



Sequential Bayesian Registration for Functional Data

Yoonji Kim¹ · Oksana A. Chkrebtii¹ · Sebastian A. Kurtek¹

Received: 7 June 2024 / Accepted: 10 May 2025 / Published online: 27 May 2025
© The Author(s) 2025

Abstract

In many modern applications, discretely-observed data may be naturally understood as a set of functions. Functional data often exhibit two confounded sources of variability: amplitude (y -axis) and phase (x -axis). The extraction of amplitude and phase, a process known as registration, is essential in exploring the underlying structure of functional data in a variety of areas, from environmental monitoring to medical imaging. Critically, such data are often gathered sequentially with new functional observations arriving over time. Despite this, existing registration procedures do not sequentially update inference based on the new data, requiring model refitting. To address these challenges, we introduce a Bayesian framework for sequential registration of functional data, which updates statistical inference as new sets of functions are assimilated. This Bayesian model-based sequential learning approach utilizes sequential Monte Carlo sampling to recursively update the alignment of observed functions while accounting for associated uncertainty. Distributed computing significantly reduces computational cost relative to refitting the model using an iterative method such as Markov chain Monte Carlo on the full data. Simulation studies and comparisons reveal that the proposed approach performs well even when the target posterior distribution has a challenging structure. We apply the proposed method to three real datasets: (1) functions of annual drought intensity near Kaweah River in California, (2) annual sea surface salinity functions near Null Island, and (3) a sequence of repeated patterns in electrocardiogram signals.

Keywords Bayesian updating · Function registration · Sequential Monte Carlo · Square-root velocity function

1 Introduction

In various real world problems, the goal of statistical analysis is to discover and explore patterns in the trajectories formed by the temporal evolution of a variable of interest. This type of data is commonly referred to as functional data, and is often recorded on a very fine temporal grid. The use of multivariate statistical methods to analyze functional data is inappropriate for two main reasons: (1) failure to account for the underlying infinite-dimensional structure of the data, and (2) inability to appropriately model strong temporal dependence within each functional observation (Ferraty and Vieu 2006; Lajos and Kokoszka 2012). This has given rise to the

field of functional data analysis (FDA), which provides a comprehensive framework for statistical modeling, summarization, analysis and visualization of data that comes in the form of functions (Ramsay and Dalzell 1991; Ramsay and Silverman 2005).

An important and common feature of functional data is that sampling is often automated or conducted over long periods of time, so that new observed functions arrive sequentially. In general, there are two different approaches to perform statistical analysis for such an expanding collection of data in the finite or infinite-dimensional settings. The first is to implement the full spectrum of statistical analysis every time a new observation arrives, referred to as *batch learning*. The second is to update the existing analysis by accounting for the new data, referred to as *sequential learning* or online learning (Hoi et al. 2021). Most existing FDA techniques are designed for batch learning, meaning that they are performed once a given number of functional data is collected, and the analysis must be repeated on the entire sample as more data arrives. This fails to account for the sequential way in which functional data is often gathered, with

✉ Oksana A. Chkrebtii
oksana@stat.osu.edu

Yoonji Kim
yoonj2kim@gmail.com

Sebastian A. Kurtek
kurtek.1@stat.osu.edu

¹ Department of Statistics, The Ohio State University, 1958
Neil Avenue, Columbus, Ohio 43210, USA

the sample size increasing over time in many application domains such as environmental monitoring or biomedical imaging. For example, trajectories of annual temperature, or other measures related to the environment, are formulated through sets of repeated measurements on an annual basis; in medicine, biosignals such as electrocardiogram (ECG) signals or gait measurements contain repetitions of a particular pattern wherein each repetition can be interpreted as an observation. We provide a visualization of sequential learning for trajectories of annual drought intensity near Kaweah River in California in Figure 1 (see Section 5.4 for details). In such scenarios, new functional observations are added to existing data sequentially, so the statistical analysis pipeline must be modified to allow for updating and monitoring of inferential results as the collection of data expands.

1.1 Sequential learning

A sequential learning method seeks to update current inferential results to include new data. The Bayesian approach is well-suited to this problem because it (1) provides a systematic way to assimilate new data by updating the posterior distribution as new data arrives, and (2) allows the user to keep track of structured uncertainty. In most scenarios, the posterior distribution does not have a closed form, so inference is based on estimates of posterior features obtained from posterior samples. Perhaps the most widely used sampling-based method for Bayesian inference is Markov chain Monte Carlo (MCMC), which is a batch learning algorithm; as such, every time new data arrives, MCMC sampling must be repeated using the full data, resulting in inefficient computation. On the other hand, sequential Bayesian learning via, e.g., sequential Monte Carlo (SMC) can assimilate new data as it arrives. Unlike MCMC, which targets a fixed posterior density, SMC defines a sequence of intermediate target densities, each represented with a set of weighted samples, or *particles*, that are perturbed and reweighted to represent the next density in the sequence (Andrieu et al. 1999; Storvik 2002; Lopes and Tsay 2011). When these intermediate target densities correspond to posteriors under different data availability scenarios, SMC becomes a sequential learning algorithm.

In addition to the efficiency of sequentially updating inference as new data arrives, SMC can exploit distributed computing to speed up implementation relative to MCMC, because each SMC particle is updated independently. SMC also performs well for sampling from challenging target posteriors, e.g., in the presence of posterior multimodality, because the intermediate sequence of target densities can act as a bridge between the prior and a challenging posterior distribution (Schweizer 2012; Paulin et al. 2019; Dai et al. 2022). The potential for particle degeneracy, i.e., when particle weights become very small in certain situations and

lead to large sampling variance in the SMC estimator, can be ameliorated through resampling or via techniques such as block sampling (Gilks and Berzuini 2001; Chopin 2002; Del Moral et al. 2006; Kantas et al. 2013; Beskos et al. 2015). While SMC methods for sequential Bayesian learning have been used for assimilating multivariate data (Lopes and Tsay 2011; Liu and West 2001), our focus in this work is on a natural inferential problem arising in FDA.

1.2 Functional data registration

A common challenge in FDA is the presence of two confounded sources of variability: amplitude and phase (Ramsay and Li 1998). Examining Figure 1, we note that the functions contain similar shape features, e.g., number of local extrema, but the timing of the features is not the same along the temporal x -axis across all observations. For example, drought intensity tends to increase sharply early in each hydrological year. Then, around the month of December, drought intensity decreases following a small peak. Thus, the variation in the data can be attributed to two sources: (1) the magnitude of function values (drought intensity) termed amplitude (y -axis) variability, and (2) the timing of amplitude features (e.g., local extrema) termed phase (x -axis) variability. Phase variation may be an inherent feature of data or the result of measurement error, and can be regarded as nuisance or a quantity of interest depending on the application. Importantly, phase variation cannot be ignored when performing statistical analysis of functional data as this may lead to misleading results (Marron et al. 2015). Instead, amplitude and phase components of functions should first be estimated through a process called registration. A common registration approach is to consider the observed functions as deformed versions of an unknown template function, and to extract their phase components via horizontal synchronization to an estimate of this template.

There are many methods for functional data registration (Ramsay and Li 1998; Tang and Müller 2008; Srivastava et al. 2011). Here, we briefly review a small subset. Landmark-based registration focuses on synchronization of (a small number of) points, which represent important features of the observed functions, e.g., local extrema (Ramsay and Silverman 2005; Kneip et al. 2000). While conceptually simple, these approaches rely on a faithful specification of landmarks, which is often difficult and time consuming. Metric-based registration uses a distance on the function space of interest to achieve horizontal synchronization of entire functions, i.e., it does not require landmark specification. However, the distance must satisfy a key invariance property (see Section 2.1 for details), which is not the case for the standard \mathbb{L}^2 distance commonly used in FDA. As an alternative, Srivastava et al. (2011) proposed the extended Fisher-Rao (eFR) metric, which is equivalent to the \mathbb{L}^2 dis-

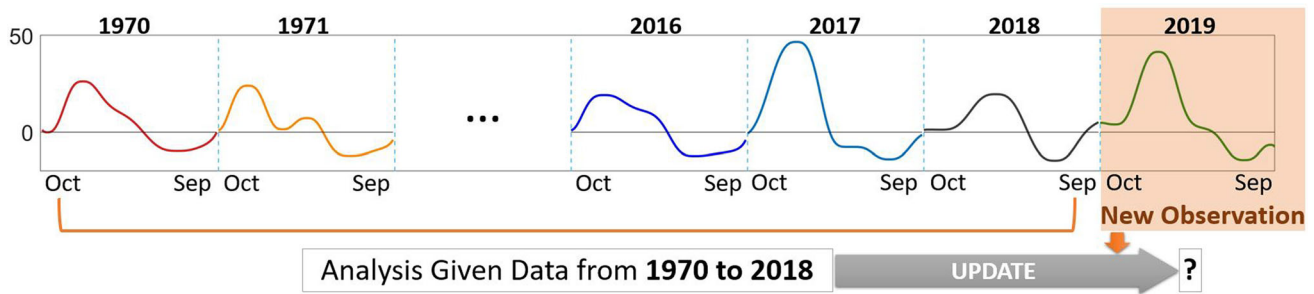


Fig. 1 Visualization of sequential learning for trajectories of annual drought intensity near Kaweah River in California during years 1970 to 2019

tance under a simple transformation of the original functions, called the square-root velocity transformation. The resulting metric-based registration method is commonly referred to as elastic. Bayesian model-based registration of functional data has been explored relatively recently. In this setting, the main challenge lies in specifying an appropriate prior distribution over the phase component of functional data. Telesca and Inoue (2008) were the first to approach registration from this perspective and modeled phase via (constrained) B-splines. Lu et al. (2017) explicitly considered the geometry of the representation space of the phase component and specified a Gaussian process prior on this space. Horton et al. (2021) also modeled phase via a Gaussian process, but additionally allowed the incorporation of landmark information in the registration process. Finally, Cheng et al. (2016) and Bharath and Kurtek (2020) used the Dirichlet distribution as a prior model on consecutive increments of discretized phase functions. An extension of Bayesian registration to sparse/fragmented functional data was developed by Matuk et al. (2021). Importantly, all of the aforementioned model-based approaches rely on batch learning, and in particular MCMC, for inference.

1.3 Summary of proposed approach and paper organization

Motivated by data collection scenarios such as the one presented in Figure 1, we propose a novel sequential Bayesian learning approach for registration of functional data. The proposed approach updates the posterior over all unknown model components, including the template function, all of the phase components, and the observation error variance, when a new function arrives. This framework leverages SMC to efficiently update the joint posterior distribution over the template function and the phase components associated with each observation as new data arrives. To the best of our knowledge, this is the first sequential inference strategy for this statistical problem.

The proposed computational approach is applicable to general registration models for functional data, though the specific model we consider is built on the state-of-the-art

framework of Lu et al. (2017). In this model, a template function captures amplitude variation across observations, and individual phase functions account for the phase of each observed function. For each phase, we employ a low-dimensional piecewise linear prior model (Bharath and Kurtek 2020). We further address the challenge posed by the increasing dimension of the state space corresponding to the addition of an unknown phase for the incoming data. In addition to using a low-dimensional prior on phase, we propose an efficient stochastic initialization strategy for the new components of each SMC particle.

