScoSh model are shown in Fig. 8.

2. **Covid Hospitalization Data**: This data[1] contains the number of daily new COVID hospitalizations at hospitals in 31 European countries, which serve as our predictors. The observation period is from January 1, 2020, to October 13, 2022, so each predictor f_i contains 1016 elements. The responses y_i are the total number of deaths in the respective countries that occurred during the observation period. Our goal is to utilize these hospitalization curves to predict the number of fatalities in a country. The data and the results are presented in Fig. 9.
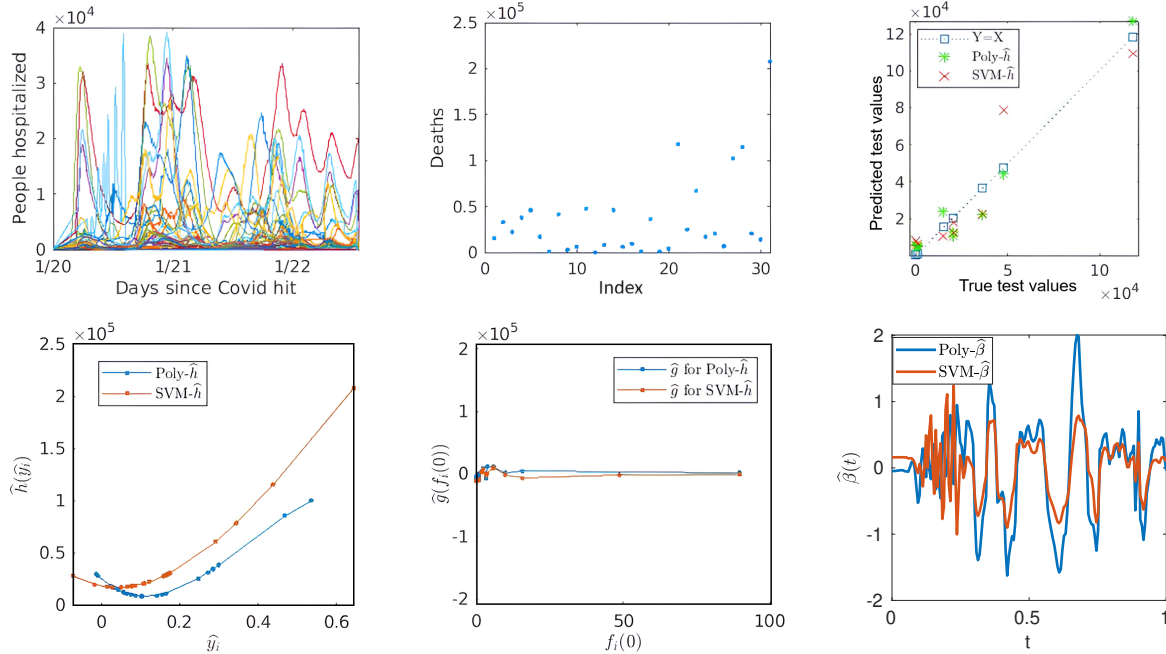

Figure 9: Covid hospitalization results – Top: the daily hospitalization curves (left), corresponding fatality counts (middle), and predicted responses versus true responses (right). Bottom: the estimated parameters under SI-ScoSh – $\widehat{h}$ (left), $\widehat{g}$ (middle), and $\widehat{\beta}$ (right).

Under the SI-ScoSh model – a quadratic g, a cubic h, and a β using the first six Fourier basis elements – provide the best performance (test set prediction $R^2 > 0.92$). The estimates in the bottom of Fig. 9 show that $\widehat{g}$ is relatively constant when compared to

25

$\widehat{h}$, indicating most correlation is captured by β and h. This is because all countries start from a point of zero hospitalizations, *i.e.*, $\{f_i(0)\}$ are all zero. Other models like SI-ScoF ($\mathbb{L}^2$ & FR), ScoSh and ScoF ($\mathbb{L}^2$ & FR) fail to capture significant relationships with prediction R^2's less than 0.2. For details, please refer to the Supplementary Material.

