Joint Registration and Conformal Prediction for Partially Observed Functional Data

Fangyi Wang, Sebastian Kurtek and Yuan Zhang
Department of Statistics, The Ohio State University

Abstract

Predicting missing segments in partially observed functions is challenging due to infinite-dimensionality, complex dependence within and across observations, and irregular noise. These challenges are further exacerbated by the existence of two distinct sources of variation in functional data, termed amplitude (variation along the y-axis) and phase (variation along the x-axis). While registration can disentangle them from complete functional data, the process is more difficult for partial observations. Thus, existing methods for functional data prediction often ignore phase variation. Furthermore, they rely on strong parametric assumptions, and require either precise model specifications or computationally intensive techniques, such as bootstrapping, to construct prediction intervals. To tackle this problem, we propose a unified registration and prediction approach for partially observed functions under the conformal prediction framework, which separately focuses on the amplitude and phase components. By leveraging split conformal methods, our approach integrates registration and prediction while ensuring exchangeability through carefully constructed predictor-response pairs. Using a neighborhood smoothing algorithm, the framework produces pointwise prediction bands with finite-sample marginal coverage guarantees under weak assumptions. The method is easy to implement, computationally efficient, and suitable for parallelization. Numerical studies and real-world data examples clearly demonstrate the effectiveness and practical utility of the proposed approach.

Keywords: phase variation, split conformal prediction, square-root slope transform, neighborhood smoothing

1 Introduction

Accurate prediction of future trajectories given historical functional data plays a crucial role in many real-world applications. For instance, given previous (complete) daily observations of traffic flow rate, and a partial observation (up to some time) for a new day, prediction of traffic flow rate for the rest of that day can help optimize transportation networks and reduce congestion during rush hour [Chiou, 2012]. Similarly, given complete historical daily maximum temperature measurements at a particular location, one may be interested in a daily maximum temperature forecast for the rest of the current year, which can help mitigate societal risks related to extremely cold weather. Both of these examples are considered in Section 5. There are many other applications where such prediction problems are of primary interest, including growth rate prediction for people or other natural objects, forecasting of pollutant density, etc. These predictions can be used to anticipate future trends and help inform biomedical, environmental and social decisions.

Functional data is challenging to analyze, because it is inherently infinite-dimensional, exhibits complex dependencies within and potentially across observations, and is often observed at discrete time points with heteroscedastic noise. Functional observations encompass two sources of variation, one along the x-axis called *phase* and one along the y-axis called *amplitude*. A common practice during statistical analyses is to first separate amplitude and phase variations. This is done via a process called *registration*, which "warps" the functions' domains to align their key features with respect to time [Kneip and Gasser, 1992, Ramsay and Silverman, 2005, Srivastava et al., 2011]. However, registration is often treated as a pre-processing step, with subsequent statistical modeling and inference applied to "well-aligned" data. Thus, registration uncertainty is often ignored in downstream tasks. Furthermore, when the data contains partial observations, registration is much more

challenging since the domains of partial and complete functions are fundamentally different [Bauer et al., 2021, Bryner and Srivastava, 2021].

To tackle the aforementioned challenges, we propose a framework for prediction of partially observed functions wherein registration is directly incorporated. Comparing to a sequential approach of "registration then prediction", our approach bypasses the need for registration of partial observations to complete ones, and propagates registration uncertainty to the prediction step. We adapt *conformal prediction*, a distribution-free uncertainty quantification method that provides valid finite-sample coverage validity [Vovk et al., 2005, Lei et al., 2018]. This requires a careful construction of exchangeable predictor-response data pairs, which are not naturally defined in our context. In addition, we employ a nonparametric, neighborhood smoothing prediction algorithm [Zhang et al., 2017], which offers flexibility without strong modeling assumptions on the relationship between predictors and responses. Thus, our main contributions are as follows.

1. We propose a novel framework that integrates registration into conformal prediction of the amplitude component for partial functional data. This results in more accurate prediction intervals, as compared to procedures that do not utilize registration, when functional data contains phase variation. We additionally define a separate conformal prediction procedure for the phase component, which incorporates a monotonicity constraint. Our approach builds on the elastic functional data analysis (EFDA) framework for registration [Srivastava et al., 2011, Srivastava and Klassen, 2016].

2. Our method offers distribution-free uncertainty quantification with a finite-sample guarantee. For conformal prediction of partially observed functions, we carefully construct predictors and responses from raw data that maintain exchangeability throughout the procedure, and leverage ideas from split conformal methods to in-

tegrate registration. Unlike traditional parametric approaches, our method imposes fairly weak assumptions on the data generating process. We employ a nonparametric neighborhood smoothing prediction algorithm that provides more flexibility than competitors such as functional linear regression.

3. Compared to nonparametric bootstrap or Bayesian uncertainty quantification, our approach is much faster, making it well-suited for large-scale applications.

1.1 Related work

There exist many approaches for functional data prediction, albeit none of them consider the challenges posed by phase variation. They include functional regression [Chiou, 2012, Maity, 2017], time series models [Didericksen et al., 2012, Leroux et al., 2018], neural networks [Yin et al., 2021], or a combination of different aforementioned methods [Jiao et al., 2023]. For uncertainty quantification, these approaches generally use closed-form confidence intervals based on variance estimates [Leroux et al., 2018], or nonparametric bootstrap residual sampling [Chiou, 2012, Jiao et al., 2023]. However, these methods require strong parametric assumptions, and the resulting prediction intervals' coverage rate converges to the desired nominal level asymptotically with no finite-sample guarantee. Matuk et al. [2022] and Earls and Hooker [2017] define Bayesian models for joint registration and prediction, where prediction intervals are constructed from posterior samples. Nonetheless, the high computational cost of these approaches hinders their application to large datasets.

Conformal prediction provides prediction intervals that have a finite-sample coverage guarantee without imposing assumptions on the data generating process [Saunders et al., 1999, Vovk et al., 2005]. It is generally computationally more efficient when compared to Bayesian methods which utilize Markov chain Monte Carlo to sample form the poste-

4

rior distribution. Since being introduced, conformal prediction has been widely applied in a variety of scenarios, ranging from regression and classification for multivariate data [Lei et al., 2018, Cauchois et al., 2021] to prediction tasks involving more complex data structures, e.g., survival analysis [Candès et al., 2023, Gui et al., 2024], time series [Zaffran et al., 2022, Angelopoulos et al., 2023], and matrix data [Gui et al., 2023, Shao and Zhang, 2023]. However, relatively few works have focused on the application of conformal prediction to functional data. Lei et al. [2015] use a basis projection for functional observations and obtain prediction sets for basis coefficients; they do not consider partially observed functions. Diquigiovanni et al. [2022] and Ajroldi et al. [2023] apply conformal prediction to multivariate functional data and functional time series using a nonconformity score based on functional depth. Recently, Diana et al. [2023] and De Magistris et al. [2024] apply conformal prediction to spatial functional data for uncertainty quantification. All of the aforementioned methods treat registration as a pre-processing step, which can affect prediction accuracy when the data contains phase variation.

The rest of the article is organized as follows. Section 2 discusses preliminaries including full conformal prediction for partial functional data without registration and a brief introduction of the elastic functional data analysis (EFDA) framework. Section 3 defines the joint registration and prediction method. Comprehensive simulations and real-world data examples are presented in Sections 4 and 5 to demonstrate the validity and efficiency of the proposed method. We conclude with a brief discussion in Section 6. Lemmas and proofs are included in appendices.

2 Preliminaries

2.1 Full functional conformal prediction without registration

Consider $f_1, \ldots, f_{n+1} \overset{iid}{\sim} \mathbb{P}_{\mathcal{F}}$, where $\mathbb{P}_{\mathcal{F}}$ is a probability distribution on the function space $\mathcal{F}$, where $\mathcal{F} = \{f : [0,1] \to \mathbb{R} | f \text{ is absolutely continuous}\}$. We use $[0,1]$ as the function domain without loss of generality. The observed portion of the $(n+1)^{\text{th}}$ sample is denoted by $f_{n+1}^{\mathcal{J}}$, where $\mathcal{J} \subseteq [0,1]$. For simplicity, we assume $\mathcal{J} = [0, U]$, where $U \sim \pi_u$ is a random stopping point and π_u is a probability distribution on $[0,1]$. This can be generalized to other observation patterns (see Section 3.2). Given $f_1, \ldots, f_n$ and $f_{n+1}^{\mathcal{J}}$, our goal is to construct pointwise prediction intervals for f_{n+1} on a fixed, uniform grid of time points, $\mathcal{T} := \{t_1, \ldots, t_T\}$, where $t_1 = 0$, $t_T = 1$ and $t_{k+1} - t_k = 1/(T-1)$, $k = 1, \ldots, T-1$.

For each time point $t \in \mathcal{T}$ and any user specified significance level $\alpha \in (0,1)$, we want the prediction interval $\mathcal{I}_t$ for $f_{n+1}(t)$ to satisfy finite-sample coverage validity,

$$\mathbb{P}(f_{n+1}(t) \in \mathcal{I}_t) \geq 1 - \alpha. \tag{1}$$

In conformal prediction, (1) is guaranteed through three key principles [Vovk et al., 2005]:

1. **exchangeable** predictor-response data pairs $((X_1, Y_1), \ldots, (X_{n+1}, Y_{n+1})) \sim \mathbb{P}_{\mathcal{X} \times \mathcal{Y}}$, i.e., given $((X_1, Y_1), \ldots, (X_n, Y_n))$ as well as X_{n+1}, the aim is to predict Y_{n+1};

2. a **permutation symmetric algorithm** that inputs augmented data $((X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, y))$ and fits model $\hat{\mu}_y : \mathcal{X} \to \mathcal{Y}$, i.e., $\hat{\mu}((x_{\pi(1)}, y_{\pi(1)}), \ldots, (x_{\pi(n)}, y_{\pi(n)})) = \hat{\mu}((x_1, y_1), \ldots, (x_n, y_n))$ for all permutations $\pi : [1:n] \leftrightarrow [1:n]$;

3. a **nonconformity score** that measures the discrepancy of each observation relative to the fit, e.g., the absolute residual $S_i := |Y_i - \hat{\mu}_y(X_i)|$.

To apply conformal prediction to partially observed functional data, we need to construct exchangeable $\{(X_i, Y_i)\}$, as there are no natural predictor and response variables. The prediction target $f_{n+1}(t)$ is analogous to the response Y_{n+1}, and thus, we define $Y_i(t) := f_i(t)$, $i = 1, \ldots, n+1$, i.e., we are predicting f_{n+1} at time t, $\forall\, t \in \mathcal{T}$. Since we observe $f_{n+1}^{\mathcal{J}}$, it can be viewed as the new feature X_{n+1}. To define $X_1, \ldots, X_n$ that are exchangeable with X_{n+1}, we cut $f_1, \ldots, f_n$ at $t = U$, and set $X_i := f_i^{\mathcal{J}}$, $i = 1, \ldots, n+1$. Since $f_1, \ldots, f_{n+1}$ are i.i.d. samples from $\mathbb{P}_{\mathcal{F}}$, $Y_1(t), \ldots, Y_{n+1}(t)$ are exchangeable (note that evaluating each function f_i at time t is deterministic, and therefore preserves exchangeability). On the other hand, there are two sources of randomness in X_{n+1}: f_{n+1} and $U \sim \pi_u$. To ensure exchangeability of predictors, we introduce Assumption 1, which can be interpreted as Missing Completely at Random (MCAR) [Rubin, 1976].