The rest of this paper is organized as follows. Section 2 provides a brief review of elastic registration and introduces the Bayesian registration model. Sections 3 and 4 describe SMC methods for state spaces of fixed and increasing dimension and the proposed SMC Bayesian registration algorithm enabling sequential inference as new functional data is observed, respectively. Section 5 presents simulations and real data examples. We close with a brief discussion in Section 6. Appendix A in the supplement contains the detailed sequential Bayesian registration algorithm (Algorithm 1). Appendices B and C present additional registration results for annual sea surface salinity functions considered in Section 5.5 and segmented PQRST complexes considered (and defined) in Section 5.6, respectively. Appendix D compares effective sample size for MCMC-based batch learning and the proposed sequential approach. Finally, Appendix E considers comprehensive sensitivity analyses. The code to reproduce results in this manuscript can be found at <https://github.com/yoony2kim/Bayesian-sequential-registration>.

2 Functional registration model

We first provide an overview of the elastic registration framework. For brevity, we only present concepts relevant to the Bayesian hierarchical model for elastic registration of functions, which is based on the square-root velocity function representation; we refer the interested reader to Srivastava and Klassen (2016) for further details. The choice of this representation is motivated by the desirable properties of the

extended Fisher-Rao metric as described in Section 2.1. The presented model is built on the general structure of the observation model in Lu et al. (2017), while utilizing different priors including a low-dimensional prior distribution over phase.

2.1 Elastic registration

We restrict our attention to absolutely continuous functions with domain $[0, 1]$, resulting in the representation space $\mathcal{F} = \{f : [0, 1] \rightarrow \mathbb{R} \mid f \text{ is absolutely continuous}\}$. The domain can be further extended to any $[a, b] \subset \mathbb{R}$, $a < b$. The phase component of a function $f \in \mathcal{F}$ is denoted by γ and is an element of $\Gamma = \{\gamma : [0, 1] \rightarrow [0, 1] \mid \gamma(0) = 0, \gamma(1) = 1, 0 < \dot{\gamma} < \infty\}$, where $\dot{\gamma}$ is the derivative of γ . The main goal of registration is to estimate the phase components $\gamma_1, \dots, \gamma_n$ of a set of functions f_1, \dots, f_n , such that $f_i \circ \gamma_i$, $i = 1, \dots, n$ are horizontally synchronized, i.e., their features are well-aligned. The composition of f and γ is usually referred to as domain warping. Metric-based registration utilizes a distance on \mathcal{F} to quantify the quality of alignment between two functions, and estimation of phase is carried out by minimizing this distance over elements of Γ . The chosen distance must satisfy $d(f_1, f_2) = d(f_1 \circ \gamma, f_2 \circ \gamma)$ for $f_1, f_2 \in \mathcal{F}$ and any $\gamma \in \Gamma$, i.e., it must be invariant to simultaneous domain warping. Crucially, the commonly used \mathbb{L}^2 distance does not satisfy this property.

Srivastava et al. (2011) proposed a formulation of metric-based registration using the extended Fisher-Rao (eFR) Riemannian metric. The eFR distance is preserved under simultaneous domain warping, i.e., $d_{eFR}(f_1, f_2) = d_{eFR}(f_1 \circ \gamma, f_2 \circ \gamma)$ for $f_1, f_2 \in \mathcal{F}$ and any $\gamma \in \Gamma$. Since d_{eFR} , and thus the resulting registration problem, are not computationally tractable, Srivastava et al. introduced a transformation that allows the distance to be computed in closed form. The *square-root velocity function (SRVF)* representation given by the mapping $Q : \mathcal{F} \rightarrow \mathcal{Q}$ is defined as $Q(f) = \text{sign}(\dot{f})\sqrt{|\dot{f}|} =: q$ for $f \in \mathcal{F}$; the derivative of f , \dot{f} , is approximated using finite differencing. Given the starting point $f(0)$, the mapping Q is bijective and the original function can be reconstructed using $f(t) = Q^{-1}(f(0), q)(t) = f(0) + \int_0^t q(s)|q(s)|ds$. Further, under this mapping, the eFR metric simplifies to the \mathbb{L}^2 metric, i.e., $d_{eFR}(f_1, f_2) = d_{\mathbb{L}^2}(Q(f_1), Q(f_2))$ for $f_1, f_2 \in \mathcal{F}$, and the resulting space of SRVFs \mathcal{Q} is a subset of $\mathbb{L}^2([0, 1], \mathbb{R})$. The domain warping $f \circ \gamma$ of a function $f \in \mathcal{F}$ by $\gamma \in \Gamma$ can be mapped to the SRVF space via $(q, \gamma) := (q \circ \gamma)\sqrt{\dot{\gamma}}$ ($q = Q(f)$). Finally, the amplitude of a function f can be formally defined through its SRVF q as the equivalence class $[q] = \{(q, \gamma) \mid \gamma \in \Gamma\}$; the set of all amplitudes is the quotient space \mathcal{Q}/Γ .

The SRVF representation is then used to define the metric-based registration problem as follows. For registration of two functions, $f_1, f_2 \in \mathcal{F}$, we set f_1 as the

reference function and find the optimal phase component of f_2 that minimizes the \mathbb{L}^2 distance between their SRVFs $q_1, q_2 \in \mathcal{Q}$, i.e., $\gamma^* = \arg \min_{\gamma \in \Gamma} d_{\mathbb{L}^2}(q_1, (q_2, \gamma))$. When multiple functions f_1, \dots, f_n are given, we register them to (an estimate of) a representative of the mean equivalence class, referred to as a template, rather than an arbitrarily chosen reference function. The SRVF of the template function, denoted by q_μ , is estimated using $[q_\mu] = \arg \min_{[q_\mu] \in \mathcal{Q}/\Gamma} \sum_{i=1}^n \min_{\gamma \in \Gamma} d_{\mathbb{L}^2}(q_\mu, (q_i, \gamma))^2$; the solution is an entire equivalence class. For identifiability, one generally selects $q_\mu \in [q_\mu]$ such that the average of the phase components, estimated via $\gamma_i^* = \arg \min_{\gamma_i \in \Gamma} d_{\mathbb{L}^2}(q_\mu, (q_i, \gamma_i))$, $i = 1, \dots, n$, is the identity $\gamma_{id}(t) = t$. For visualization, the SRVF of the template, q_μ , is mapped to \mathcal{F} using Q^{-1} . Figure 2 presents an illustration of metric-based registration under the elastic framework.

2.2 Observation model

Denote the functional data by $f_1, \dots, f_n \in \mathcal{F}$ and the corresponding SRVFs by $q_1, \dots, q_n \in \mathcal{Q}$. We assume that each datum f_i is a noisy deformation of a latent template function $f_\mu \in \mathcal{F}$ whose SRVF is $q_\mu \in \mathcal{Q}$. The deformation of the i th function is expressed through the unknown phase $\gamma_i \in \Gamma$. The observation error is assumed to be additive in the SRVF space, so that the i th function's SRVF can be written as $q_i = (q_\mu, \gamma_i^{-1}) + \epsilon_i$, $i = 1, \dots, n$, where ϵ_i is an error process, and $(q, \gamma) = (q \circ \gamma)\sqrt{\dot{\gamma}}$ denotes the SRVF of the domain warping by γ of a function $f \in \mathcal{F}$ with SRVF $q \in \mathcal{Q}$. At the implementation stage, the functional data is discretized across the time domain over a fine grid $[t] = (t_1 = 0, \dots, t_M = 1)^\top$, where $t_1 < t_2 < \dots < t_M$. Assuming that the error follows a Gaussian process with white noise covariance structure, the observation model is

$$q_i([t])|q_\mu, \gamma_i, \sigma^2 \stackrel{ind}{\sim} N((q_\mu, \gamma_i^{-1})([t]), \sigma^2 I_M), \quad i = 1, \dots, n, \tag{1}$$

where I_M is the $M \times M$ identity matrix. While we assume that the variance is the same across all functions, this assumption may be relaxed; other covariance structures can also be accommodated by the model.

2.3 Signal components - template function and phase

We next present the process model involving the unknown quantities q_μ and γ_i , $i = 1, \dots, n$ (and σ^2). We assume that the SRVF template q_μ is a linear combination of B basis functions,

$$q_\mu(t) = \sum_{b=1}^B c_b \phi_b(t). \tag{2}$$

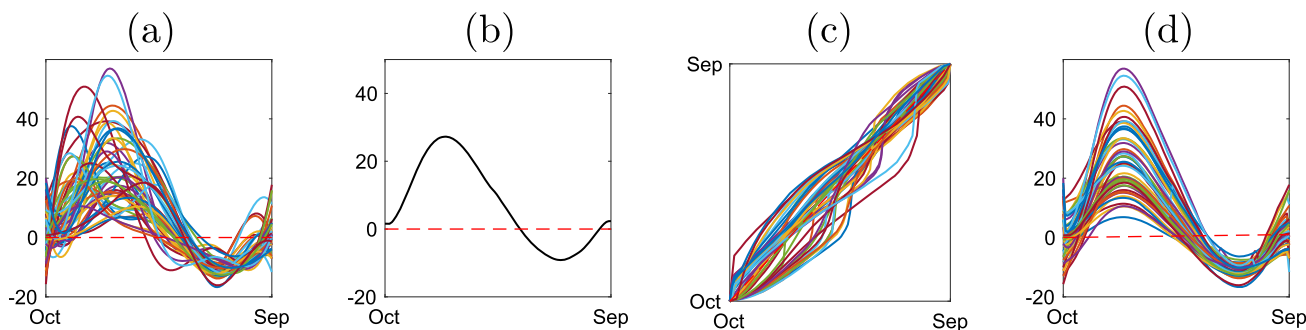


Fig. 2 Metric-based registration of trajectories of annual drought intensity near Kaweah River in California during years 1970 to 2019. (a) Observed functional data. (b) Estimated template function. (c) Estimated phase components. (d) Registered functions using the phase components in (c)

The number and type of basis functions $\phi_1, \dots, \phi_B : [0, 1] \rightarrow \mathbb{R}$ depend on the desired number of features, e.g., the number of local extrema, and the smoothness of the SRVF template. We wish to choose B that is large enough to reproduce in the template prominent features present in the functional data, but not so large as to capture variation due to noise. Specific choices of basis are discussed in Section 5 where we apply this modeling framework to simulated and real data; in all experiments, we use cubic B-splines, which resulted in satisfactory registration results. Other popular choices of bases, e.g., Fourier, which are not employed in our analyses, are discussed in Appendix E.3 in the supplement in the context of sensitivity analyses. Each phase component, γ_i , is a strictly increasing function with $\gamma_i(0) = 0$ and $\gamma_i(1) = 1$. We adopt a piecewise linear model introduced by Bharath and Kurtek (2020), which, while potentially low-dimensional, is still flexible enough to capture phase variation among functional data. The M_γ -dimensional partition of the domain $[0, 1]$ is prespecified as $0 = s_1 < s_2 < \dots < s_{M_\gamma-1} < s_{M_\gamma} = 1$, and the phase model $\gamma_i = d^{-1}(\mathbf{d}_i)$, $i = 1, \dots, n$, is the linear interpolation of phase with increments $\mathbf{d}_i = d(\gamma_i) = (\gamma_i(s_2), \dots, \gamma_i(s_m) - \gamma_i(s_{m-1}), \dots, 1 - \gamma_i(s_{M_\gamma-1}))^\top$.