3. **Covid Infection Data**: This dataset[2] contains the number of new COVID-19 infections per day in each of 41 countries. These daily infection rate functions serve as the predictors. (see top left of Fig 10). The total number of people hospitalized during the entire period is the response for each country. The raw dataset has been smoothed but not centered or phase-shifted.
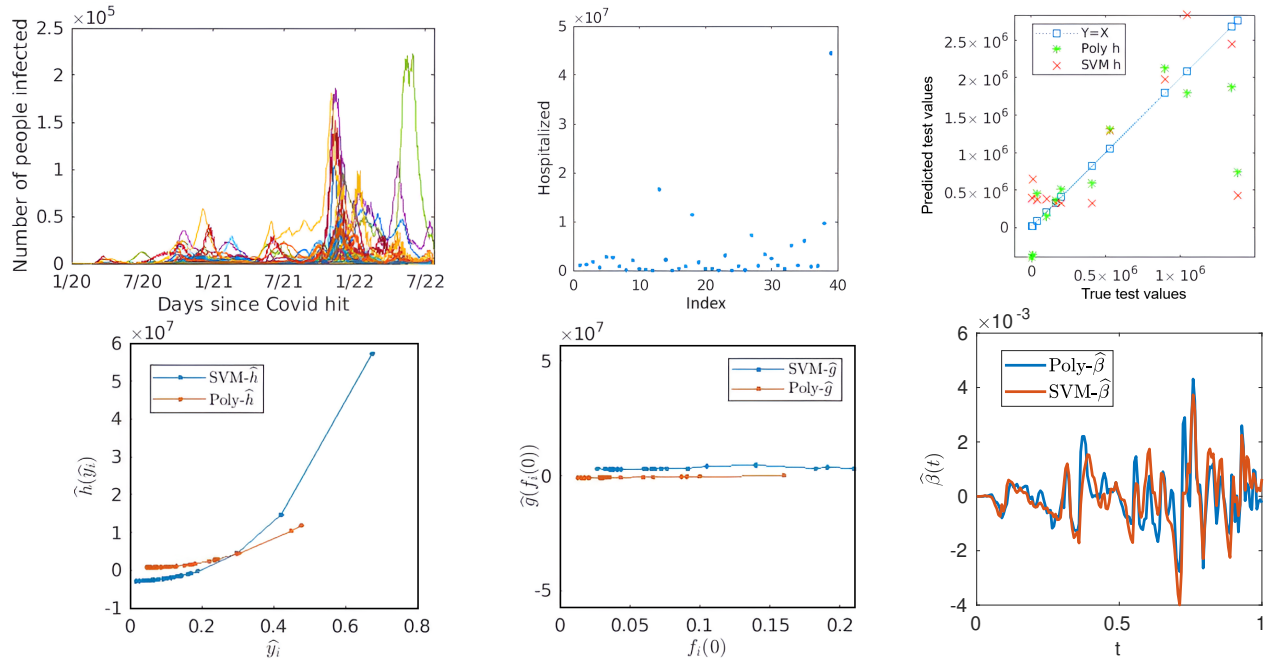

Figure 10: Covid infection results – Top: the daily infection curves (left), corresponding hospitalization counts (middle), and predicted responses versus true responses (right). Bottom: the estimated parameters under SI-ScoSh – $\widehat{h}$ (left), $\widehat{g}$ (middle), and $\widehat{\beta}$ (right).

We apply the SI-ScoSh and SI-ScoF(FR & $\mathbb{L}^2$) models and their simpler versions for a prediction performance comparison. The SI-ScoSh model predicts the test responses with $R^2 = 0.89$, but the SI-ScoF(FR) model captures a far less statistically significant

---

[2]https://ourworldindata.org/covid-hospitalizations

predictor-response relationship ($R^2 = 0.4$). Models without the index functions provide even worse performance. The $\mathbb{L}^2$ version of the ScoF model performs worse than the FR version. Like the previous example, g, the index function does not play an important role here. Please refer to the Supplementary Material for detailed results.