Assumption 1. *U is independent of f_i, $\forall\, i = 1, \ldots, n+1$.*

Assumption 1 ensures joint exchangeability of $\{(X_i, Y_i(t))\}$ for all $t \in \mathcal{T}$. This assumption is fairly mild and holds in many real-world data scenarios where the truncation time point does not depend on the observation process, e.g., monitoring of environmental measurements.

With exchangeable $\{(X_i, Y_i(t))\}$ and a properly chosen symmetric algorithm and non-conformity score, we can implement Full Functional Conformal Prediction (`FFCP`) as follows. For each $y \in \mathcal{Y}_{trial}$, a set of trial values for $Y_{n+1}(t)$, we set $Y_{n+1}(t) = y$ and fit $\hat{\mu}_y$ to the augmented data $((X_1, Y_1(t)), \ldots, (X_n, Y_n(t)), (X_{n+1}, y))$. We then compute $S_i = |Y_i(t) - \hat{\mu}_y(X_i)|$ for $i = 1, \ldots, n+1$ and include y in the prediction interval $\mathcal{I}_t$ if S_{n+1} is within the $1 - \alpha$ quantile of the empirical distribution of $\{S_1, \ldots, S_{n+1}\}$, i.e., $\mathcal{I}_t := \{y : |y - \hat{\mu}_y(X_{n+1})| = S_{n+1} \leq \mathcal{Q}_{1-\alpha}(\{S_1, \ldots, S_{n+1}\})\}$, where $\mathcal{Q}_\beta(\cdot)$ denotes the lower β quantile function. We repeat this procedure for each $t \in \mathcal{T}$, and get the set of prediction intervals $\mathcal{I}_{t_1}, \ldots, \mathcal{I}_{t_T}$ for $f_{n+1}(t_1), \ldots, f_{n+1}(t_T)$, where each prediction interval $\mathcal{I}_{t_k}$ satisfies
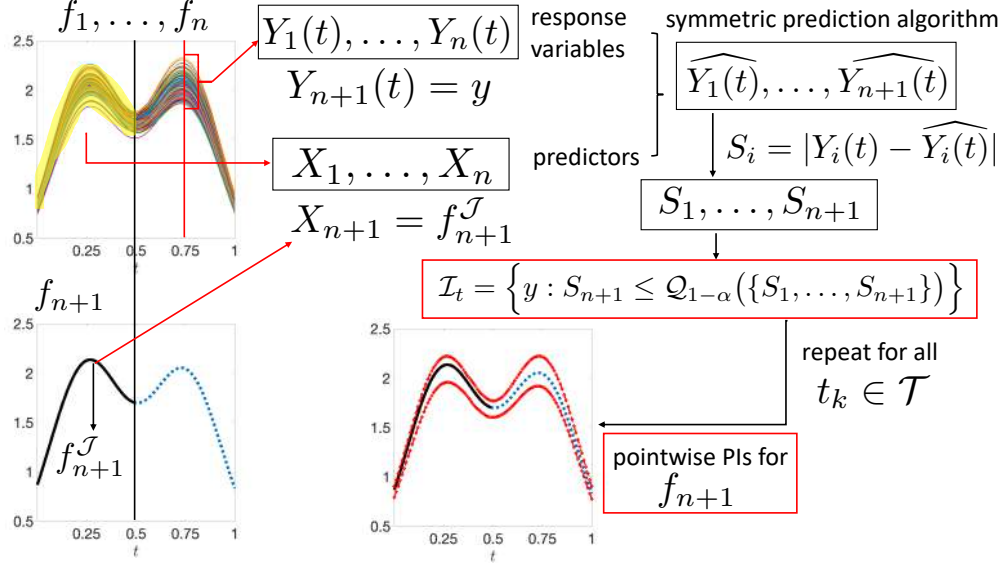
7

Figure 1: Full Functional Conformal Prediction (`FFCP`) algorithm.

the coverage validity in (1). The specific choice of the symmetric algorithm, which plays a critical role in our functional conformal prediction framework, will be discussed in detail in Section 3.1. The full `FFCP` algorithm is presented in Figure 1.

To provide a clearer understanding of the practical performance of `FFCP`, we present a small simulation study. We consider two-peak functions without and with phase variation (see examples in top and bottom of Figure 2(a), respectively). To assess the performance of `FFCP`, we use $B = 1000$ Monte Carlo samples with $n = 100$, truncation time point $U = 0.5$ for f_{n+1}, and $1 - \alpha = 0.9$. Evaluation is based on two criteria,

1. **coverage validity**: empirical pointwise coverage rate

$$p_k = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\left\{ f_{n+1}^{(b)}(t_k) \in \mathcal{I}_{t_k}^{(b)} \right\}, \ k = 1, \ldots, T; \tag{2}$$

2. **prediction accuracy**: pointwise average prediction interval (PI) length

$$\ell_k = \frac{1}{B} \sum_{b=1}^{B} \text{length}(\mathcal{I}_{t_k}^{(b)}), \ k = 1, \ldots, T. \tag{3}$$

Results for functional data without (with) phase variation are given in the top (bottom) row of Figure 2. Panels (b) and (c) show pointwise coverage rates p_k, with pointwise
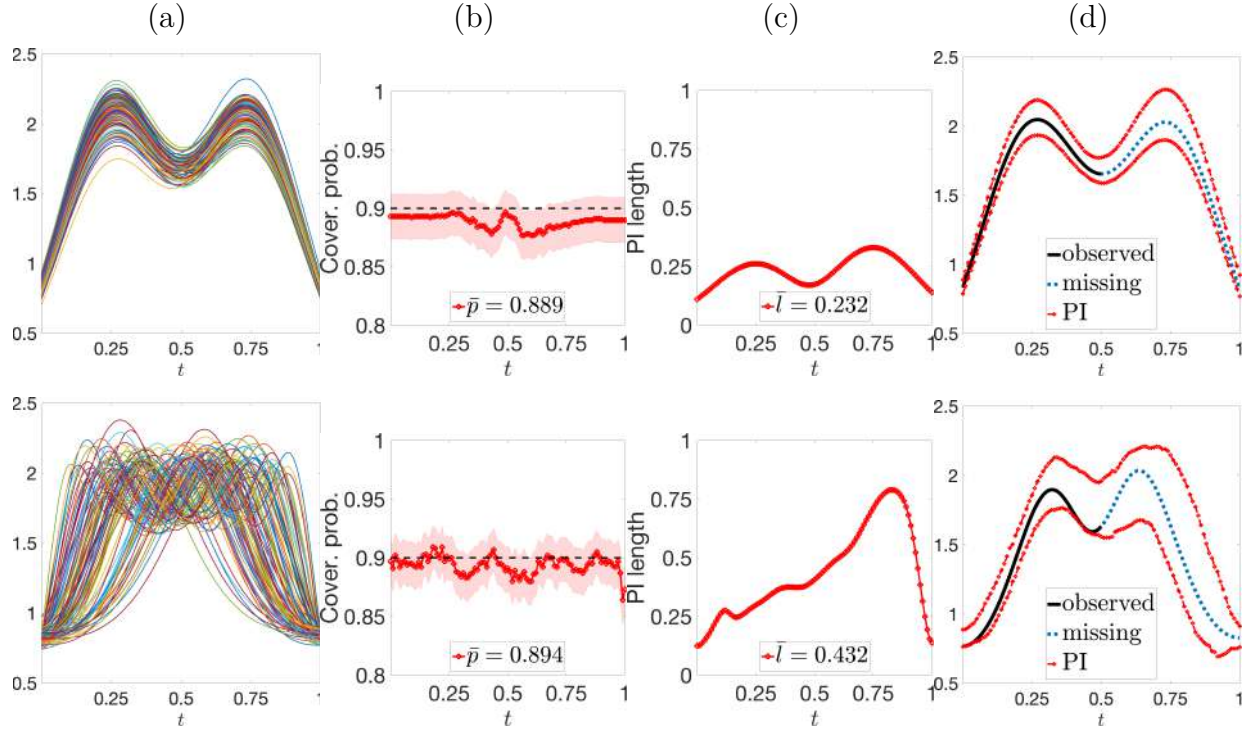
8

Figure 2: FFCP for two-peak functions without (row 1) and with (row 2) phase variation. (a) Data. (b)&(c) Coverage rates p_k with 95% CIs (shaded region) and average PI lengths ℓ_k, respectively. (d) Partial observation $f_{n+1}^{\mathcal{J}}$ (black), ground truth $f_{n+1}^{[0,1]\backslash\mathcal{J}}$ (dashed blue), and pointwise PIs (red).

95% confidence intervals (CIs) (shaded region), and pointwise average PI lengths ℓ_k for $k = 1, \ldots, T$, respectively. The time-averaged coverage rate $\bar{p} = \sum_{k=1}^{T} p_k/T$ and PI length $\bar{\ell} = \sum_{k=1}^{T} \ell_k/T$ are also reported. We see that FFCP guarantees finite-sample coverage validity for any $t \in \mathcal{T}$ for functions without and with phase variation. However, pointwise PI lengths tend to be much larger when phase variation is present in the data. Panel (d) demonstrates pointwise prediction intervals $\mathcal{I}_{t_1}, \ldots, \mathcal{I}_{t_T}$ for $f_{n+1}^{(b)}$ for a randomly chosen Monte Carlo sample $b \in \{1, \ldots, B\}$. We show the partial observation $f_{n+1}^{\mathcal{J}}$ (black), the ground truth missing segment $f_{n+1}^{[0,1]\backslash\mathcal{J}}$ (dashed blue), and the lower and upper boundaries of the pointwise PIs (red). When there is no phase variation in the data, the PIs form a prediction band that accurately captures the geometric features, i.e., two peaks and

9

one valley, of the underlying function f_{n+1}. However, when phase variation is present, the prediction band is less effective at capturing such geometric features, since their timing varies considerably across observations. As a result, $Y_i(t)$ fails to provide useful information for predicting $Y_{n+1}(t)$. These results indicate that FFCP is better suited for functional data without phase variation. To improve FFCP in the presence of phase variation, we will incorporate registration, based on the elastic functional data analysis (EFDA) framework.

2.2 Registration via EFDA

Phase variation in functional data is characterized through *warping functions*, which are elements of $\Gamma := \{\gamma : [0,1] \to [0,1] | \gamma(0) = 0, \gamma(1) = 1, \dot\gamma > 0\}$. Given $\gamma \in \Gamma$, the domain warping of a function $f \in \mathcal{F}$ is given by composition: $f \circ \gamma$. To register a function f_2 with respect to another function f_1, we need a distance $d(\cdot, \cdot)$ that is preserved under simultaneous warping: $d(f_1, f_2) = d(f_1 \circ \gamma, f_2 \circ \gamma)$. It is well-known that the standard $\mathbb{L}^2$ distance, given by $d_2(f_1, f_2) := \left(\int_0^1 (f_1(t) - f_2(t))^2 dt \right)^{1/2}$, does not satisfy this property. To address this, one can instead use (the extension of) the Fisher-Rao (FR) Riemannian metric that is preserved under simultaneous warping. However, the resulting Riemannian distance cannot be computed in closed form. Srivastava et al. [2011] proposed the square-root slope function (SRSF) that reduces the FR metric to the $\mathbb{L}^2$ metric. For a function $f \in \mathcal{F}$, its SRSF $q : [0,1] \to \mathbb{R}$ is given by $Q(f) = q := \text{sign}(\dot f)\sqrt{|\dot f|}$, where $Q : \mathcal{F} \to \mathcal{Q}$. Given $f(0)$, the inverse of Q is given by $Q^{-1}(q, f(0))(t) = f(0) + \int_0^t q(s)|q(s)|ds$. The space of SRSFs $\mathcal{Q}$ is a subset of $\mathbb{L}^2([0,1], \mathbb{R})$. Finally, the domain warping of a function f by γ, $f \circ \gamma$, is given by the following transformation of its SRSF: $(q \circ \gamma)\sqrt{\dot\gamma}$. Thus, given two functions f_1, $f_2 \in \mathcal{F}$ and their SRSFs q_1, $q_2 \in \mathcal{Q}$, $d_{FR}(f_1, f_2) = d_2(q_1, q_2) = d_2((q_1 \circ \gamma)\sqrt{\dot\gamma}, (q_2 \circ \gamma)\sqrt{\dot\gamma})$.