2.4 Prior model

Prior choice for the template coefficients $\mathbf{c} = (c_1, \dots, c_B)^\top$ is application-specific, and adaptive models may be considered (see, e.g., Lang and Brezger 2004). For generality, we assume a multivariate normal model with mean 0_B and diagonal covariance Σ_c . This is equivalent to a B -dimensional Gaussian process prior on the SRVF template function (Lu et al. 2017). The prior hyperparameters for \mathbf{c} are taken to be fixed. Sensitivity analysis to different choices of Σ_c is presented in Appendix E.1 in the supplement.

The vector of phase increments \mathbf{d}_i must be restricted to the $(M_\gamma - 1)$ -dimensional simplex. Thus, we assign a Dirichlet

distribution as a prior model:

$$\mathbf{d}_i | \mathbf{u}_i \stackrel{ind}{\sim} Dir(\kappa \mathbf{u}_i), \quad i = 1, \dots, n, \tag{3}$$

where $\mathbf{u}_i = (u_{i(2)}, \dots, u_{i(m-1)} - u_{i(m)}, \dots, 1 - u_{i(M_\gamma-1)})^\top$ and $u_{i(2)}, \dots, u_{i(M_\gamma-1)}$ are order statistics of a random sample drawn from the uniform distribution on $[0, 1]$. The constant κ serves as a concentration parameter and is fixed. Note that this prior is centered around identity warping γ_{id} , which corresponds to the case where the template is not warped. Finally, phase increment vectors for the n functions are assumed to be *a priori* independent. In Appendix E.2 in the supplement, we study sensitivity of the posterior to different choices of the partition size M_γ .

For the auxiliary parameter σ^2 , the error variance in the SRVF space, we assume an inverse-gamma prior distribution, which is conjugate for the multivariate normal likelihood. The shape and scale hyperparameters α_σ and β_σ are fixed.

3 Sequential Monte Carlo

Sequential Monte Carlo (SMC) refers to a class of sampling algorithms targeting a sequence of prespecified distributions (Gordon et al. 1993). In the Bayesian inferential setting, these consist of either a sequence of posterior distributions or some transformation thereof. Suppose the target distributions have densities η_τ , $\tau \in \mathbb{N}_+$ over the state variables θ_τ , and are defined on a measurable space $(E_\tau, \mathcal{E}_\tau)$. SMC is a sequential version of importance sampling that generates a set of weighted samples, which are used to approximate features of each intermediate target distribution (Gordon et al. 1993). This allows the user to track uncertainty while updating inference recursively, or sequentially annealing challenging posterior distributions. For an intermediate probability density η_τ , samples are first randomly drawn from a different distribution, termed the importance distribution with density denoted by g_τ , which is easy to sample from and is available in closed form. These samples, also known as

particles, are then reweighted to reflect the shape of η_τ . To make the description more precise, let $\theta_\tau^{(j)}$, $j = 1, \dots, J$ represent J samples drawn from the importance distribution. Then, their corresponding weights are computed as $w_\tau^{(j)} \propto \eta_\tau(\theta_\tau^{(j)})/g_\tau(\theta_\tau^{(j)})$ and subsequently normalized, so that the pairs $\{(\theta_\tau^{(j)}, w_\tau^{(j)}), j = 1, \dots, J\}$ form a collection of weighted samples from $\eta_\tau(\cdot)$. Crucially, the next density $\eta_{\tau+1}$ in the sequence can then be sampled recursively starting from the weighted pairs from the importance density $g_{\tau+1} = \eta_\tau$, and so on. We present a brief overview of SMC for two different Bayesian inference scenarios that are utilized in the proposed registration approach: (1) a state space with increasing dimension and target distributions having fixed marginal densities across the sequence, and (2) a fixed-dimensional state space and subsequent target distributions that are similar.

3.1 SMC for state space with increasing dimension

SMC on a state space E_τ of increasing dimension, i.e., $\dim(E_\tau) < \dim(E_{\tau+1})$, is often of interest, such as when the target sequence of distributions consists of posteriors over an increasing number of unknown model components. Assume that we are given a set of weighted samples, $\{(\theta_\tau^{(j)}, w_\tau^{(j)}), j = 1, \dots, J\}$, drawn from the distribution with density η_τ , and we aim to modify the weights and particles such that they approximate the next target distribution in the sequence with density $\eta_{\tau+1}$. Suppose that the state variable at time $\tau + 1$ is obtained by appending a new variable $\tilde{\theta}$ to the previous state, $\theta_{\tau+1} = (\theta_\tau, \tilde{\theta})$, and the marginal density of θ_τ at time $\tau + 1$ is equal to the density at time τ : $\eta_\tau(\theta_\tau) = \eta_{\tau+1}(\theta_\tau)$. Interest lies in approximating $\eta_{\tau+1}$ using $\{(\theta_\tau^{(j)}, w_\tau^{(j)}), j = 1, \dots, J\}$ as well as random samples from the conditional distribution of $\tilde{\theta}$ given θ_τ . Liu and Chen (1998) present an SMC sampler for such a scenario, which is described below.

Since the marginal distribution of θ_τ does not change from time τ to $\tau + 1$, we update the existing samples to $\theta_{\tau+1}^{(j)} = (\theta_\tau^{(j)}, \tilde{\theta}^{(j)})$, $j = 1, \dots, J$ by appending $\tilde{\theta}^{(j)}$ generated via $\tilde{K}_{\tau+1}(\tilde{\theta}^{(j)} | \theta_\tau^{(j)})$. This defines a Markov transition kernel $K_{\tau+1}(\theta_{\tau+1}^{(j)} | \theta_\tau^{(j)}) := \eta_\tau(\theta_\tau^{(j)})\tilde{K}_{\tau+1}(\tilde{\theta}^{(j)} | \theta_\tau^{(j)})$ on $E_{\tau+1} \times \mathcal{E}_{\tau+1}$. An efficient choice for $\tilde{K}_{\tau+1}$ is one that targets the conditional density of $\tilde{\theta}$ given θ_τ . The unnormalized weights of $\theta_{\tau+1}^{(j)}$, $j = 1, \dots, J$ are updated using

$$w_{\tau+1}^{(j)} \propto \frac{\eta_{\tau+1}(\theta_{\tau+1}^{(j)})}{g_{\tau+1}(\theta_{\tau+1}^{(j)})} = w_\tau^{(j)} \frac{\eta_{\tau+1}(\theta_{\tau+1}^{(j)})}{\eta_\tau(\theta_\tau^{(j)})\tilde{K}_{\tau+1}(\tilde{\theta}^{(j)} | \theta_\tau^{(j)})}, \tag{4}$$

where $g_{\tau+1}$ denotes the importance density for sampling $\theta_{\tau+1}^{(j)}$; once computed, the weights can be normalized. This

framework is widely used for dynamical systems with state space models increasing in dimension (Liu and Chen 1998).

If we recursively update weighted samples targeting a long sequence of distributions with increasing state space dimension, at some point the weighted samples may poorly estimate the target distribution due to the curse of dimensionality. For example, when using J weighted samples to approximate the target distribution at time τ , fewer particles retain large weights as the dimension increases with τ for subsequent target distributions. This is known as the degeneracy problem. To measure potential degeneracy, we compute the effective sample size (ESS) using the magnitude of the normalized weights, $ESS = \left(\sum_{j=1}^J (w_\tau^{(j)})^2\right)^{-1}$. If ESS is small, a large portion of the weighted samples have small weights suggesting particle degeneracy. One solution is to remove particles with small weights and duplicate ones with large weights through resampling (Gilks and Berzuini 2001; Chopin 2002). A common approach draws random samples from the multinomial distribution with weights serving as parameters; the resampled particles are then assigned equal weights (Gordon et al. 1993). The overall procedure is as follows: (1) generate samples $\theta_{\tau+1}$ via Markov transition kernel $K_{\tau+1}$, (2) update the weights using Equation 4, and (3) resample the weighted samples if $ESS < J/2$.

3.2 SMC for state space with fixed dimension

Another scenario of interest is when the sequence of distributions η_τ , $\tau \in \mathbb{N}_+$ are defined on the same space E , and thus have fixed dimension. This occurs, for example, when we wish to sample from a challenging target posterior distribution by building a sequence, or bridge, of intermediate distributions that are similar to one another, such as annealed versions of the posterior density. Assume that we have a set of weighted samples from a distribution at time τ and we aim to perturb them toward the target distribution at time $\tau + 1$. We assume that the dimensions of the state variables at times τ and $\tau + 1$ are the same, i.e., $\dim(\theta_\tau) = \dim(\theta_{\tau+1})$, and that the adjacent distributions are similar to each other, i.e., $\eta_\tau(\theta_\tau) \approx \eta_{\tau+1}(\theta_{\tau+1})$. Given a set of weighted samples at time τ , $\{(\theta_\tau^{(j)}, w_\tau^{(j)}), j = 1, \dots, J\}$, we perturb them toward the target distribution $\eta_{\tau+1}$ via a transition kernel $K_{\tau+1}(\theta_{\tau+1}^{(j)} | \theta_\tau^{(j)})$ defined on $E \times \mathcal{E}$. A natural choice for $K_{\tau+1}$ is an MCMC kernel. The weight update is given by

$$w_{\tau+1}^{(j)} \propto \frac{\eta_{\tau+1}(\theta_{\tau+1}^{(j)})}{g_{\tau+1}(\theta_{\tau+1}^{(j)})} = \frac{\eta_{\tau+1}(\theta_{\tau+1}^{(j)})}{\int g_\tau(\theta_\tau) K_{\tau+1}(\theta_{\tau+1}^{(j)} | \theta_\tau) d\theta_\tau}, \tag{5}$$

$j = 1, \dots, J.$

This requires updating the importance density which is, however, often intractable as it involves integration that does not have a closed form solution. To circumvent this issue,

Del Moral et al. (2006) calculate the weights through the use of a backward kernel $L_\tau(\theta_\tau^{(j)} | \theta_{\tau+1}^{(j)})$, defined on $E \times \mathcal{E}$, as

$$w_{\tau+1}^{(j)} \propto w_\tau^{(j)} \frac{\eta_{\tau+1}(\theta_{\tau+1}^{(j)}) L_\tau(\theta_\tau^{(j)} | \theta_{\tau+1}^{(j)})}{\eta_\tau(\theta_\tau^{(j)}) K_{\tau+1}(\theta_{\tau+1}^{(j)} | \theta_\tau^{(j)}), \quad j = 1, \dots, J, \tag{6}$$

if the transition kernel has a closed form. If $K_{\tau+1}$ is an MCMC kernel that is invariant to $\eta_{\tau+1}$, and the adjacent target distributions are similar, Del Moral et al. (2006) proposed using the following approximate backward kernel that avoids explicit evaluation of $K_{\tau+1}$ and L_τ in the weight update: $L_\tau(\theta_\tau^{(j)} | \theta_{\tau+1}^{(j)}) = \frac{\eta_{\tau+1}(\theta_\tau^{(j)}) K_{\tau+1}(\theta_{\tau+1}^{(j)} | \theta_\tau^{(j)})}{\eta_{\tau+1}(\theta_{\tau+1}^{(j)})}$, $j = 1, \dots, J$. For this choice of backward kernel, the weight update in Equation 6 simplifies to

$$w_{\tau+1}^{(j)} \propto w_\tau^{(j)} \frac{\eta_{\tau+1}(\theta_\tau^{(j)})}{\eta_\tau(\theta_\tau^{(j)})}, \quad j = 1, \dots, J, \tag{7}$$

and the weights are then normalized.