5   Extension to a Multiple Index Model

Following Ferraty et al. [2013], we can extend the SI-ScoSh model from a single index to a multiple index model according to:

$$y_i = \sum_{j=1}^{K} \left\{ g_j(f_i(0)) + h_j \left(\sup_{\gamma_{i,j} \in \Gamma} \langle \beta_j, q_i \star \gamma_{i,j} \rangle \right) + \epsilon_{ij} \right\} \; , \; i = 1, \cdots, n \qquad (8)$$

The estimation proceeds by treating the problem as a single-index model and estimating $\{\beta_1, h_1, g\}$. Then, we calculate the residuals and use them as responses for the next single index model, leading to the estimation of $\{\beta_2, h_2, g_2\}$. We continue until the improvement in prediction performance becomes small.

Rainfall vs Morning Humidity: We illustrate this model using a weather dataset. The predictor functions are daily humidity at 9 am every ten days over the course of the period Jan-1-2014 to Dec-31-2015 for 49 counties in Australia. The response variable for each county is the total amount of rain over the same period.The raw dataset[3] has been smoothed (with a moving average) to reduce noise. The results from the application of the multi-index ScoSch model are presented in Fig. 11.

   The results show that the first layer $\{h_1, \beta_1, g_1\}$ captures approximately a third of the correlation between shapes of the predictors and the response, but as we add more layers,

---

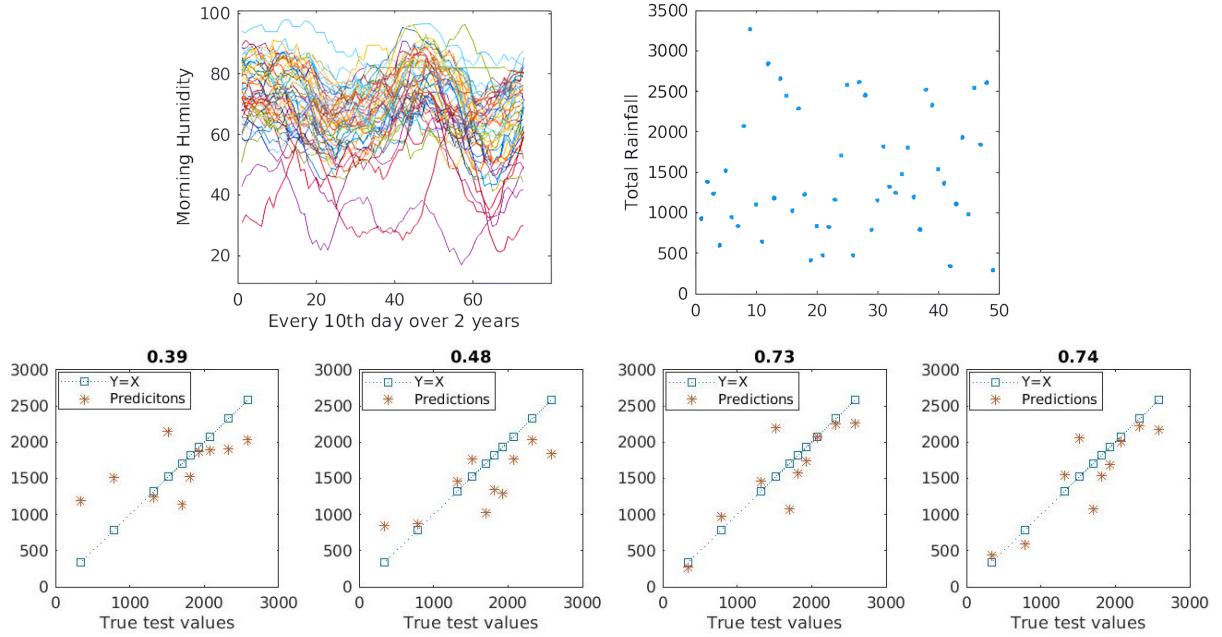[3]https://rattle.togaware.com/weatherAUS.csv