Given these results, pairwise registration of f_2 to f_1, with corresponding SRSFs q_2 and

q_1, is given by the optimization problem

$$\gamma^* = \underset{\gamma \in \Gamma}{\arg\min}\, d_2\big(q_1, (q_2 \circ \gamma)\sqrt{\dot{\gamma}}\big), \tag{4}$$

which can be solved using dynamic programming [Robinson, 2012]. Then, $f_2 \circ \gamma^*$ is registered to f_1, and the amplitude distance between them is given by $d_a(f_1, f_2) = d_2(q_1, (q_2 \circ \gamma^*)\sqrt{\dot{\gamma^*}})$; γ^* is the relative phase of f_2 with respect to f_1. Multiple registration of functions $f_1, \ldots, f_n$ is performed pairwise using (4) with respect to a suitable template function. Let $q_1, \ldots, q_n$ be the SRSFs of $f_1, \ldots, f_n$, and define the *sample Karcher mean* as the template, which is given by

$$\bar{q} := \underset{q \in \mathcal{Q}}{\arg\min} \sum_{i=1}^{n} \underset{\gamma_i \in \Gamma}{\min}\, d_2^2(q, (q_i \circ \gamma_i)\sqrt{\dot{\gamma_i}}). \tag{5}$$

See Algorithms 2 and 3, and Section 3.4 in Srivastava et al. [2011] for details of the sample Karcher mean computation, and the complete multiple registration algorithm. The corresponding Karcher mean in $\mathcal{F}$ is $\bar{f} := Q^{-1}(\bar{f}(0), \bar{q})$, where $\bar{f}(0) = \frac{1}{n}\sum_{i=1}^{n} f_i(0)$. Multiple registration results in (i) $\{\gamma_i^*\}$, the relative phases with respect to the sample Karcher mean, and (ii) $\{f_i \circ \gamma_i^*\}$, the amplitudes of $\{f_i\}$.

3   Joint registration and prediction for functional data

Next, we incorporate registration into conformal prediction when functional data exhibits phase variation. In this context, the prediction target changes to the amplitude $f_{n+1} \circ \gamma_{n+1}^*$, where $\gamma_{n+1}^* \in \Gamma$ registers f_{n+1} to a template function. Thus, we define new response variables $Y_i(t) := (f_i \circ \gamma_i^*)(t)$ for each $t \in \mathcal{T}$, while keeping the predictors $X_i = f_i^{\mathcal{J}}$ unchanged. However, a critical challenge now is the choice of template function for registration that preserves exchangeability of $(X_i, Y_i(t))$.

One simple choice is to randomly select a function from the complete observations $f_1, \ldots, f_n$ as the template, e.g., f_1. However, this would break exchangeability of $(X_i, Y_i(t))$,

since the registration process would treat f_1 differently from $f_2, \ldots, f_{n+1}$. Instead, we will choose a template function that is independent of $f_1, \ldots, f_{n+1}$. This will be accomplished using the split conformal prediction approach. Specifically, we randomly split $f_1, \ldots, f_n$ into two independent training and calibration sets of sizes n_1 and $n_2 = n - n_1$, respectively. We first use the training set $f_1, \ldots, f_{n_1}$ to compute the sample Karcher mean $\bar{f}$ using (5) and the inverse SRSF mapping Q^{-1}. We then register each function in the calibration set, $f_{n_1+1}, \ldots, f_n$, to $\bar{f}$ by computing $\gamma_1^*, \ldots, \gamma_{n_2}^*$ via (4) (γ_i^* is the relative phase of f_{n_1+i} with respect to $\bar{f}$). Finally, we use the amplitudes $\tilde{f}_i := f_{n_1+i} \circ \gamma_i^*$ to construct the response variables $Y_i(t) = \tilde{f}_i(t)$, $i = 1, \ldots, n_2$. The response corresponding to f_{n+1} is $Y_{n_2+1}(t) = (f_{n+1} \circ \gamma_{n_2+1}^*)(t)$. This procedure maintains exchangeability of $(X_i, Y_i(t))$, as permutation symmetry is preserved when the template function is computed using an independent training set. After constructing exchangeable $(X_i, Y_i(t))$, they can be used as the input of FFCP described in Section 2.1. We call this procedure Split Functional Conformal Prediction (SFCP) and present it in Figure 3. Like FFCP, SFCP also provides a marginal finite-sample coverage guarantee, which is formally stated in Theorem 1. The proof of Theorem 1 is provided in Appendix A.

Theorem 1. *Under Assumption 1, for any $T \geq 2$, $0 < n_1 < n$, $\alpha \in (0, 1)$ and each $t_k \in \mathcal{T}$, the prediction set $\mathcal{I}_{t_k}$ from SFCP satisfies $1 - \alpha \leq \mathbb{P}\Big((f_{n+1} \circ \gamma_{n_2+1}^*)(t_k) \in \mathcal{I}_{t_k} \Big) \leq 1 - \alpha + n_2^{-1}$.*

Notice that the relationship between FFCP and SFCP is slightly different from the relationship between classical full and split conformal prediction. The split conformal method was originally developed to improve computational efficiency of the full conformal method. In our case, however, the training set is used to estimate the sample Karcher mean, which acts as the template for registration of the calibration set, as well as the target function. The goal is not to speed up computation, but rather to construct new response variables
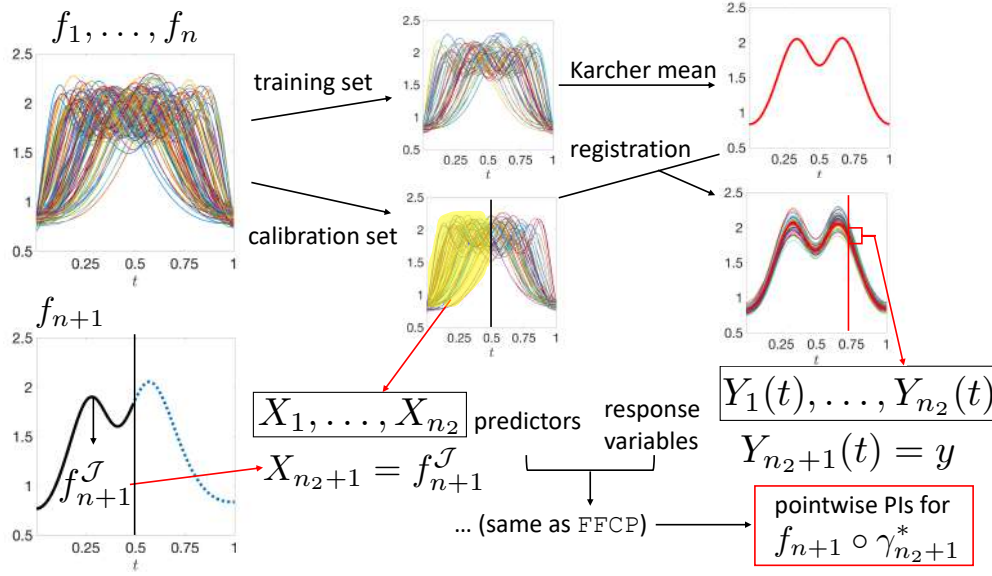
12

Figure 3: Split Functional Conformal Prediction (SFCP) algorithm.

that improve prediction accuracy in the presence of phase variation. Both FFCP and SFCP are suitable prediction methods for functional data with or without phase variation as they both guarantee finite-sample coverage. However, failing to account for phase variation generally inflates variance [Marron et al., 2015], and as a result, impairs prediction; see the simulation example in Section 2.1. Thus, we advocate the use of SFCP rather than FFCP for functional data with phase variation.

3.1 Symmetric prediction algorithm: neighborhood smoothing

A permutation symmetric algorithm is required to fit $\hat{\mu}_y$ to the augmented data composed of functional predictors X_i and scalar responses $Y_i(t) \in \mathbb{R}$. One choice is linear scalar-on-function regression, for each $t \in \mathcal{T}$, which satisfies permutation symmetry. But, it imposes strong assumptions on the relationship between predictors and responses, and the error, which might not hold for complex functional data. Here, we adapt the *neighborhood smoothing* technique from Zhang et al. [2017], which is a renaissance of the Nadaraya-Watson estimator [Nadaraya, 1964, Watson, 1964] for non-conventional data types. We

13

later find that a preexisting work [Ferraty et al., 2007] bears the similar spirit. It enforces minimal assumptions on the predictor-response relationship. Specifically, the neighborhood smoothing estimator for $Y_i(t)$ is given by

$$\widehat{Y_i(t)} := \frac{\sum_{i' \neq i} K(h^{-1}d(X_i, X_{i'}))Y_{i'}(t)}{\sum_{i' \neq i} K(h^{-1}d(X_i, X_{i'}))}, \quad i = 1, \ldots, n+1, \tag{6}$$

where $K(\cdot)$ is a kernel function, e.g., triangular or Gaussian kernel, h is a bandwidth parameter, and $d(X_i, X_{i'})$ is a distance between functions X_i and $X_{i'}$. Higher weights are assigned to $Y_{i'}$ for predictors $X_{i'}$ that are closer to X_i as measured via distance $d(\cdot, \cdot)$; h controls the weights' concentration. This approach is computationally efficient, because (6) can be computed for all $t \in \mathcal{T}$ using a one-pass evaluation of a symmetric distance matrix $D \in \mathbb{R}^{(n+1) \times (n+1)}$ with $D_{i,j} := d(X_i, X_j)$ [Shao and Zhang, 2023]. This prediction algorithm can be used in both FFCP and SFCP. In our implementation, we use the Gaussian kernel for $K(\cdot)$. In addition, we must choose the distance $d(\cdot, \cdot)$ between functional predictors and bandwidth parameter h in the kernel, both of which can affect prediction accuracy.

Choice of distance. The $\mathbb{L}^2$ distance $d_2(\cdot, \cdot)$ is most commonly used in functional data analysis. Another choice is the FR Riemannian distance $d_{FR}(\cdot, \cdot)$ as defined in Section 2.2. These two choices are computationally efficient when evaluating the distance matrix D. Alternatively, we can use the amplitude distance $d_a(\cdot, \cdot)$ introduced in Section 2.2, which registers predictors thus potentially leading to better prediction accuracy, but is slower to compute due to optimization over Γ.