4 SMC algorithm for registration of functional data

In this section, we present a sequential Bayesian registration approach for the model introduced in Section 2, noting that it can be straightforwardly extended to a wider variety of models for functional data, for which the state space similarly increases in dimension as new data arrives; a detailed algorithm is provided in Appendix A in the supplement. In the following, we will use the subscript $1 : n$ to denote a set of objects indexed from 1 through n , so that, for example, $f_{1:n} = \{f_1, \dots, f_n\}$ represents the first n observed functions and $\mathbf{d}_{1:n} = \{\mathbf{d}_1, \dots, \mathbf{d}_n\}$ represents the increments of the phase components for the first n functions. We introduce a sequential Monte Carlo algorithm that updates the posterior distribution over the template function, all of the phase components, and the error variance when the state space of the phase components increases as new functions arrive. Let $\dots, \pi(\mathbf{c}, \mathbf{d}_{1:n}, \sigma^2 | f_{1:n}), \pi(\mathbf{c}, \mathbf{d}_{1:n+1}, \sigma^2 | f_{1:n+1}), \pi(\mathbf{c}, \mathbf{d}_{1:n+2}, \sigma^2 | f_{1:n+2}), \dots$ denote the sequence of target posterior densities over all of the parameters given an increasing number of functions, and suppose that we have a large number J of weighted samples $\{(\mathbf{c}^{(j)}, \mathbf{d}_{1:n}^{(j)}, \sigma^{2(j)}, w^{(j)}), j = 1, \dots, J\}$ approximating the posterior $\pi(\mathbf{c}, \mathbf{d}_{1:n}, \sigma^2 | f_{1:n})$, where \mathbf{c} is the vector of basis coefficients defining the template function and σ^2 is the observation error variance. The goal is to update these particles and weights to generate a new weighted sample

$\{(\mathbf{c}^{(j)}, \mathbf{d}_{1:n+1}^{(j)}, \sigma^{2(j)}, w^{(j)}), j = 1, \dots, J\}$, approximating the posterior $\pi(\mathbf{c}, \mathbf{d}_{1:n+1}, \sigma^2 | f_{1:n+1})$, as a new function f_{n+1} becomes part of the data.

4.1 Stochastic initialization of new phase components

Since additional parameters are required in the model when a new function, f_{n+1} , is observed, we consider the increasing state space dimension setting described in Section 3.1. The first step (lines 4-10 in Algorithm 1) augments the previous particles with phase component $\mathbf{d}_{n+1}^{(j)}$ for the new function f_{n+1} , and updates their weights. To enhance the efficiency of the sampling algorithm, we initialize this new component in a region of high posterior probability. The initialization kernel leaves the existing particle components unchanged and draws the phase increment $\mathbf{d}_{n+1}^{(j)}$ from a distribution centered around the increment $d(\hat{\gamma}_{n+1}^{(j)})$ obtained by optimally aligning f_{n+1} to the template $q_\mu^{(j)} = \sum_{b=1}^B c_b^{(j)} \phi_b$ as follows.

1. Compute the phase function $\hat{\gamma}_{n+1}^{(j)}$ that provides the optimal alignment to the template $q_\mu^{(j)} = \sum_{b=1}^B c_b^{(j)} \phi_b$ by solving the optimization problem $\hat{\gamma}_{n+1}^{(j)} = \arg \min_{\gamma \in \Gamma} \|q_\mu^{(j)} - (q_{n+1} \circ \gamma) \sqrt{\gamma}\|^2$ via the Dynamic Programming algorithm (Bertsekas 2000). Each resulting $\hat{\gamma}_{n+1}^{(j)}$ is a piecewise linear function that is finely sampled on the domain $[0, 1]$. For the next step, we further approximate $\hat{\gamma}_{n+1}^{(j)}$ over the prespecified partition $0 = s_1 < s_2 < \dots < s_{M_\gamma-1} < s_{M_\gamma} = 1$ using a least squares procedure.
2. Draw a random sample \mathbf{d}^p from $Dir((\kappa_{ini}/(M_\gamma - 1)) \mathbf{1}_{M_\gamma-1})$ and initialize the new phase component by composition via $\gamma^{(j)*} = \hat{\gamma}_{n+1}^{(j)} \circ d^{-1}(\mathbf{d}^p)$.

The sampling density (Zang 2021) of these new phase increments $\mathbf{d}_{n+1}^{(j)} = d(\gamma^{(j)*})$ is

$$\tilde{K}(\mathbf{d}_{n+1}^{(j)} | \mathbf{c}^{(j)}, f_{n+1}) = \left[\prod_{m=2}^{M_\gamma} \left(\left(\hat{\gamma}_{n+1}^{(j),-1} \right)' \circ \gamma^{(j)*} \right) (s_m) \right] Dir(\mathbf{d}^p; \frac{\kappa_{ini}}{M_\gamma - 1} \mathbf{1}_{M_\gamma-1}), \tag{8}$$

where $(\hat{\gamma}_{n+1}^{(j),-1})'$ is the derivative of $\hat{\gamma}_{n+1}^{(j),-1}$, and the concentration parameter κ_{ini} is user-selected. We provide a sensitivity analysis to the choice of κ_{ini} in Appendix E.4 in the supplement. Next, we update the weight of each particle while accounting for the increased dimension of the parameter space, following the approach of Liu and Chen (1998). Denote the likelihood by h , the prior density by p , the importance sampling density by g , and the initialization kernel density by \tilde{K} . The importance sampling

density after augmentation is related to the previous importance sampling density via $g(\mathbf{c}^{(j)}, \mathbf{d}_{1:n+1}^{(j)}, \sigma^{2(j)} \mid f_{1:n+1}) = g(\mathbf{c}^{(j)}, \mathbf{d}_{1:n}^{(j)}, \sigma^{2(j)} \mid f_{1:n}) \tilde{K}(\mathbf{d}_{n+1}^{(j)} \mid \mathbf{c}^{(j)}, f_{n+1})$. The new weights are then given by

$$\begin{aligned} \tilde{w}^{(j)} &\propto \frac{\pi(\mathbf{c}^{(j)}, \mathbf{d}_{1:n+1}^{(j)}, \sigma^{2(j)} \mid f_{1:n+1})}{g(\mathbf{c}^{(j)}, \mathbf{d}_{1:n+1}^{(j)}, \sigma^{2(j)} \mid f_{1:n+1})} \\ &\propto \frac{h(f_{1:n+1} \mid \mathbf{c}^{(j)}, \mathbf{d}_{1:n+1}^{(j)}, \sigma^{2(j)}) p(\mathbf{c}^{(j)}, \mathbf{d}_{1:n+1}^{(j)}, \sigma^{2(j)})}{g(\mathbf{c}^{(j)}, \mathbf{d}_{1:n}^{(j)}, \sigma^{2(j)} \mid f_{1:n}) \tilde{K}(\mathbf{d}_{n+1}^{(j)} \mid \mathbf{c}^{(j)}, f_{n+1})} \\ &\propto w^{(j)} \frac{h(f_{n+1} \mid \mathbf{c}^{(j)}, \mathbf{d}_{n+1}^{(j)}, \sigma^{2(j)}) p(\mathbf{d}_{n+1}^{(j)})}{\tilde{K}(\mathbf{d}_{n+1}^{(j)} \mid \mathbf{c}^{(j)}, f_{n+1})}, \end{aligned} \tag{9}$$

for each $j = 1, \dots, J$, where $w^{(j)}$ denotes the weight before augmentation. Thus, the resulting weighted sample, composed of the augmented particles and updated weights, now targets the posterior density given $f_{1:n+1}$.

4.2 Particle perturbation and centering

While the weighted samples $\{(\mathbf{c}^{(j)}, \mathbf{d}_{1:n+1}^{(j)}, \sigma^{2(j)}, \tilde{w}^{(j)}), j = 1, \dots, J\}$ already approximate the target posterior $\pi(\mathbf{c}, \mathbf{d}_{1:n+1}, \sigma^2 \mid f_{1:n+1})$, the presence of particles with small weights as well as low particle diversity, can result in SMC estimators with large variance. Thus, we resample the particles if ESS falls below $J/2$ using a multinomial distribution with parameters $\tilde{w}^{(j)}, j = 1, \dots, J$, assigning equal weights to the resampled particles. Furthermore, since resampling can result in multiple copies of particles, we use MCMC perturbations to further diversify the particle locations. Numerical experiments suggest that performing this step guards against particle degeneracy in subsequent updates. Since perturbation occurs on a state space of fixed dimension, we use the tools described in Section 3.2. We begin by perturbing particle components $\mathbf{c}^{(j)}$ and $\mathbf{d}_{1:n+1}^{(j)}$, given $\sigma^{2(j)}, j = 1, \dots, J$ using a Metropolis-Hastings (MH) transition kernel (lines 17-35 in Algorithm 1). We adopt the approximate backward kernel approach of Del Moral et al. (2006) to avoid computing the density of the MH kernel in the weight update calculation. Because the additional perturbation still targets the desired posterior $\pi(\mathbf{c}, \mathbf{d}_{1:n+1}, \sigma^2 \mid f_{1:n+1})$, i.e. $\eta_{\tau+1} = \eta_{\tau}$ in Equation 7, the weight update simplifies to $\tilde{w}^{(j)} = \tilde{w}^{(j)}$. This argument also enables the application of multiple subsequent MH perturbation steps to further diversify the sample, if desired. Our implementation performs 30 MH perturbation steps. The proposal distribution for the template coefficients (line 18 in Algorithm 1) is multivariate normal $N(\mathbf{c}^{(j)}, \widehat{\Sigma}_c)$ centered at the j th particle, $\mathbf{c}^{(j)}$, with empirical covariance $\widehat{\Sigma}_c = \frac{1}{J-1} \sum_{j=1}^J \tilde{w}^{(j)} \mathbf{c}^{(j)} \mathbf{c}^{(j)\top}$. Phase increment proposals are generated by composing $\gamma^{(j)}$ with the linear interpola-

tion of a $Dir(\frac{\kappa}{M_{\gamma}-1} \mathbf{1}_{M_{\gamma}-1})$ realization, where κ is fixed (lines 26-29 in Algorithm 1). An analogous update involving $\sigma^{2(j)}$ is performed using a Gibbs sampling kernel following the centering step described below.