Figure 11: Top: the morning humidity curves for different counties (left) and the total rainfall in the respective counties (right). Bottom: Results from successive index models with R^2 values on top. These values improve from 0.39 for $K = 1$ to 0.74 for $K = 4$.

the prediction performance R^2 increases to around 0.74. Further addition of layers does not improve performance. This result contrasts SI-ScoF(FR), where R^2 improves less than 0.1 for each extra layer. A detailed table is presented in the Supplementary Material.

6   Conclusion

Functional data has two components: phase and shape, and they may contribute at different levels in a functional regression model. This paper develops a novel approach, termed a ScoSh model, that uses only the shapes of functions and ignores their original phases in scalar predictions. Furthermore, it optimizes the phases inside the regression models rather than as preprocessing, as is often done currently. This formalization leads to new definitions of *regression phase* and *regression mean*. The model also imposes an index function to result in a SI-ScoSh model. The two novel components - removal of dependence

on the predictor phase and using a nonlinear index function - show improved performance in various situations. Several simulated and real-data experiments demonstrate the model and its superiority.

The proposed SI-ScoSh model is appropriate when the phase components of predictors carry little or no information. This is often the case in image analysis and neuroimaging, where phases correspond to different parameterizations of neuroanatomical objects. However, in general, the phase components may contain helpful information, and discarding them would degrade prediction performance. In that situation, a more flexible model would be to separate the phases (from shapes) and use them as separate predictors themselves. This idea has been left for future explorations.

References

K. Ahn, J. D. Tucker, W. Wu, and A. Srivastava. Regression models using shapes of functions as predictors. *Comp. Statistics & Data Analysis*, 151:107017, 2020.

A. Ait-Saïdi, F. Ferraty, R. Kassa, and P. Vieu. Cross-validated estimations in the single-functional index model. *Statistics*, 42(6):475–494, 2008.

U. Amato, A. Antoniadis, and I. De Feis. Dimension reduction in functional regression with applications. *Computational Statistics & Data Analysis*, 50(9):2422–2446, 2006.

E. Boj, P. Delicado, and J. Fortiana. Distance-based local linear regression for functional predictors. *Computational Statistics & Data Analysis*, 54(2):429–437, 2010.

E. Boj, A. Caballé, P. Delicado, A. Esteve, and J. Fortiana. Global and local distance-based generalized linear models. *Test*, 25:170–195, 2016.

P. Delicado. Comments on: Shape-based functional data analysis. *TEST*, 33(1):62–65, 2024.

I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis, with Applications in R. Second Edition*. John Wiley and Sons, Chichester, 2016.

J. Du, I. L. Dryden, and X. Huang. Size and shape analysis of error-prone shape data. *Journal of the American Statistical Association*, 110(509):368–379, 2015.

P. H. C. Eilers, B. Li, and B. D. Marx. Multivariate calibration with single-index signal regression. *Chemometrics and Intelligent Lab. Systems*, 96(2):196–202, 2009.

F. Ferraty, A. Goia, E. Salinelli, and P. Vieu. Functional projection pursuit regression. *Test*, 22:293–320, 2013.

A. Ghosal, W. Meiring, and A. Petersen. Fréchet single index models for object response regression. *Electronic Journal of Statistics*, 17(1), 2023.

G. M. James and B. W. Silverman. Functional adaptive model estimation. *Journal of the American Statistical Association*, 100(470):565–576, 2005.

G. M. James, J. Wang, and J. Zhu. Functional linear regression that's interpretable. *.;*, pages 2083–2108, 2009.

D. G. Kendall, D. Barden, T. K. Carne, and H. Le. *Shape and shape theory*. Wiley, Chichester, New York, 1999.

E. R. Lee and B. U. Park. Sparse estimation in functional linear regression. *Journal of Multivariate Analysis*, 105(1):1–17, 2012.