Choice of bandwidth parameter. To tune the bandwidth parameter, we adapt the method of Liang et al. [2024], which ensures coverage validity while selecting a model that minimizes PI length. Specifically, we run SFCP with neighborhood smoothing as the prediction method for a set of candidate bandwidth values $h \in \mathcal{H}$ and obtain prediction intervals $\mathcal{I}_{t_k}^h$ for $k = 1, \ldots, T$. We can set $\mathcal{H}$ to be a grid of fixed values, or the lower β quantile

14

of the empirical distribution of the distances $\{D_{i,j}\}_{1 \le i < j \le n+1}$, for a grid of $\beta \in (0,1)$. We select the optimal bandwidth h^* (i) globally, $h^* := \arg\min_{h \in \mathcal{H}} \frac{1}{T} \sum_{k=1}^{T} \text{length}(\mathcal{I}_{t_k}^h)$, or (ii) locally, $h_{t_k}^* := \arg\min_{h \in \mathcal{H}} \text{length}(\mathcal{I}_{t_k}^h)$, $k = 1, \dots, T$.

3.2 Other observational regimes for partial functional data

`FFCP` and `SFCP` are not limited to settings where a single continuous segment of f_{n+1} on $\mathcal{J} = [0, U]$ is observed. The approaches can be trivially generalized to the case of $\mathcal{J} = [U_1, U_2]$, where $U_1, U_2 \sim \pi_{\boldsymbol{u}}$, $0 \le U_1 < U_2 \le 1$ and $\pi_{\boldsymbol{u}}$ is a probability density on $[0,1]^2$. Two other generalizations apply to fragmented and sparse functional data.

Fragmented. The observed portions of f_{n+1} are spread across random disjoint subintervals, $\mathcal{J} = \cup_{j=1}^{J} \mathcal{J}_j$, where each $\mathcal{J}_j = [U_{j,1}, U_{j,2}]$. This is often seen in applications involving segmented or intermittent functional observations, such as X-ray measurements of bone mineral density [Bachrach et al., 1999]. In this case, the predictors are $X_i = \boldsymbol{f}_i := (f_i^{\mathcal{J}_1}, \dots, f_i^{\mathcal{J}_J})$, $i = 1, \dots, n+1$. One can then define a distance for neighborhood smoothing as $d_{\text{prod}}(X_i, X_{i'}) := \sum_{j=1}^{J} \lambda_j d(f_i^{\mathcal{J}_j}, f_{i'}^{\mathcal{J}_j})$, where λ_j are weights with $\sum_{j=1}^{J} \lambda_j = 1$. The simplest choice is $\lambda_j = 1/J \ \forall \ j$. Alternatively, one can choose λ_j to be proportional to the length of each subinterval j.

Sparse. Here, f_{n+1} is observed at a set of discrete time points, $\mathcal{J} = \{t_1, \dots, t_{N_t}\} \subseteq [0,1]$. Such observations are common when full continuous measurements are difficult or impractical to collect, e.g., longitudinal data in clinical trials [Yao et al., 2005]. Now, the predictors are $X_i := (f_i(t_1), \dots, f_i(t_{N_t})) \in \mathbb{R}^{N_t}$, and we use the Euclidean distance in $\mathbb{R}^{N_t}$ for neighborhood smoothing: $d_e(X_i, X_{i'}) := \sqrt{\sum_{k=1}^{N_t} (f_i(t_k) - f_{i'}(t_k))^2}$.

3.3 Conformal prediction of relative phase

SFCP focuses on predicting the amplitude component $f_{n+1} \circ \gamma_{n_2+1}^*$, but it is also useful for prediction of the relative phase $\gamma_{n_2+1}^*$. Phase prediction allows assessment of uncertainty in the relative timing of amplitude features of f_{n+1} with respect to the sample Karcher mean $\bar{f}$. We use the same functional predictors as before, $X_i = f_{n_1+i}^{\mathcal{J}}$, $i = 1, \ldots, n_2 + 1$, which include amplitude and phase variation. However, we cannot predict the target function $\gamma_{n_2+1}^*$ pointwise at each $t \in \mathcal{T}$ independently, because the resulting PIs $\{\mathcal{I}_{t_k}\}$ would not be monotonically increasing (recall that $\dot{\gamma}(t) > 0 \; \forall \; t$). Instead, we predict $\gamma_{n_2+1}^*$ jointly for all $t \in \mathcal{T}$, by considering $\boldsymbol{Y}_{n_2+1} = (\gamma_{n_2+1}^*(t_1), \ldots, \gamma_{n_2+1}^*(t_T)) \in \mathbb{R}^T$ as the prediction target. Thus, we define the other response variables as $\boldsymbol{Y}_i = (\gamma_i^*(t_1), \ldots, \gamma_i^*(t_T))$ for $i = 1, \ldots, n_2$. This results in exchangeable $(X_i, \boldsymbol{Y}_i)$ (see Lemma 3 in Appendix A). We again employ the neighborhood smoothing estimator in (6), which yields $\widehat{\boldsymbol{Y}_i} \in \mathbb{R}^T$, $i = 1, \ldots, n_2 + 1$. Note that the discretization grid $\mathcal{T}$ for prediction of relative phase using SFCPP can be different from the discretization grid used for prediction of amplitude using SFCP. Due to computational considerations, we use a much coarser (equally spaced) time grid for relative phase prediction.

Since $\boldsymbol{Y}_i \in \mathbb{R}^T$ is a discretized version of γ_i^*, we use the FR distance on Γ [Srivastava and Klassen, 2016] as the nonconformity score. To efficiently compute the FR distance, we again employ the SRSF representation: for a $\gamma \in \Gamma$, its SRSF is $q^\gamma = \sqrt{\dot{\gamma}}$. Under the SRSF representation, the space of warping functions is the positive orthant of the Hilbert sphere and the FR metric simplifies to the $\mathbb{L}^2$ metric. Thus, the distance between two warping functions γ_i and γ_j is $d_w(\gamma_i, \gamma_j) := \cos^{-1}(\int_0^1 q_i^\gamma(t) q_j^\gamma(t) dt)$, resulting in the nonconformity score $S_i = d_w(\boldsymbol{Y}_i, \widehat{\boldsymbol{Y}_i})$ for $i = 1, \ldots, n_2 + 1$. The prediction set is given by $\mathcal{I}_\gamma := \{\boldsymbol{y} : S_{n_2+1} \leq \mathcal{Q}_{1-\alpha/(T-2)}(\{S_1, \ldots, S_{n_2+1}\})\}$, where $\boldsymbol{y} = (y_1, \ldots, y_T) \in \mathbb{R}^T$ are
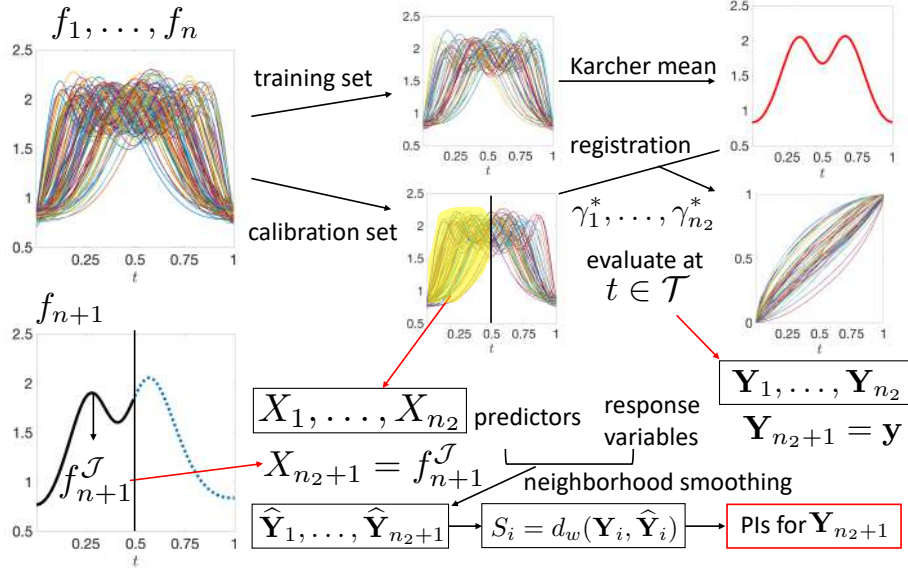
Figure 4: Split Functional Conformal Prediction for relative Phase, (SFCPP) algorithm.

trial vectors for $\boldsymbol{Y}_{n_2+1}$. Here, S_{n_2+1} is compared to the $1 - \alpha/(T-2)$ quantile of the empirical distribution of the nonconformity scores, because $\gamma(0) = 0$ and $\gamma(1) = 1$ for all $\gamma \in \Gamma$. Due to the required constraints on γ (fixed end points and monotonicity), the T-dimensional search for $\boldsymbol{y}$ reduces to a $T-2$-dimensional search for $\boldsymbol{y} = (0, y_2, \ldots, y_{T-1}, 1)$ with $0 < y_2 < \cdots < y_{T-1} < 1$. Figure 4 illustrates the algorithm, which we refer to as Split Functional Conformal Prediction for relative Phase (SFCPP).

4 Simulations

We simulate i.i.d. $f_1, \ldots, f_{n+1}$ without phase variation from (i) a homogeneous population of two-peak functions: $f_i(t) = \boldsymbol{Z}_{i1} \exp\{-(t - 0.25)^2/0.072\} + \boldsymbol{Z}_{i2} \exp\{-(t - 0.25)^2/0.072\}$, $\boldsymbol{Z}_i := (\boldsymbol{Z}_{i1}, \boldsymbol{Z}_{i2}) \overset{iid}{\sim} N(2, 0.1\mathbf{I_2})$; (ii) a heterogeneous population of one- and two-peak functions: one-peak functions are $f_i(t) = Z_i \exp\{-(t - 0.5)^2/0.25\}$, $Z_i \overset{iid}{\sim} N(2, 0.1)$, while two-peak functions are as in (i). We induce phase variation in $\{f_i\}$ by simulating $\gamma_i = F_{a,b}$, where $F_{a,b}$ is the cumulative distribution function of a Beta(a, b) with $a, b \overset{iid}{\sim} \text{Unif}(1, 3)$, and computing $f_i \circ \gamma_i$. To evaluate coverage validity, prediction accuracy and computational
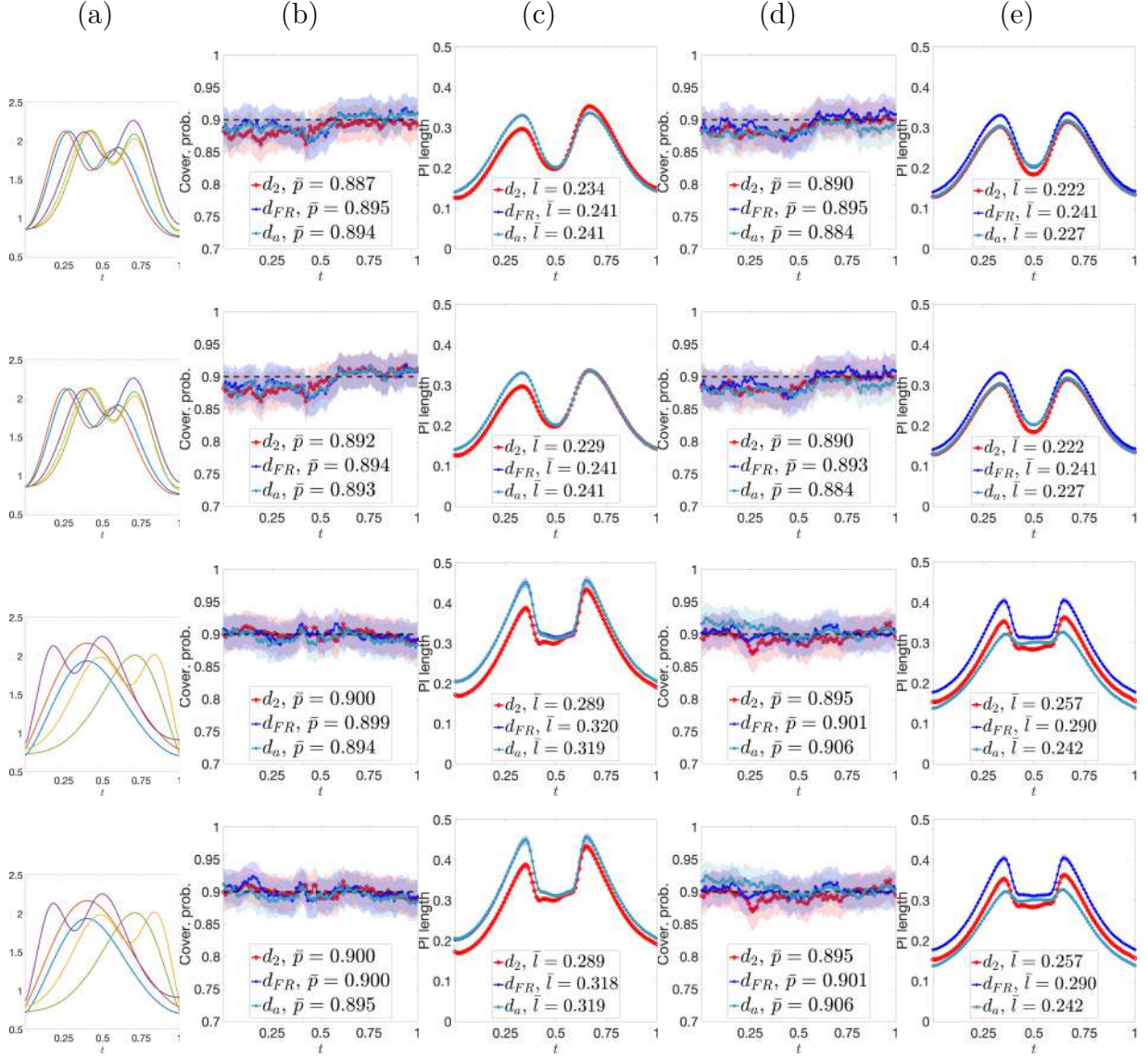
17