After each full update, we perform a centering step to ensure identifiability of the template function (line 36 in Algorithm 1; see Section 2.1 for a brief discussion). This step constrains the mean of the phase components of the data with respect to the template function to be the identity γ_{id} . To do this, we first compute the sample average of the phase components $\gamma_{1:n+1}^{(j)} = d^{-1}(\mathbf{d}_{1:n+1}^{(j)})$ (Srivastava et al. 2011) and then apply its inverse to each template function $q_{\mu}^{(j)} = \sum_{b=1}^B c_b^{(j)} \phi_b$ and each phase $\gamma_{1:n+1}^{(j)}$. We then update the weights accordingly. As this step utilizes a deterministic kernel, and the importance density and likelihood are invariant to simultaneous warping, the weight update is based on the ratio of prior densities evaluated at the centered particles and their uncentered counterparts. To do this, we first use the centered template function and phase components to construct the centered particles $\{(\tilde{\mathbf{c}}^{(j)}, \tilde{\mathbf{d}}_{n+1}^{(j)}), j = 1, \dots, J\}$. We then update the weights (line 37 in Algorithm 1) using

$$\tilde{w}^{(j)} \leftarrow \tilde{w}^{(j)} \frac{p(\tilde{\mathbf{c}}^{(j)}, \tilde{\mathbf{d}}_{1:n+1}^{(j)})}{p(\mathbf{c}^{(j)}, \mathbf{d}_{1:n+1}^{(j)})}, \quad j = 1, \dots, J, \tag{10}$$

which are normalized subsequently. Assuming that the initial samples are centered (this is the standard approach in MCMC-based batch learning for registration of functional data), the sample average of the phase components does not deviate very much from the identity after each new function is observed. This fact, coupled with our specification of diffuse priors, ensures that the weights do not change much due to this centering step. Thus, to improve computational efficiency, one may choose to not update the weights of the particles after the centering step.

After updating the template and phase components of the weighted particles, the components corresponding to the error variance σ^2 are perturbed via a Gibbs update (lines 38-39 in Algorithm 1). This is done by sampling $\tilde{\sigma}^{2(j)}$, for each $j = 1, \dots, J$, from the full-conditional distribution, $\pi(\sigma^2 \mid \tilde{\mathbf{c}}^{(j)}, \tilde{\mathbf{d}}_{1:n+1}^{(j)}, f_{1:n+1})$, which is inverse-gamma with parameters $\alpha_{\sigma} + \frac{(n+1)M}{2}$ and $\beta_{\sigma} + \sum_{i=1}^{n+1} \sum_{m=1}^M (q_i(t_m) - (\tilde{q}_{\mu}^{(j)}, \tilde{\gamma}_i^{-1(j)})(t_m))^2$, where $\tilde{q}_{\mu}^{(j)} = \sum_{b=1}^B \tilde{c}_b^{(j)} \phi_b$ and $\tilde{\gamma}_i^{(j)} = d^{-1}(\tilde{\mathbf{d}}_i^{(j)})$, $i = 1, \dots, n+1$. Again adopting the approximate backward kernel approach of Del Moral et al. (2006), the weights remain approximately unchanged after the Gibbs update.

Following the above resampling, perturbation, and centering steps with associated weight updates, we obtain a collection of J weighted samples $\{(\tilde{\mathbf{c}}^{(j)}, \tilde{\mathbf{d}}_{1:n+1}^{(j)}, \tilde{\sigma}^{2(j)}, \tilde{w}^{(j)}), j =$

$1, \dots, J\}$ approximating the target posterior distribution $\pi(\mathbf{c}, \mathbf{d}_{1:n+1}, \sigma^2 \mid f_{1:n+1})$.

5 Simulations and real data examples

We begin in Section 5.1 by discussing visualization and summarization of posterior uncertainty resulting from our analyses. The following sections describe in detail the application of the proposed sequential Bayesian elastic registration framework to two simulated examples and three real data studies. For these analyses we assume that the template function is a weighted sum of cubic B-splines with equally-spaced knots, as they are sufficiently flexible to capture the expected variation in the template. Our model also accommodates the use of other basis sets. The number of bases is chosen to reflect the complexity of amplitude features in the data without over-fitting. For the simulated examples, we use the same number of basis functions to model the template as were used to simulate the data. In the following examples, we set $\kappa_{ini} = 100$ for the initialization kernel of each new phase component in the SMC algorithm. Our implementation employs parallel computing with 12 workers.

5.1 Summarization and visualization of target posterior distribution

An important consideration is how to visualize marginal posteriors over template and phase in their respective function spaces. To extract posterior features we first map the particle components $\tilde{\mathbf{c}}^{(j)}$ and $\tilde{\mathbf{d}}_i^{(j)}$ to the template SRVF $\tilde{q}_\mu^{(j)} = \sum_{b=1}^B \tilde{c}_b^{(j)} \phi_b \in \mathcal{Q}$ and the phase $\tilde{\gamma}_i^{(j)} = d^{-1}(\tilde{\mathbf{d}}_i^{(j)})$, $i = 1, \dots, n$, $j = 1, \dots, J$, respectively. The posterior mean of q_μ is estimated using the weighted sample average $\sum_{j=1}^J \tilde{w}^{(j)} \tilde{q}_\mu^{(j)} =: \bar{q}_\mu$, while its variance is estimated using the weighted sample average of the squared eFR distance of each particle from the posterior mean, i.e., $\sum_{j=1}^J \tilde{w}^{(j)} \int_0^1 (\tilde{q}_\mu^{(j)}(t) - \bar{q}_\mu(t))^2 dt$. Pointwise estimates may be used to visualize summaries of uncertainty at specific times, and posterior summaries for phase are computed similarly.

We use smooth line plots of multiple superimposed trajectories across time, known as *spaghetti plots*, to visualize marginal posteriors over q_μ and $\gamma_{1:n}$ by plotting each SMC particle across time with partial transparency proportional to the weights of the individual particles. This enables visualization of posterior features such as dispersion and clustering, which are difficult to summarize in the functional setting. Visualizing the uncertainty in the *shape* of the template function can be more intuitive in the original function space \mathcal{F} rather than the SRVF space \mathcal{Q} . However, since SRVFs do not contain translation information, the spaghetti plot represen-

tation in the original space can appear misleading because starting points must be chosen arbitrarily. To circumvent this issue, we propose to visualize the leading sample principal components computed using the template particles to illustrate uncertainty in template shape. We first compute the $M \times M$ pointwise covariance of the marginal posterior of the template function in the SRVF space, $\Sigma_q = \sum_{j=1}^J \tilde{w}^{(j)} (\tilde{q}_\mu^{(j)}([t]) - \bar{q}_\mu([t])) (\tilde{q}_\mu^{(j)}([t]) - \bar{q}_\mu([t]))^\top$ where \bar{q}_μ is the marginal posterior mean, and perform singular value decomposition, $\Sigma_q = U \Delta V^\top$, where the diagonal entries of Δ are the singular values of Σ_q and the columns of U are orthonormal bases u_1, \dots, u_M . We may then obtain positive/negative directions along the k th principal component (in units of standard deviation) from the marginal posterior mean by $\bar{q}_\mu + / - \alpha \sqrt{\Delta_k} u_k$, where α is a positive number and Δ_k is the k th diagonal entry of Δ . We can visualize the uncertainty in template shape by mapping the functions represented by these vectors to the original function space \mathcal{F} via the inverse mapping Q^{-1} , where the template function value at $t = 0$ is taken as the sample average $\frac{1}{n} \sum_{i=1}^n f_i(0)$. This visualization approach is used in Section 5.6.

5.2 Simulated example 1

We simulate 100 functions by deforming a ground truth template q_μ using randomly sampled phases γ_i , $i = 1, \dots, 100$. The template is generated using a linear combination of eight B-spline basis functions and the phases as piecewise linear functions with Dirichlet increments on a uniform partition of size $M_\gamma = 5$ with concentration parameter $\kappa = 50$. The functional data is then generated via $f_i = Q^{-1}(f_i(0) = 0, q_i)$, with $q_i([t]) = (q_\mu, \gamma_i^{-1})([t]) + \epsilon_i([t])$ and $\epsilon_i([t]) \stackrel{iid}{\sim} N(0_M, 0.03 I_M)$, where 0_M is a vector of $M = 100$ zeros. A subset of the 100 simulated functions is shown in Figure 3(a) with the corresponding subset of ground truth phase components in red in Figure 3(b). We fit the model described in Section 2 with prior hyperparameters $\Sigma_c = 20 I_8$, $\kappa = 5$, $\alpha_\sigma = 4$, $\beta_\sigma = 0.01$, and a uniform partition of size $M_\gamma = 5$.

We first obtain an MCMC sample of size 10,000 from the posterior distribution of the parameters given the first 30 functions, $f_{1:30}$, and use these as equally weighted particle inputs to the subsequent SMC update. The weighted samples are updated recursively using the proposed SMC algorithm adding one function at a time. The resulting summaries of the posterior distribution given the full data $f_{1:100}$ are shown in Figure 3. Panel (b) shows the estimated posterior mean phases in black, noting that they are very similar to the ground truth. The marginal posterior variances obtained using the weighted sample average of the squared \mathbb{L}^2 distances of each particle from the marginal posterior mean are very small ($< 10^{-4}$). Panel (c) illustrates the posterior uncertainty in the template function in the SRVF space \mathcal{Q} using