Y. Li, N. Wang, and R. J. Carroll. Generalized functional linear models with semiparametric single-index interactions. *Journal of the American Statistical Association*, 105(490):621–633, 2010.

L. Lin, B. St. Thomas, H. Zhu, and D. B. Dunson. Extrinsic local regression on manifold-valued data. *Journal of the American Statistical Association*, 112(519):1261–1273, 2017.

L. Lin, N. Mu, P. Cheung, and D. Dunson. Extrinsic gaussian processes for regression and classification on manifolds. *Bayesian Analysis*, 14(3), 2019.

J. S. Marron, J. O. Ramsay, L. M. Sangalli, and A. Srivastava. Statistics of time warpings and phase variations. *Electronic Journal of Statistics*, 8(2):1697–1702, 2014.

J. S. Marron, J. O. Ramsay, L. M. Sangalli, and A. Srivastava. Functional Data Analysis of Amplitude and Phase Variation. *Statistical Science*, 30(4):468 – 484, 2015.

B. D. Marx and P. H. Eilers. Generalized linear regression on sampled signals and curves: a p-spline approach. *Technometrics*, 41(1):1–13, 1999.

M. W. McLean, G. Hooker, A.-M. Staicu, F. Scheipl, and D. Ruppert. Functional generalized additive models. *Journal of Computational and Graphical Statistics*, 23(1):249–269, 2014.

J. S. Morris. Functional regression. *Annual Review of Statistics and Its Application*, 2: 321–359, 2015.

A. Petersen and H.-G. Müller. Fréchet regression for random objects with euclidean predictors. *The Annals of Statistics*, 47(2):691–719, 2019.

J. O. Ramsay and B. W. Silverman. Fitting differential equations to functional data: Principal differential analysis. *Functional data analysis*, pages 327–348, 2005.

T. W. Randolph, J. Harezlak, and Z. Feng. Structured penalties for functional linear models—partially empirical eigenvectors for regression. *Electronic journal of statistics*, 6:323, 2012.

P. T. Reiss and R. T. Ogden. Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, 102(479):984–996, 2007.

X. Shi, M. Styner, J. Lieberman, J. G. Ibrahim, W. Lin, and H. Zhu. Intrinsic regression models for manifold-valued data. In *International conference on medical image computing and computer-assisted intervention*, pages 192–199. Springer, 2009.

H.-Y. Shin and H.-S. Oh. Robust geodesic regression. *International Journal of Computer Vision*, 130(2):478–503, 2022.

A. Srivastava and E. P. Klassen. *Functional and shape data analysis*, volume 1. Springer, 2016.

A. Stoecker, L. Steyer, and S. Greven. Functional additive models on manifolds of planar shapes and forms. *Journal of Computational and Graphical Statistics*, 32(4):1600–1612, 2023.

P. Thomas Fletcher. Geodesic regression and the theory of least squares on riemannian manifolds. *International journal of computer vision*, 105:171–185, 2013.

D. Tsagkrasoulis and G. Montana. Random forest regression for manifold-valued responses. *Pattern Recognition Letters*, 101:6–13, 2018.

J. D. Tucker, W. Wu, and A. Srivastava. Generative models for functional data using phase and amplitude separation. *Computational Statistics & Data Analysis*, 61:50–66, 2013.

Y. Wu, C. Huang, and A. Srivastava. Shape-based functional data analysis. *Test*, 33(1): 1–47, 2024.

Z. Zhang, E. Klassen, and A. Srivastava. Phase-amplitude separation and modeling of spherical trajectories. *Journal of Computational and Graphical Statistics*, 27(1):85–97, 2018.

Y. Zhao, R. T. Ogden, and P. T. Reiss. Wavelet-based lasso in functional linear regression. *Journal of comp. and graphical statistics*, 21(3):600–617, 2012.