Figure 5: Row 1: homogeneous population, global tuning. Row 2: same as 1, but with local tuning. Row 3: heterogeneous population, global tuning. Row 4: same as 3, but with local tuning. (a) Random subsample of 5 functions. (b) Coverage rates with 95% CIs (shaded regions), $U = 0.25$. (c) Average PI lengths, $U = 0.25$. (d) Same as (b), but $U = 0.75$. (e) Same as (c), but $U = 0.75$. In (b)-(e), $d_2(\cdot, \cdot)$ is in red, $d_{FR}(\cdot, \cdot)$ in dark blue and $d_a(\cdot, \cdot)$ in light blue.

efficiency, we use $B = 500$ Monte Carlo samples with $n = 100$, size of time grid $T = 100$ and $1 - \alpha = 0.9$ in all simulations.

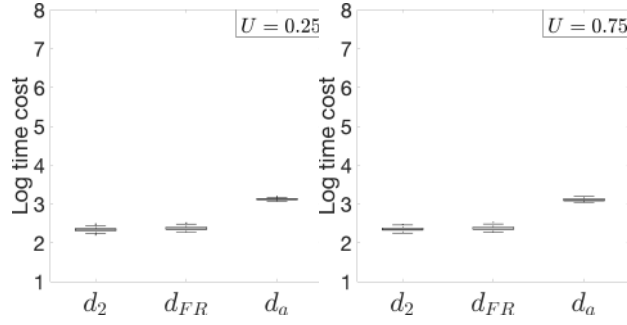**Simulation 1: choice of distance metric and bandwidth parameter for SFCP.** We

18

Figure 6: Log time cost in seconds for a single prediction for (a) $U = 0.25$ and (b) $U = 0.75$.

compare three distances for neighborhood smoothing, $d_2(\cdot, \cdot)$, $d_{\mathrm{FR}}(\cdot, \cdot)$ and $d_a(\cdot, \cdot)$, and two bandwidth tuning methods, global and local. We use $\mathcal{H} = \{\mathcal{Q}_\beta(\{D_{i,j}\}_{1 \le i < j \le n+1}), \ \beta = 0.1, \ldots, 0.9\}$. Figure 5 shows the results. Rows 1&2 (3&4) use functions from a homogeneous (heterogeneous) population with phase variation. Rows 1&3 (2&4) use global (local) tuning for h. Panel (a) shows a random subsample of five functions for each case. Panels (b)-(e) compare pointwise coverage rates (with 95% CIs as shaded regions) and average PI lengths for truncation time points $U = 0.25$ ((b)&(c)) and $U = 0.75$ ((d)&(e)) in f_{n+1}, respectively. Our method provides valid coverage in all cases. When $U = 0.25$, using $d_2(\cdot, \cdot)$ yields the smallest time-averaged PI length $\bar{\ell}$. On the other hand, when functions are sampled from a heterogeneous population and $U = 0.75$ (rows 3&4(e)), $d_a(\cdot, \cdot)$ yields PIs with smallest $\bar{\ell}$. This is because, when a larger portion of the predictors is observed, $d_a(\cdot, \cdot)$ is effective at registering their geometric features, resulting in better predictive power.

To assess computational efficiency, Figure 6 shows log time cost in seconds for a single prediction using the three distances. As expected, $d_2(\cdot, \cdot)$ and $d_{FR}(\cdot, \cdot)$ are faster than $d_a(\cdot, \cdot)$. Summarily, using $d_2(\cdot, \cdot)$ is most computationally efficient and results in best prediction accuracy in most cases. Further, local and global bandwidth tuning result in valid coverage rates, comparable PI lengths and computational cost. Thus, we use $d_2(\cdot, \cdot)$ with local tuning in all subsequent numerical experiments.

Simulation 2: SFCP after smoothing. In some real-world data scenarios, it is desirable to mitigate the effects of noise on prediction by applying smoothing. Thus, to evaluate the effect of functional data smoothing on coverage validity and prediction accuracy of SFCP, we consider two methods: (i) projection onto ten Fourier basis functions, and (ii) moving average smoothing with window size of 12. For this simulation, we added i.i.d. pointwise Gaussian noise to each function from the homogeneous population (with and without phase variation), i.e., $\epsilon_i(t) \overset{i.i.d.}{\sim} N(0, 0.01)$, $i = 1, \ldots, n + 1$, $t = 1, \ldots, T$. After adding noise, we generate the partial function using a truncation time point $U = 0.5$. To maintain exhangeability, we apply smoothing in the following manner. First, we construct the predictors $\{X_i\}$ as before. Then, we separately smooth the predictors and the complete functions resulting in $\{X_i^F\}$ and $\{f_i^F\}$ for Forier basis projection, or $\{X_i^M\}$ and $\{f_i^M\}$ for the moving average smoother; these then serve as inputs into SFCP. Thus, the prediction target becomes $f_{n+1}^F \circ \gamma_{n_2+1}^*$ or $f_{n+1}^M \circ \gamma_{n_2+1}^*$, respectively.

Figure 7(a) shows a randomly sampled noisy functional observation (blue) and its smoothed versions (Fourier basis projection in red and moving average smoother in yellow). Panels (b)&(c) show pointwise coverage rates (with 95% CIs as shaded regions) and average PI lengths, respectively. Panels (d)&(e) show examples of PIs for $f_{n+1}^F \circ \gamma_{n_2+1}^*$ and $f_{n+1}^M \circ \gamma_{n_2+1}$, respectively, with a point prediction in blue (midpoint of PIs at each time point). SFCP yields PIs with valid coverage in all cases. The PIs constructed after Fourier basis projection are shorter close to $t = 0$ and $t = 1$ than those generated after applying the moving average smoother. This is not surprising due to the boundary effect associated with the latter approach. Overall, the time averaged PI length $\bar{l}$ is a bit smaller for Fourier basis projection. Finally, both methods yield prediction bands that are effective at capturing the geometric features, i.e., two peaks and one valley, of the underlying noiseless function.
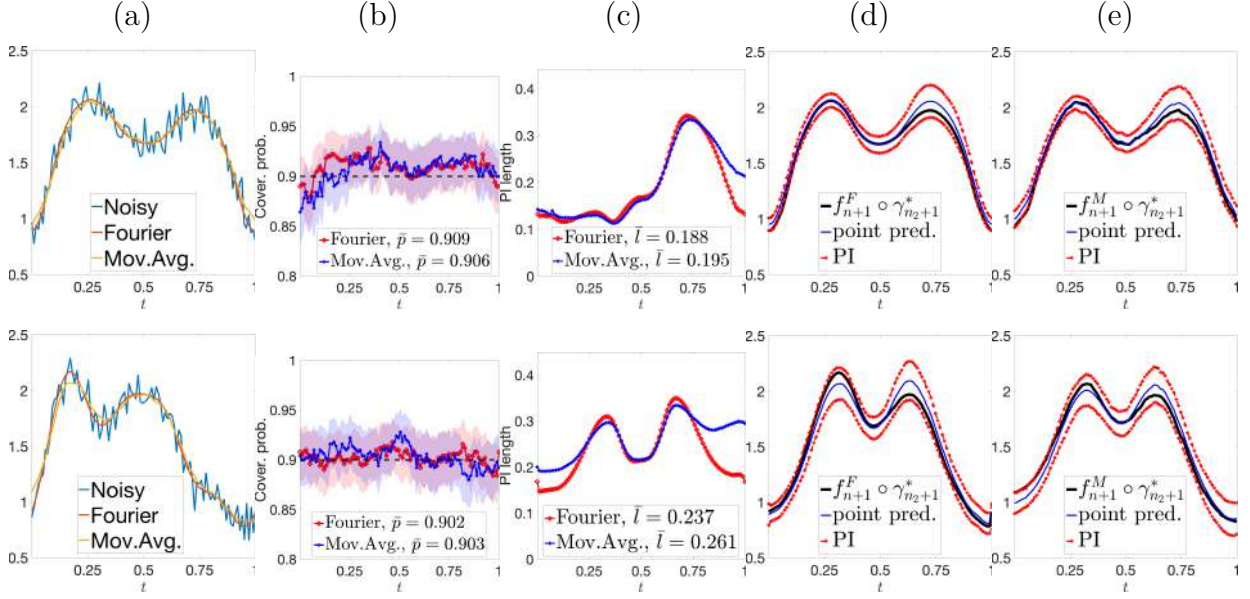
Figure 7: SFCP for smoothed functions, $U = 0.5$. Row 1: without phase variation. Row 2: with phase variation. (a) Noisy observation (blue), smoothed function after applying Fourier basis projection (red) and moving average smoother (yellow). (b)&(c) Coverage rates with 95% CIs (shaded regions) and average PI lengths (Fourier basis projection in red and moving average smoother in blue), respectively. (d)&(e) Target function (black), point prediction (blue) and pointwise PIs (red) after applying Fourier basis projection and moving average smoother, respectively.