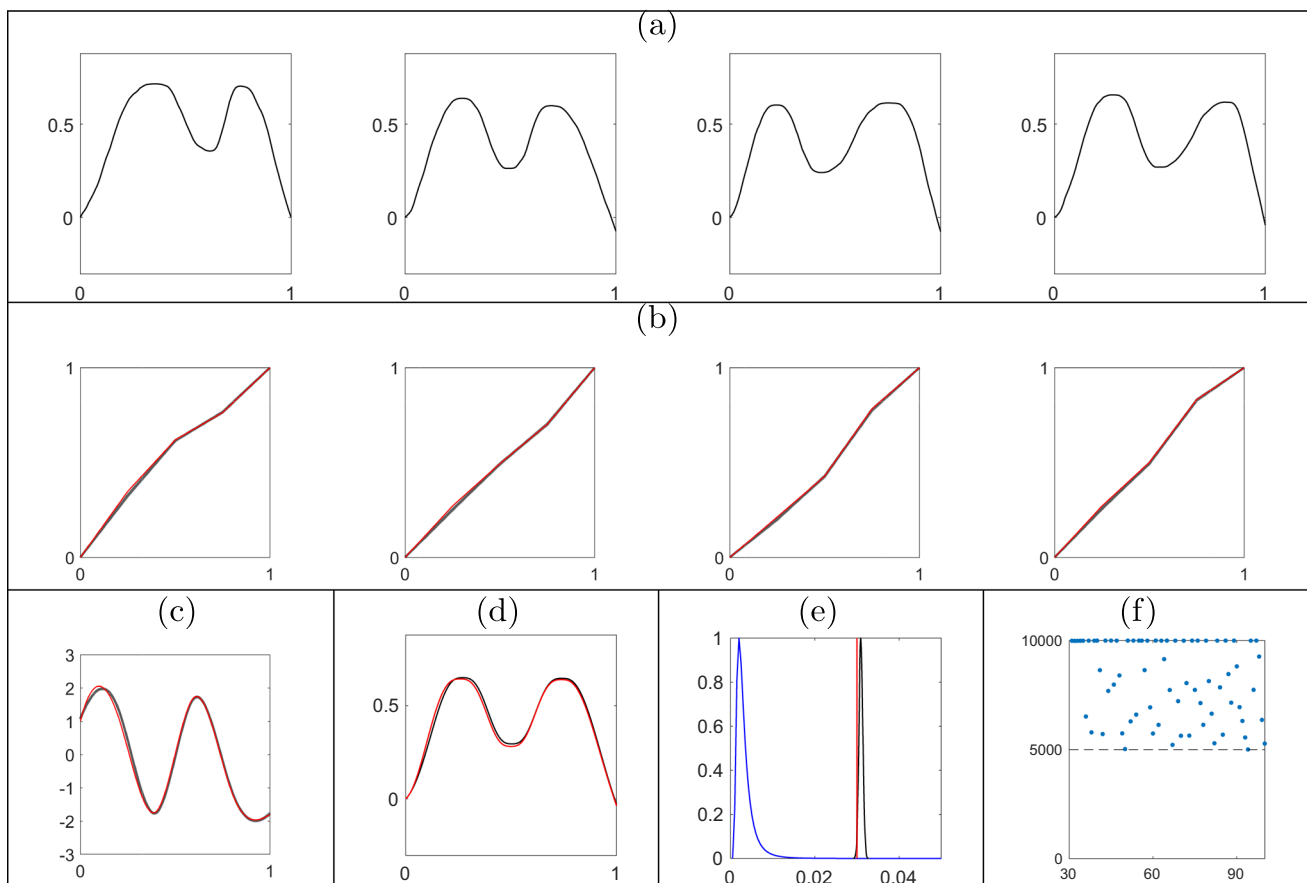


Fig. 3 Illustration of the proposed approach based on simulated data. Results in the second and third rows were generated via recursive updates starting with $f_{1:30}$ and ending with $f_{1:100}$. (a) Subset of simulated functions $f_{1:100}$. (b) Marginal posterior means of phase components for the functions in (a) are shown in the same column in black with ground truth in red. (c) Summary of the marginal posterior for the template SRVF using spaghetti plots with transparency (in black) and the ground truth template function (in red). (d) The marginal posterior

mean of the template function (in black) and the ground truth template function (in red) in the original function space \mathcal{F} . (e) Kernel density estimate of the marginal posterior for the error variance σ^2 in black. The prior distribution of σ^2 and the ground truth are shown in blue and red, respectively. (f) ESS of weighted samples for a sequence of posterior distributions, $\pi(\cdot | f_{1:n})$, $n = 31, \dots, 100$, with n on x -axis and ESS on y -axis

spaghetti plots (in black) with transparency and the ground truth template function in red. The magnitude of uncertainty in the template is small across the domain. The posterior mean template in the original function space \mathcal{F} is shown in panel (d) in black and the ground truth template in red. The eFR distance between the posterior mean SRVF template and the ground truth is only 0.1080. Panel (e) shows that the true error variance lies within the 99% central credible region of the marginal posterior for σ^2 . Panel (f) shows updated ESS values after resampling the particles when the ESS of the previous particles fell below the threshold of $J/2$ (line 41 of Algorithm 1 in Appendix A in the supplement). The figure suggests that ESS remains relatively large as we add more functions to the data. After augmenting the previous particles with a new phase component and updating the weights using

Equation 9, the ESS sometimes falls below the threshold, but is still moderately large ($> J/3$).

To compare estimation accuracy, we fit the model to the full data $f_{1:100}$ using both the sequential and MCMC-based batch learning methods. Our implementation of MCMC uses Gibbs and adaptive Metropolis-Hastings updates for a total of 50,000 iterations, with a burn-in period of 40,000. Table 1 reports the estimated mean squared errors for two posterior summaries of the template basis coefficients and the phase increments, obtained based on 100 runs of the batch and sequential learning algorithms. The proposed sequential approach outperforms MCMC-based batch learning in all scenarios. While not presented in this table, similar results were obtained when comparing estimates from intermediate posterior distributions in the sequence to batch learning posterior estimates given the same set of functional data.

Table 1 Estimated mean squared errors of the posterior mean and posterior mode for the template basis coefficients \mathbf{c} and phase increments $\mathbf{d}_{1:100}$ based on 100 replications of the batch learning (MCMC) and sequential (SMC) algorithm given $f_{1:100}$. The best performance is shown in bold

Method	Template Function (\mathbf{c})		Phase Components ($\mathbf{d}_{1:100}$)	
	Posterior Mean	Posterior Mode	Posterior Mean	Posterior Mode
SMC	0.1712	0.1491	0.0120	0.0079
MCMC	0.2200	0.2457	0.0121	0.0108

Furthermore, we see no evidence of degraded accuracy as the number of assimilated functions grows (see Appendix D in the supplement).

We further compare the computational efficiency of the two methods. For the proposed sequential approach, we start with posterior samples given $f_{1:30}$, recursively add one function to the data at a time, and update the weighted samples using the proposed algorithm. We do this for the full sequence of posterior distributions until all of the data, $f_{1:100}$, is used. For the MCMC-based batch learning approach, we draw samples from each posterior distribution in the sequence by re-running the full algorithm each time. As before, we use a total of 50,000 MCMC iterations, with a burn-in period of 40,000. The computation time (in seconds) needed to obtain a Monte Carlo sample of size 10,000 from each posterior distribution in the sequence given $f_{1:n}$, $n = 31, \dots, 100$ is decreased by $> 80\%$ on average. The computation time to obtain a posterior sample via MCMC-based batch learning given $f_{1:30}, \dots, f_{1:100}$ is 322,906 seconds. This is substantially longer than the 58,568 seconds for the sequential approach, which includes (i) MCMC-based batch learning to first obtain equally weighted particles from the posterior distribution given $f_{1:30}$, and (ii) SMC-based updates of the posterior given $f_{1:31}, \dots, f_{1:100}$. This significant gain in computational efficiency is expected due to the sequential nature of the proposed method. A comparison of ESS based on the two approaches is presented in Appendix D in the supplement.

5.3 Simulated example 2

Next, we assess the performance of the proposed sequential approach for registration of functional data when the target posterior density is multimodal. In what follows we reserve the term “multimodal” to describe posterior densities rather than the functional observations, some of which will have two peaks. We simulate seven functions: the first six have two peaks and a valley and are simulated in the same fashion as Simulated Example 1, but with an additional random scaling (sampled from a uniform distribution on $[0.7, 1.4]$) applied to each of $f_{1:6}$. The seventh function, f_7 , is simulated such that it only has one peak. We expect the marginal posterior of the template function given $f_{1:7}$ to be unimodal with the mode representing a template that has two peaks,

since most of the data has this form. However, since phase is relative with respect to the template, the marginal posterior samples of the phase component for function f_7 should cluster into two distinct groups: one that registers the single peak in function f_7 to the left peak of the template, and one that registers it to the right peak of the template. Thus, we expect the marginal posterior of phase for functions $f_{1:6}$ to be unimodal and for function f_7 to be bimodal. Figures 4(a1) and (a2) show a subset of the simulated two-peak functions, $f_{1:6}$, and the single-peak function, f_7 , respectively. All prior hyperparameters in the model are set to the same values as in Simulated Example 1, except for $\alpha_\sigma = 40$.

We initialize the SMC algorithm using MCMC samples from the posterior distribution given $f_{1:3}$. We then assimilate the data $f_{4:7}$, one function at a time, and update the weighted samples sequentially to target the posteriors $\pi(\cdot | f_{1:4}), \dots, \pi(\cdot | f_{1:7})$. The results of this simulation are presented in Figure 4. As shown in panels (d)-(e) and (f)-(g), the template components of the particles indeed appear to have two peaks before and after the addition of f_7 to the data. It is evident that posterior uncertainty in (f) is greater than in (d), which is expected due to the very different shape of f_7 . The marginal posteriors of phase for functions $f_{1:6}$, shown in panel (b1), suggest a unimodal marginal posterior; the marginal posteriors of phase have very small pointwise variances and thus we only visualize the marginal posterior means here. However, the particles of the phase component corresponding to function f_7 in panel (b2) clearly fall into two distinct groups as the peak of function f_7 can be registered to either of the two peaks in the marginal posterior of the template function, leading to a bimodal marginal posterior density. One mode corresponds to phase functions that fall above the identity with pointwise weighted sample average in red in panel (b2); this corresponds to the registration of the single peak in f_7 to the first peak in the template as seen in (c2). The second mode corresponds to phase functions that fall below the identity with pointwise weighted sample average in blue in (b2); this corresponds to the registration of the single peak in f_7 to the second peak in the template as seen in (c2). It is well-known that, in practice, MCMC can be quite inefficient in sampling from multimodal posterior distributions. On the other hand, the proposed sequential algorithm is able to sample from both modes in this posterior density relatively easily.

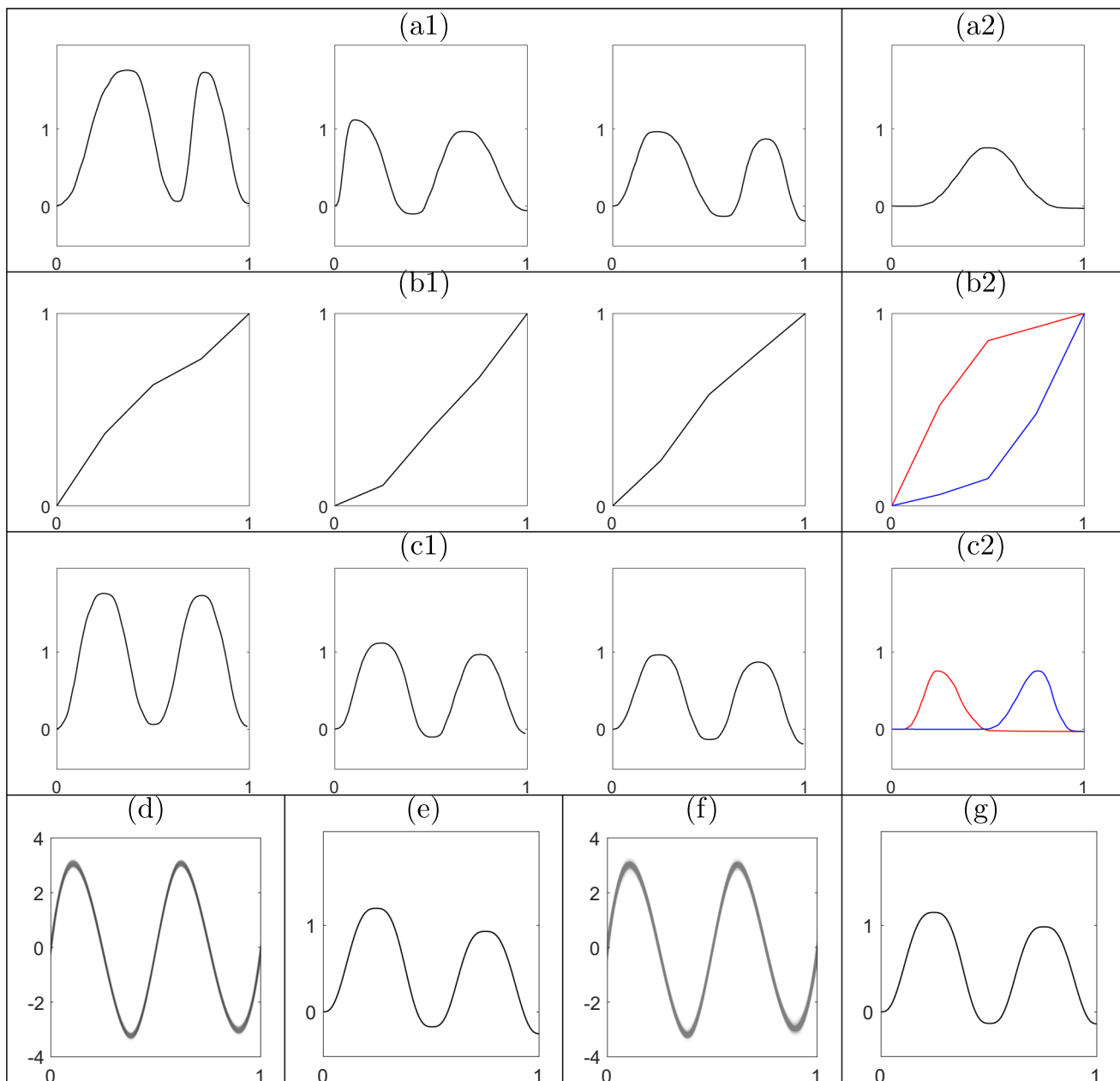


Fig. 4 Illustration of the proposed approach under a bimodal target posterior. (a1) Subset of simulated data $f_{1:6}$ and (a2) simulated data f_7 with one peak. (b1) & (b2) Posterior mean estimates of the corresponding phases. (b2) Weighted mean for each of two modes in the marginal posterior. (c1) & (c2) Registered functions $f_{1:6}$ and f_7 , respectively, using