Simulation 3: comparison to other prediction methods. Next, we compare SFCP and FFCP to two state-of-the-art functional regression methods, implemented in the R package refund [Goldsmith et al., 2024]: (i) scalar-on-function regression (SoF), with predictors $f_i^{\mathcal{J}}$ and responses $f_i(t)$ fitted independently for each $t \in \mathcal{T}$, and (ii) function-on-function regression (FoF), with predictors $f_i^{\mathcal{J}}$ and responses f_i. To ensure numerical stability of the optimization procedures for SoF and FoF, we apply functional principal component analysis (FPCA) to the predictors, fixing the dimension to 8 or a number that explains 90% of the total variance, whichever is smaller. In addition to the pointwise coverage rates p_k, we also evaluate the empirical overall coverage rate, $p = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{f_{n+1}^{(b)}(t_k) \in \mathcal{I}_{t_k}^{(b)} \ \forall \ k \in [1:T]\}$,
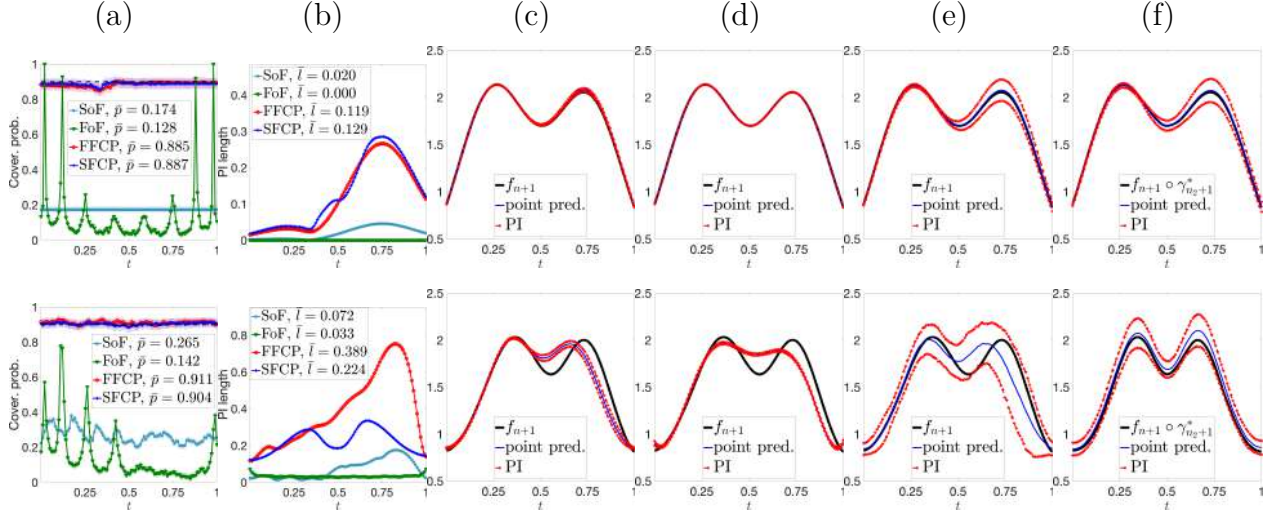
Figure 8: Row 1: without phase variation. Row 2: with phase variation. (a) Pointwise coverage rates with 95% CIs (shaded regions) and (b) average PI lengths for SoF (light blue), FoF (green), FFCP (red) and SFCP (dark blue). (c)-(f) Target function (black), point prediction (blue), pointwise PIs (red) for SoF, FoF, FFCP and SFCP, respectively.

which checks whether the entire target function is within the pointwise prediction band.

Figure 8 shows results for functions generated from a homogeneous population without (row 1) and with (row 2) phase variation, using a truncation time point $U = 0.5$ for f_{n+1}. Panels (a)&(b) show pointwise coverage rates (with 95% CIs as shaded regions) and average PI lengths, respectively. Panels (c)-(f) show examples of PIs (red) and a point prediction (blue) from SoF, FoF, FFCP and SFCP, respectively. Note that the prediction target for the first three methods is f_{n+1}, but for SFCP it is the amplitude of f_{n+1}. SoF and FoF generate PIs with very small lengths, but fail to provide valid coverage. PIs from FFCP and SFCP have valid coverage. However, in the presence of phase variation, SFCP yields PIs with smaller length than FFCP; qualitatively, SFCP results in much better pointwise PIs in this case. Table 1 reports the overall coverage rate p. In absence of phase variation, SoF and FoF have $p = 0$, whereas FFCP and SFCP have $p \approx 0.7$. When phase variation is present

22

in the data, the overall coverage decreases to 0.362 for `FFCP`, but remains stable for `SFCP` at $p = 0.674$. Computationally, `SFCP` is faster than `SoF` and `FFCP`, but slower than `FoF`, when phase variation is not present in the data. When phase variation is present, `SFCP` and `SoF` have comparable speed and are both faster than `FFCP`, but slower than `FoF`. A more detailed description of the computational speed comparison is provided in Appendix B.

Table 1: Overall coverage rates with standard errors in parentheses.

Data	SoF	FoF	FFCP	SFCP
No phase variation	0 (0)	0 (0)	0.6900 (0.0207)	0.6800 (0.0209)
Phase variation	0 (0)	0 (0)	0.3620 (0.0215)	0.6740 (0.0210)

Simulation 4: other observational regimes. We evaluate the performance of `SFCP` for fragmented and sparse observations. For fragmented data, we simulate $\mathcal{J}_1 = [0, 0.2]$, $\mathcal{J}_2 = [0.4, 0.6]$, $\mathcal{J}_3 = [0.8, 1]$, and use $\lambda_j = 1/3$ in the distance $d_{\mathrm{prod}}(\cdot, \cdot)$. For sparse data, we simulate $\mathcal{J} = \{0, 0.1, \ldots, 0.9, 1\}$. Here, we use data from the homogeneous population with phase variation. Row 1 (row 2) in Figure 9 shows results for the fragmented (sparse) case. Panel (a) shows observed (black) and missing (dashed blue) segments of f_{n+1}. Panels (b)&(c) show pointwise coverage rates (with 95% CIs as shaded regions) and average PI lengths, respectively. Panel (d) shows the resulting PIs (red) with a point prediction (blue) for the amplitude of f_{n+1} (ground truth in black). `SFCP` yields PIs that preserve coverage validity in both observational regimes, with very similar PI lengths.

Simulation 5: prediction of relative phase. We apply `SFCPP` to predict the relative phase component $\gamma^*_{n_2+1}$ using data from the homogeneous population with phase variation. Since `SFCPP` predicts $\gamma^*_{n_2+1}$ for all time points simultaneously, we use global tuning for the bandwidth parameter in neighboorhood smoothing. We use a coarse grid of time points
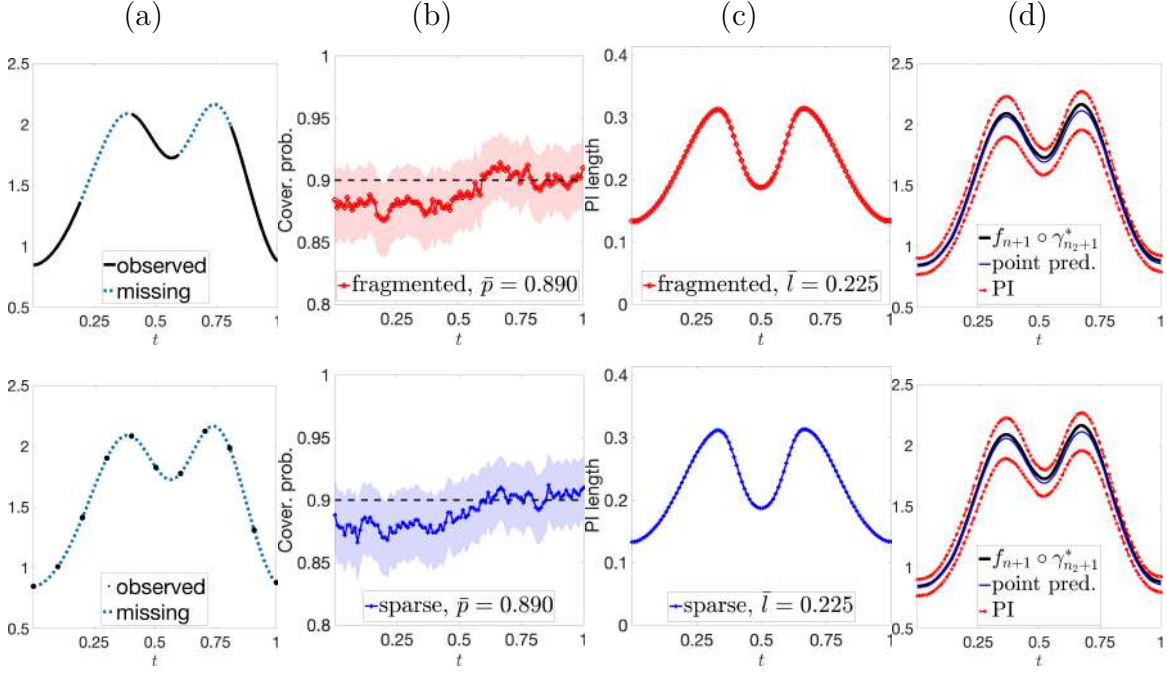
Figure 9: Row 1: fragmented regime. Row 2: sparse regime. (a) Observed (black) and missing (dashed blue) segments for f_{n+1}. (b)&(c) Coverage rates with 95% CIs (shaded regions) and average PI lengths, respectively. (d) Ground truth $f_{n+1} \circ \gamma_{n+1}^*$ (black), point prediction (blue) and pointwise PIs (red).

for prediction, $\mathcal{T} := \{0, 0.25, 0.5, 0.75, 1\}$, and set $\alpha = 0.1$ as before. Figure 10 shows prediction results for a randomly selected Monte Carlo sample. Panels (a)-(c) correspond to truncation time points $U = 0.25$, 0.5 and 0.75 in f_{n+1}, respectively. In all cases, SFCPP provides a reasonable point prediction (blue) and informative PIs (red). The coverage rates and associated 95% CIs for (a)-(c) are 0.976 (0.963, 0.989), 0.980 (0.968, 0.992) and 0.978 (0.965, 0.991), respectively; the proposed method has valid coverage. The time averaged PI lengths (average computed over the coarse time grid $\mathcal{T}$) for (a)-(c) are 0.297, 0.270 and 0.258, respectively. Overall, the PIs become narrower as we observe more of f_{n+1} matching intuition.
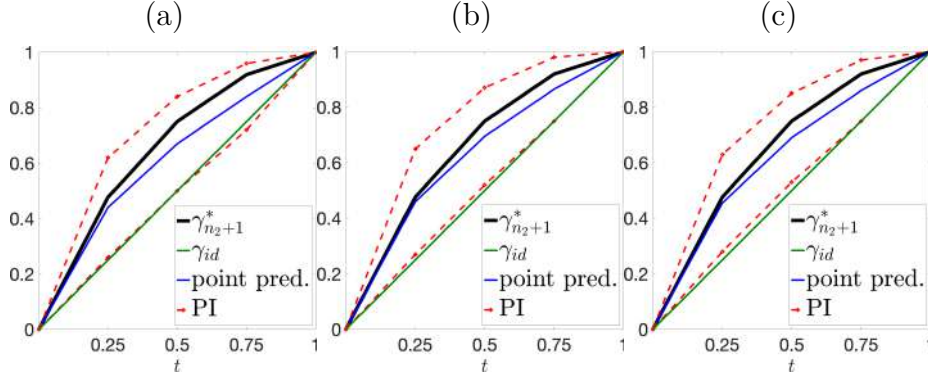
Figure 10: Ground truth $\gamma^*_{n_2+1}$ (black), point prediction (blue), PIs (red) and identity warping (green). (a)-(c) Truncation time points $U = 0.25,\ 0.5$ and 0.75, respectively.

5 Data examples

Example 1: real-world data with simulated truncation time point. We compare prediction performance of `SoF`, `FoF`, `FFCP` and `SFCP` on four functional datasets with a simulated truncation point. Rows 1-4 in Figure 11(a) show the following four datasets, respectively.

1. **Berkeley growth rate functions.** First derivative of measurements on heights in centimeters for $n = 93$ boys and girls from age 1-18 [Ramsay and Silverman, 2005].

2. **PQRST complexes.** Segmented PQRST complexes from a long electrocardiogram (ECG) signal; $n = 80$ [Kurtek et al., 2013].

3. **Gait pressure functions.** Gait dynamics during walking; $n = 50$ individuals [Kurtek et al., 2013].

4. **Traffic flow rate functions.** A smoothed version of $n = 84$ days of traffic flow rate data on National Highway 5 in Taiwan [Chiou, 2012, Jiao et al., 2023].

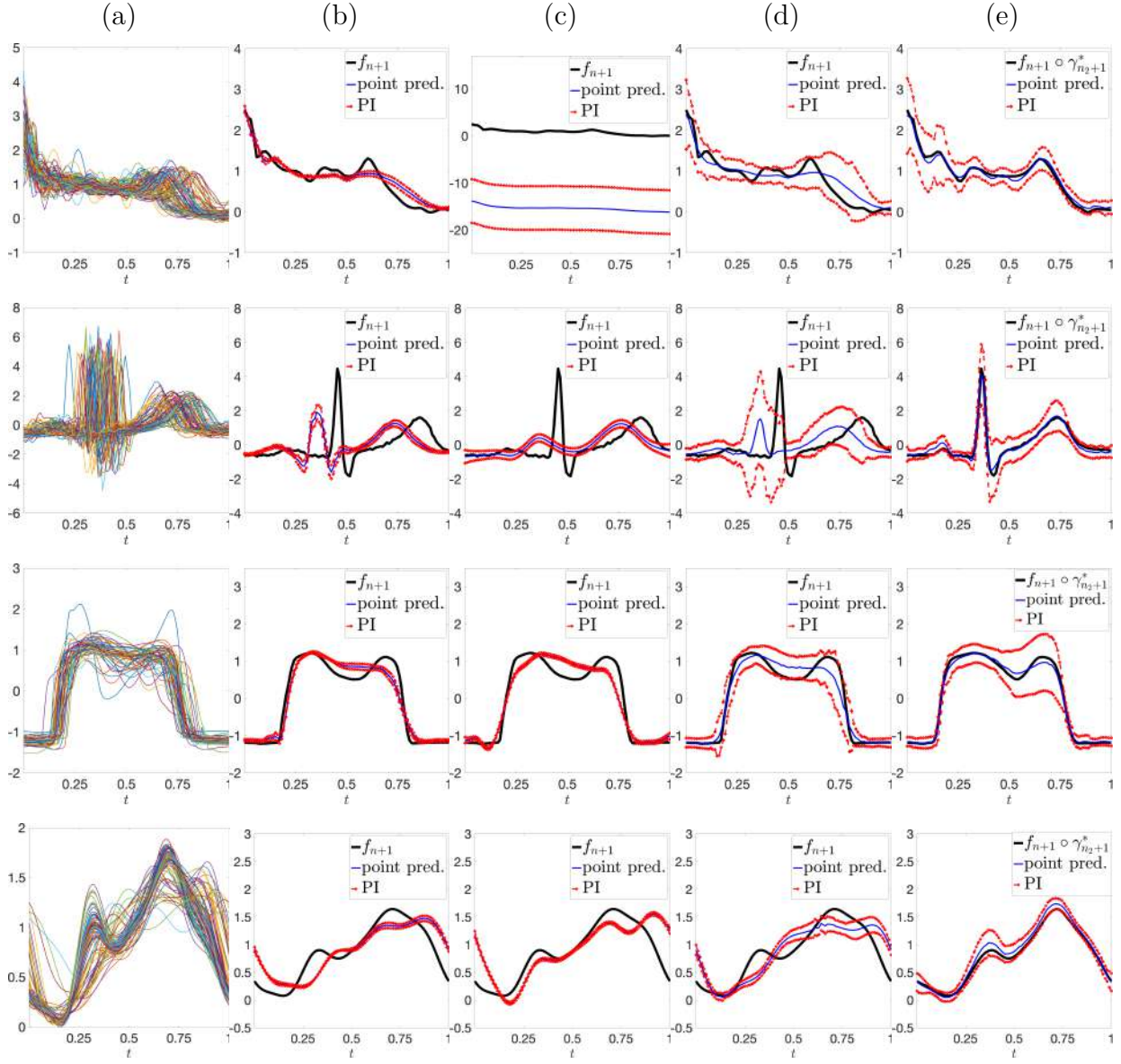In each case, we rescaled the time axis to $[0, 1]$ and normalized all functions to have unit

Figure 11: Rows 1-4: Berkeley growth rate functions, PQRST complexes, gait pressure functions and traffic flow rate functions. (a) Data. (b)-(e) Prediction results, with target function (black), point prediction (blue) and pointwise PIs (red), for SoF, FoF, FFCP and SFCP, respectively.

$\mathbb{L}^2$ norm. For each dataset, f_{n+1} is chosen at random and truncated using $U = 0.5$. Panels (b)-(e) show PIs (red) and point predictions (blue) from SoF, FoF, FFCP and SFCP, respectively; the target function is in black. Compared to the other approaches, SFCP yields much more accurate point predictions and PIs, which capture the main geometric features
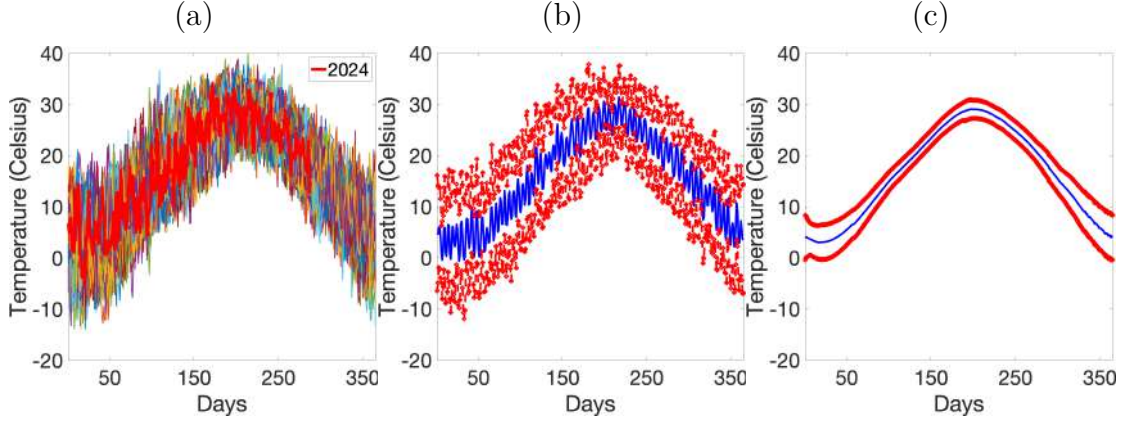
26

Figure 12: Prediction of MDT in Rhode Island. (a) Complete historical data with partial observation for 2024 in red. (b)&(c) PIs (red) and point prediction (blue) for raw and presmoothed data using Fourier basis projection, respectively.

of the target functions. In most cases, `SoF` and `FoF` generate PIs that are overly smooth and fail to capture the true target function. `FFCP` generates much wider PIs to maintain its coverage guarantee, but the resulting point predictions are not accurate.

Example 2: prediction of maximum daily temperature. Finally, we focus on forecasting maximum daily temperatures (MDTs) in Rhode Island. We use complete observations of MDT from 1950 to 2023 to predict MDT for (the remainder of) 2024, based on a partial observation from January 1 to October 15 [NCEI, 2024]. The data is shown in Figure 12(a) with the partial observation highlighted in red. We apply `SFCP` to both, raw data and data after smoothing using Fourier basis projection. Panels (b) and (c) in Figure 12 show the PIs (red) and point predictions (blue) generated by `SFCP` based on raw and smoothed data, respectively. Overall, the proposed approach performs very well. Based on raw data, it generates PIs that capture the overall MDT trend as well as (local) daily fluctuations. Based on smoothed data, the resulting prediction band captures the global MDT trend with reasonable pointwise PI lengths. The 2024 MDT trend in Rhode Island has one maximum in mid-July and one minimum in late January.

6 Discussion

We introduced a novel conformal prediction framework for partial functional data that incorporates registration. Results based on simulations and real-world data examples validate the proposed method's finite-sample coverage, high prediction accuracy and computational efficiency. Despite these advantages over competing approaches, several challenges remain that we leave as future work. First, SFCP may be less effective for highly heterogeneous populations with significant amplitude variation across subpopulations. In such cases, inaccurate estimation of the Karcher mean, a key step in our framework, may impair prediction accuracy. Addressing this issue requires strategies that account for such heterogeneity, e.g., a group conditional approach given observed labels for subpopulations. Second, SFCP relies on a grid search over trial values when constructing prediction intervals, which can be computationally expensive for a very dense grid of time points. This issue is further amplified in SFCPP, where joint prediction over multiple time points is required. Exploring alternative conformal prediction methods that avoid an exhaustive grid search can significantly improve computational efficiency and broaden the practical utility of the framework. Finally, SFCP generates pointwise prediction intervals with marginal coverage validity only. We will explore alternative formulations that utilize basis expansions with a global coverage validity guarantee.

Acknowledgments

References

Niccolò Ajroldi, Jacopo Diquigiovanni, Matteo Fontana, and Simone Vantini. Conformal prediction bands for two-dimensional functional time series. *Computational Statistics & Data Analysis*, 187:107821, 2023.

Anastasios Angelopoulos, Emmanuel Candes, and Ryan J Tibshirani. Conformal PID control for time series prediction. *Advances in Neural Information Processing Systems*, 36, 2023.

Laura K Bachrach, Trevor Hastie, May-Choo Wang, Balasubramanian Narasimhan, and Robert Marcus. Bone mineral acquisition in healthy asian, hispanic, black, and caucasian youth: a longitudinal study. *The Journal of Clinical Endocrinology & Metabolism*, 84 (12):4702–4712, 1999.

Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816–845, 2023.

Alexander Bauer, Fabian Scheipl, Helmut Küchenhoff, and Alice-Agnes Gabriel. Registration for incomplete non-gaussian functional data. *arXiv preprint arXiv:2108.05634*, 2021.

Darshan Bryner and Anuj Srivastava. Shape analysis of functional data with elastic partial matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 44(12):9589–9602, 2021.

Emmanuel Candès, Lihua Lei, and Zhimei Ren. Conformalized survival analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(1):24–45, 2023.

Maxime Cauchois, Suyash Gupta, and John C Duchi. Knowing what you know: valid and validated confidence sets in multiclass and multilabel prediction. *Journal of Machine Learning Research*, 22(81):1–42, 2021.

Jeng-Min Chiou. Dynamical functional prediction and classification, with application to traffic flow prediction. *Annals of Applied Statistics*, 6(4):1588–1614, 2012.

Anna De Magistris, Andrea Diana, and Elvira Romano. Conformal prediction for functional ordinary kriging. *arXiv preprint arXiv:2409.20084*, 2024.

Andrea Diana, Elvira Romano, and Antonio Irpino. Distribution free prediction for geographically weighted functional regression models. *Spatial Statistics*, 57:100765, 2023.