posterior mean phases from (b1) & (b2). (d) & (f) Marginal posterior over the template SRVF given $f_{1:6}$ and $f_{1:7}$, respectively; transparency reflects magnitude of weights. (e) & (g) Marginal posterior mean of the template in the original function space \mathcal{F} given $f_{1:6}$ and $f_{1:7}$, respectively

This simulation study also provides empirical evidence that the proposed method is effective at approximating the posterior distribution even when a new observation with a very different shape becomes part of the data. One concern in such a scenario for the proposed SMC algorithm is that the target posterior may change to an extent that would violate some of the assumptions required for the weight update

calculations in Section 4. In particular, marginal posterior variance for the template function increases from 0.0038 to 0.0089 after the arrival of f_7 , as seen in panels (d) and (f). At the same time, the distance between the marginal posterior means of the template for the two cases, which are shown in (e) and (g), is small (< 0.1). This shows that adding a function with a different shape does not generally change the

marginal posteriors very much, even given a relatively small number of functions prior to the sequential update (we start with only six two-peaked functions in the data). Even after more functions with one peak arrive sequentially, we expect the marginal posterior that is centered at a template function with two peaks to smoothly change toward a marginal posterior centered at a template with one peak due to additional MH-based particle perturbation steps employed in our algorithm. Finally, this example illustrates that initialization of the SMC particles can be based on MCMC samples given a relatively small number of initial functional data. In general, the number of functions required for initialization depends on the expected change in the marginal posterior distribution of the template function as new data arrive. We found through simulation that initializing the particles from an MCMC sample given 10-20 functional observations yields satisfactory results.

5.4 Real data analysis 1: drought intensity

The proposed method can be effective in analyzing annual functional data related to climate, which is often monitored in real-time. One annual measurement that exhibits common templates across years, along with phase variation, is drought intensity. We consider sequential Bayesian registration of trajectories of annual drought intensity near Kaweah River in California. Statistical analysis of drought can help understand historical patterns and the variability in timing of drought intensity that can further be accounted for by management and allocation of water to different communities. Importantly, drought intensity in California has been of great interest to address the state’s drought risk management and limited access to water; Californians rely on the West and East Sierra Nevada water resources due to intense drought in other areas of the state (Mann and Gleick 2015). Much attention in the literature has been devoted to developing new models for annual drought intensity. However, this task is challenging due to the time-variation and confounding of drought intensity magnitude and relative phase due to meteorologic and anthropogenic changes in the climate (Diftenbaugh et al. 2015; Son et al. 2021). In particular, annual drought intensity often exhibits a time lag as well as seasonal variation. Thus, interest not only lies in the magnitude of drought intensity, but also in the times at which drought periods begin and end during a particular year; this feature of drought intensity is captured through phase and can improve our understanding of variation in the drought lead time (Dikshit et al. 2021). While most existing statistical analyses of drought timing variation adopt multivariate data analysis approaches (Mohammed et al. 2020; Ferijal et al. 2021), we apply the proposed method to explore continuous trajectories of historical annual drought intensities

and to account for variation in timing of features in the trajectories via registration.

Despite the importance of monitoring the severity of drought, there is no universal definition of drought intensity (Mishra and Singh 2010). Many types of drought intensity indices have been developed for monitoring purposes, which are derived using a summary of covariates such as precipitation and temperature via a deterministic model. Time lag is a common phenomenon in these drought intensity indices, and some studies have focused on time lag estimation via multivariate time series methods (Mishra and Singh 2011). In this study, we extract phase variability from a historical record of a single drought intensity index. In particular, we use the proposed sequential registration approach to explore the variation in the magnitude and phase of annual drought intensity near Kaweah River in California from 1970 to 2019. We use the Standardised Precipitation-Evapotranspiration Index (SPEI), that is obtained via a deterministic equation of latitude, time of year, and precipitation and temperature across time and space (Kim and Grulke 2022; Vicente-Serrano et al. 2010). SPEI functional data, which is available at monthly intervals, is pre-processed via spline interpolation and segmentation by hydrological year, starting on October 1 and ending on September 30. We refer to this data as annual SPEI functions and denote it by $f_{1:50}$. A subset of the data, corresponding to hydrological years 2014 to 2019, is shown in Figure 5(a).

Since the structure of SPEI functions is relatively simple, with few peaks and valleys in each functional observation, we model the SRVF template as a linear combination of ten cubic B-spline basis functions with equally spaced knots. We specify the following prior hyperparameters for the template and phase components: $B = 10$, $\Sigma_c = 20I_B$, $\kappa = 5$, $\alpha_\sigma = 4$, $\beta_\sigma = 0.01$, and a uniform partition of size $M_\gamma = 10$. Given the small number of prominent features on each SPEI, a partition size of $M_\gamma = 10$ should be large enough to capture phase variation. We begin by obtaining 10,000 MCMC samples from the posterior given $f_{1:30}$, after a burn-in period of 40,000 iterations. The rest of the SPEI functions $f_{31:50}$ are assimilated sequentially, and the initial samples are updated recursively via the proposed SMC algorithm.

Features of the target posterior density, $\pi(\cdot|f_{1:50})$, are visualized in Figure 5. Drought is defined as an interval of negative SPEI, with the threshold shown as a red dashed line in panels (a) and (e). In (a), it appears that drought near Kaweah River generally occurs during the summer season, with significant variation in the timing of drought onset from year to year. Panel (c) shows weighted posterior samples of the SRVF template (transparency reflects the magnitude of the corresponding weight), given the annual SPEI functions $f_{1:50}$. This marginal posterior has small pointwise uncertainty across the domain. Panel (b) shows the weighted samples of phase for SPEI functions $f_{45:50}$, corresponding to

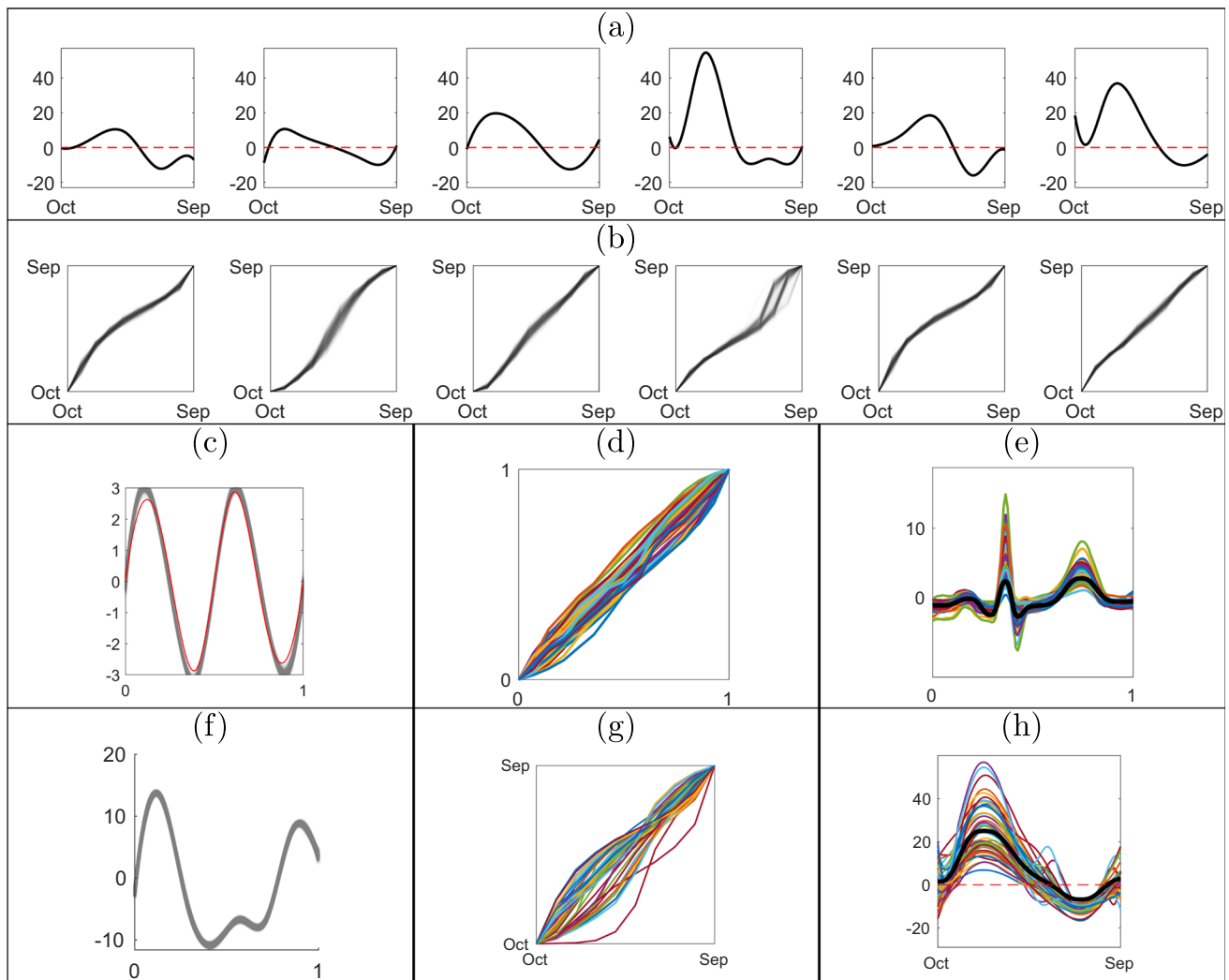


Fig. 5 Sequential Bayesian registration of SPEI functions representing drought intensity. (a) Trajectories of annual drought intensity near Kaweah River in California from 2014 to 2019 (hydrological year). Weighted posterior samples of (b) the phase components for the functions in (a), and (c) the template SRVF from the target posterior distribution given SPEI functions from 1970 to 2019. Transparency in (b) and (c) reflects magnitude of the weights. (d) Marginal posterior

means of the phase components for SPEI functions from 1970 to 2019. (e) SPEI functions (for all years) registered via the phase components in (d) are shown in colors with the marginal posterior mean of the template in the original function space in black. (f)–(g) Same as (c)–(e), but based on MCMC batch learning. The red dashed line in (a), (e) and (h) is the threshold at which a drought period begins and ends

hydrological years 2014 to 2019. The marginal posterior of the phase for year 2017 (fourth column) appears multimodal. This is due to the SPEI function in 2017 having two valleys during the drought season; either of the two valleys can be aligned to the single valley in the template.