Devin Didericksen, Piotr Kokoszka, and Xi Zhang. Empirical properties of forecasts with the functional autoregressive model. *Computational Statistics*, 27(2):285–298, 2012.

Jacopo Diquigiovanni, Matteo Fontana, and Simone Vantini. Conformal prediction bands for multivariate functional data. *Journal of Multivariate Analysis*, 189:104879, 2022.

Cecilia Earls and Giles Hooker. Variational Bayes for functional data registration, smoothing, and prediction. *Bayesian Analysis*, 12(2):557 – 582, 2017.

Frédéric Ferraty, André Mas, and Philippe Vieu. Nonparametric regression on functional data: inference and practical aspects. *Australian & New Zealand Journal of Statistics*, 49(3):267–286, 2007.

Jeff Goldsmith, Fabian Scheipl, Lei Huang, et al. *refund: Regression with Functional Data*, 2024. URL https://CRAN.R-project.org/package=refund. v 0.1-37.

Yu Gui, Rina Barber, and Cong Ma. Conformalized matrix completion. *Advances in Neural Information Processing Systems*, 36:4820–4844, 2023.

Yu Gui, Rohan Hore, Zhimei Ren, and Rina Foygel Barber. Conformalized survival analysis with adaptive cut-offs. *Biometrika*, 111(2):459–477, 2024.

Shuhao Jiao, Alexander Aue, and Hernando Ombao. Functional time series prediction under partial observation of the future curve. *Journal of the American Statistical Association*, 118(541):315–326, 2023.

Alois Kneip and Theo Gasser. Statistical tools to analyze data representing a sample of curves. *The Annals of Statistics*, pages 1266–1305, 1992.

Sebastian Kurtek, Wei Wu, Gary E. Christensen, and Anuj Srivastava. Segmentation, alignment and statistical analysis of biosignals with application to disease classification. *Journal of Applied Statistics*, 40:1270–1288, 2013.

Jing Lei, Alessandro Rinaldo, and Larry Wasserman. A conformal prediction approach to explore functional data. *Annals of Mathematics and Artificial Intelligence*, 74:29–43, 2015.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

Andrew Leroux, Luo Xiao, Ciprian Crainiceanu, and William Checkley. Dynamic prediction in functional concurrent regression with an application to child growth. *Statistics in Medicine*, 37(8):1376–1388, 2018.

Ruiting Liang, Wanrong Zhu, and Rina Foygel Barber. Conformal prediction after efficiency-oriented model selection. *arXiv preprint arXiv:2408.07066*, 2024.

Arnab Maity. Nonparametric functional concurrent regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 9(2):e1394, 2017.

James Stephen Marron, James O Ramsay, Laura M Sangalli, and Anuj Srivastava. Functional data analysis of amplitude and phase variation. *Statistical Science*, pages 468–484, 2015.

James Matuk, Karthik Bharath, Oksana Chkrebtii, and Sebastian Kurtek. Bayesian framework for simultaneous registration and estimation of noisy, sparse, and fragmented functional data. *Journal of the American Statistical Association*, 117(540):1964–1980, 2022.

Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

NCEI. National Centers for Environmental Information, 2024. https://www.ncei.noaa.gov/cdo-web/.

James O. Ramsay and Bernard W. Silverman. *Functional Data Analysis*. Springer, New York, 2005.

Daniel T. Robinson. *Functional Data Analysis and Partial Shape Matching in the Square Root Velocity Framework*. PhD thesis, Florida State University, 2012.

Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

Craig Saunders, Alex Gammerman, and Volodya Vovk. Transduction with confidence and credibility. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, volume 2, pages 722—726, Morgan Kaufmann, Los Altos, CA, 1999.

Meijia Shao and Yuan Zhang. Distribution-free matrix prediction under arbitrary missing pattern. *arXiv preprint arXiv:2305.11640*, 2023.

Anuj Srivastava and Eric P Klassen. *Functional and shape data analysis*, volume 1. Springer, 2016.

Anuj Srivastava, Wei Wu, Sebastian Kurtek, Eric Klassen, and James Stephen Marron. Registration of functional data using fisher-rao metric. *arXiv preprint arXiv:1103.3817*, 2011.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World.* Springer, 2005.

Geoffrey S Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372, 1964.

Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

Xueyan Yin, Genze Wu, Jinze Wei, et al. Deep learning on traffic prediction: Methods, analysis, and future directions. *IEEE Trans. on Intelligent Transportation Systems*, 23 (6):4927–4943, 2021.

Margaux Zaffran, Olivier Féron, Yannig Goude, Julie Josse, and Aymeric Dieuleveut. Adaptive conformal predictions for time series. In *International Conference on Machine Learning*, pages 25834–25866. PMLR, 2022.

Yuan Zhang, Elizaveta Levina, and Ji Zhu. Estimating network edge probabilities by neighbourhood smoothing. *Biometrika*, 104(4):771–783, 2017.

Appendices

A   Lemmas and proofs

Exchangeability lemmas. We state and prove three exchangeability lemmas that are used to establish the coverage guarantee of PIs generated using FFCP, SFCP and SFCPP.

Lemma 1 (Exchangeability, FFCP). *Suppose Assumption 1 holds. For any $t \in \mathcal{T}$, let $X_i = f_i^{\mathcal{J}}$ and $Y_i(t) = f_i(t)$ for $i = 1, \ldots, n+1$. Then $\{(X_i, Y_i(t))\}$, $i = 1, \ldots, n+1$ are exchangeable.*

Lemma 2 (Exchangeability, SFCP). *Suppose Assumption 1 holds. For any $t \in \mathcal{T}$, let $X_i = f_{n_1+i}^{\mathcal{J}}$ and $Y_i(t) = \tilde{f}_i(t) = (f_{n_1+i} \circ \gamma_i^*)(t)$ for $i = 1, \ldots, n_2 + 1$. Then $\{(X_i, Y_i(t))\}$, $i = 1, \ldots, n_2 + 1$ are exchangeable.*

Lemma 3 (Exchangeability, SFCPP). *Suppose Assumption 1 holds. Let $X_i = f_{n_1+i}^{\mathcal{J}}$ and $\boldsymbol{Y}_i = (\gamma_i^*(t_1), \ldots, \gamma_i^*(t_T))$ for $i = 1, \ldots, n_2 + 1$. Then $\{(X_i, \boldsymbol{Y}_i)\}$, $i = 1, \ldots, n_2 + 1$ are exchangeable.*

Proof. We first prove Lemma 2. Let $\bar{f}$ be the sample Karcher mean of functions $f_1, \ldots, f_{n_1}$ in the training set. For any $t \in \mathcal{T}$, define mapping $g : \mathcal{F} \times \mathcal{F} \times \mathbb{R} \to \mathcal{X} \times \mathcal{Y}$, $g(f_{n_1+i}, \bar{f}, U) = (X_i, Y_i(t))$, $i = 1, \ldots, n_2 + 1$. Given U, $\bar{f}$ and f_{n_1+i}, g is a deterministic procedure, which involves (i) truncating f_{n_1+i} at $t = U$ to generate X_i, (ii) registering f_{n_1+i} to $\bar{f}$ via (4), and (iii) evaluating the registered f_{n_1+i} at $t \in \mathcal{T}$; steps (ii) and (iii) result in $Y_i(t)$, $t \in \mathcal{T}$. Under Assumption 1, g is symmetric with respect to f_{n_1+i}, $i = 1, \ldots, n_2 + 1$. Further, again using Assumption 1, for any $t \in \mathcal{T}$, $(X_i, Y_i(t))|(\bar{f}, U)$, $i = 1, \ldots, n_2 + 1$ are i.i.d., which implies $(X_i, Y_i(t))$, $i = 1, \ldots, n_2 + 1$ are exchangeable.

Proofs of Lemmas 1 and 3 follow similar arguments as the proof of Lemma 2. For Lemma

1, FFCP only involves truncating $\{f_i\}$ at $t = U$ to generate X_i and evaluating $\{f_i\}$ at $t \in \mathcal{T}$ to generate $Y_i(t)$ for $i = 1, \ldots, n+1$. Thus, under Assumption 1, $(X_i, Y_i(t))|U, \ i = 1, \ldots, n+1$ are i.i.d.. For Lemma 3, and again using Assumption 1, since registering $f_{n_1+i}, \ i = 1, \ldots, n_2+1$ to $\bar{f}$ is deterministic and symmetric, the resulting warping functions $\gamma_i^*|\bar{f}, \ i = 1, \ldots, n_2 + 1$ are i.i.d., and thus $\boldsymbol{Y}_i|\bar{f}$ and $(X_i, \boldsymbol{Y}_i)|(\bar{f}, U), \ i = 1, \ldots, n_2 + 1$ are also i.i.d. $\qquad\square$

Proof of Theorem 1. Lemma 2 ensures exchangeability of $\{(X_i, Y_i)\}, \ i = 1, \ldots, n_2 + 1$. Given exchangeable predictors and responses, a permutation symmetric algorithm, and absolute residual as nonconformity score, the coverage guarantee in Theorem 1 follows directly from standard conformal prediction literature [Vovk et al., 2005, Barber et al., 2023]. The anti-conservative bound follows from the proof in Lei et al. [2018], assuming no ties in $\{S_i\}$. $\qquad\square$

B   Comparison of computational cost

To compare computational efficiency of `SoF`, `FoF`, `SFCP` and `FFCP`, we recorded log time cost for a single prediction of a target function based on $n = 100$ complete functions sampled at $T = 100$ time points. For `SFCP`, we used an equal number of training and calibration samples, i.e., $n_1 = n_2 = 50$. Figure 13 shows the results for data without (a) and with (b) phase variation. In general, `FoF` is faster than other methods. This is because `FoF` performs prediction for f_{n+1} simultaneously at all time points. In contrast, the other methods predict the target function in a pointwise manner. Note that computational efficiency of the pointwise approaches can be improved via a parallel implementation across time points. Also, `SFCP` is faster than `FFCP` in both cases. Unlike standard split conformal methods, `SFCP` does not avoid the grid search. However, the computational cost depends on
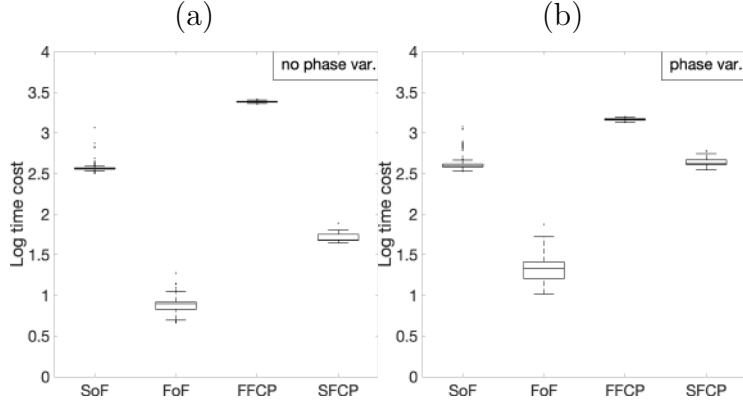
Figure 13: (a)&(b) Log time cost in seconds for a single prediction for functions without and with phase variation, respectively.

the number of predictor-response pairs. SFCP only uses $n_2 = 50$ samples in the calibration set to generate $\{X_i, Y_i(t)\}$ for prediction, which is much smaller than the $n = 100$ samples used in FFCP. SFCP is slower when phase variation is present in the data than when it is not. The Karcher mean is estimated via an iterative algorithm, which requires more iterations to converge in the presence of phase variation.