Figure 5(d) illustrates annual phase variability of the SPEI functions $f_{1:50}$ via the marginal posterior means of the phase components. There appears to be significant phase variation throughout each hydrological year as most of the posterior marginal phase functions deviate from the identity element. The most phase variability is generally observed around the time when the peak in SPEI functions occurs; overall, there

is more phase variation from October to April than from May to September. Such estimates of the relative phases aid in the analysis of variation in drought duration as well as the timing of drought onset across years. Figure 5(e) shows the SPEI functions after registration via the marginal posterior means of the phases, which are illustrated using colors. It is evident that amplitude variability in the peak of the SPEI functions is much larger than amplitude variability in the valley; the magnitude of most severe drought (minimum of the SPEI functions) is similar across all years. Panel (e) shows the posterior mean template in the original function space with a thick black line, which effectively captures the prominent

features (shape) of the annual SPEI functions. We also visualize posterior samples given $f_{1:50}$ obtained via MCMC. Panels (f)-(h) show the marginal posterior of the template SRVF, the marginal posterior means of the phase components, and the registered functions (using the posterior means in (g)) with the marginal posterior mean of the template in the original function space in black, respectively. While the results based on our approach and MCMC-based batch learning are quite similar, there are some differences in the posterior means of the phase components. We hypothesize that this is due to issues with MCMC convergence.

While multivariate analysis methods can be used to analyze discrete drought intensity data, the proposed approach is better suited for inference on the smoothly-varying annual SPEI functions. Often, synchronization of SPEI functions, and extraction of phase variability, is done on the basis of (a few) manually identified landmarks, e.g., peaks and valleys. However, as seen in Figure 5(a), systematic selection of landmark points on SPEI functions is difficult as the number of significant features on each function can vary. On the other hand, the sequential registration framework enables full horizontal synchronization of the SPEI functions, without specification of such landmarks and while also providing uncertainty estimates.

5.5 Real data analysis 2: sea surface salinity

Next, we consider trajectories of annual sea surface salinity (SSS) near Null Island. Salinity is an important variable in understanding global climate change as it regulates the movement of currents and heat carried within currents based on water density (Durack 2015). As water evaporation and precipitation change over time, the magnitude of SSS fluctuates annually according to seasonal variation and the movement of currents. There are recent studies focusing on using phase variation in SSS to understand ocean dynamics, e.g., on the lag between fluctuations of salinity and precipitation, a major factor that affects ocean salinity (Bingham et al. 2012). Bingham et al. (2021) compare two sets of SSS observations, which were recorded using different equipment, via estimates of phase variability. Thus, estimation and assessment of phase variation in annual SSS measurements is an important statistical task, and can be subsequently used to understand ocean dynamics and to explore relationships between SSS and other oceanic factors. In previous studies, Bingham et al. (2012, 2021) approximated seasonal SSS variability through a one-dimensional measurement of phase corresponding to the time at which maximum SSS magnitude is observed. This, again, is akin to landmark-based registration of SSS functions, with the point of maximum SSS magnitude on each function serving as a single landmark. While intuitive and simple, this approach fails to register prominent features of the SSS functions not captured by the single landmark.

Thus, to estimate more flexible and informative phase components of SSS functions, we apply the proposed sequential Bayesian registration approach.

A climate pattern over the Pacific Ocean that is known to influence global oceanographic variables is El Niño-Southern Oscillation (ENSO). ENSO is a phenomenon that fluctuates among three states: warm state or El Niño, neutral state, and cool state or La Niña; the criteria used to determine which of these three states is occurring in a given year are sea surface temperature, wind speed, surface pressure, and related measurements. This climate pattern has a close relationship with SSS, e.g., SSS tends to decrease (increase) near the equatorial Pacific during El Niño (La Niña) (Zhu et al. 2014; Hackert et al. 2020). In this study, we explore amplitude and phase patterns in SSS functions for the three ENSO states. To do this, we first pre-process monthly EN4 observations of SSS via spline interpolation and segmentation across years 2000 to 2017 (British Crown Copyright Met Office 2021; Good et al. 2013). We further categorize each year to one of the three ENSO states, and apply the proposed sequential approach to separately register SSS functions for each state. In particular, there are four SSS functions measured during El Niño, $f_{1:4}^E$, ten SSS functions measured during the neutral state, $f_{1:10}^N$, and four SSS functions measured during La Niña, $f_{1:4}^L$; the data is shown in Figure 6(a). Note that the SSS functions within each ENSO state tend to share similar amplitude features with clear phase variation. The majority of the El Niño SSS functions have a single distinct peak; the neutral and La Niña SSS functions tend to have more amplitude features.

As in the drought intensity study, we model the SRVF template using a linear combination of ten cubic B-splines basis functions, which provides considerable flexibility for template estimation. The other prior hyperparameters for the template and phase components are $\Sigma_c = 20I_B$, $\kappa = 5$, $\alpha_\sigma = 4$, $\beta_\sigma = 0.01$, and a uniform partition of size $M_\gamma = 10$ (a partition of this size provides sufficient flexibility for phase estimation). For each ENSO state, we initialize the sampler using 10,000 MCMC samples from the posterior given the first three functions after a burn-in period of 40,000 iterations. The rest of the SSS functions for each ENSO state are then added to the data sequentially, and the initial samples are updated recursively using the proposed SMC algorithm.

Estimated relative phase and amplitude for the data in Figure 6(a), based on posterior densities $\pi(\cdot | f_{1:4}^E)$, $\pi(\cdot | f_{1:10}^N)$ and $\pi(\cdot | f_{1:4}^L)$, are shown in Figure 6(b) and (c), respectively. Panels (b) and (c) show the marginal phase posterior means, and registered SSS functions, respectively. Phase variability within each ENSO state is fairly small and registered functions appear to be horizontally synchronized as desired.

Figure 7 shows the marginal posterior of the template function for each ENSO state: El Niño in red, neutral in blue,

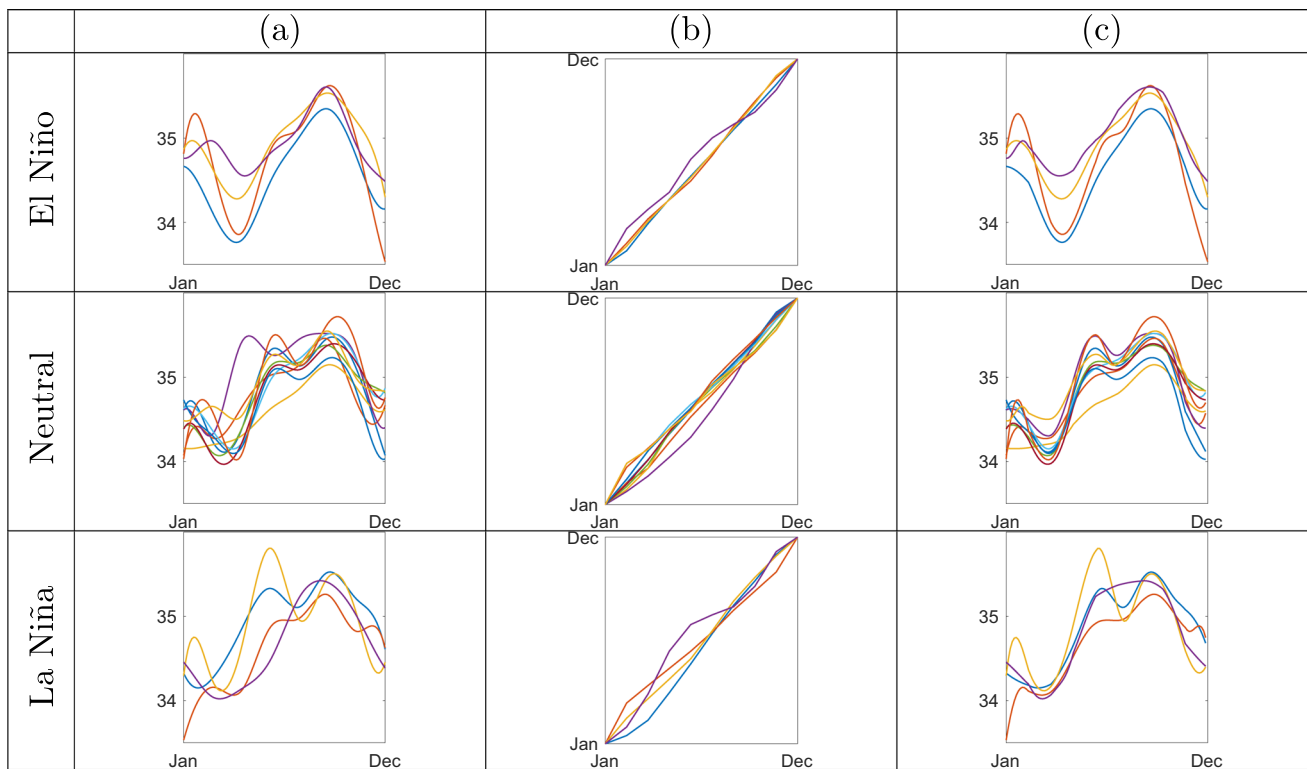


Fig. 6 Sequential Bayesian registration of SSS functions for three ENSO states. (a) Trajectories of annual SSS near Null Island from 2000 to 2017, plotted by ENSO state. (b) Marginal posterior means of phase components for the SSS functions in (a). (c) SSS functions registered using the phases in (b)

and La Niña in yellow. In (a), the marginal posterior mean of the SRVF template is shown as a solid curve with a band corresponding to ± 2 pointwise marginal posterior standard deviations. The marginal posterior means of the template function in the original function space are illustrated in (b). The marginal posterior means of the template function clearly capture the prominent features of the SSS functions in each of the three ENSO states; the El Niño template has a single valley followed by a fairly sharp peak, while the neutral and La Niña templates have shallower valleys followed by a peak that is spread over a longer time period. In particular, the estimated template functions for the neutral and La Niña states appear to be more similar in shape than the El Niño template. These shape differences can be quantified by computing pairwise \mathbb{L}^2 distances between (rescaled to have norm one) SRVFs of the marginal posterior means of the template function for each state. The distances between the El Niño and La Niña/neutral states are both relatively large: 0.7662/0.6867. On the other hand, the distance between the La Niña and neutral states is much smaller: 0.2978. The presented amplitude/phase estimation results for annual SSS functional data, based on the proposed sequential registration framework, can provide insights into the association between amplitude and phase variation in SSS functions across the three ENSO states.

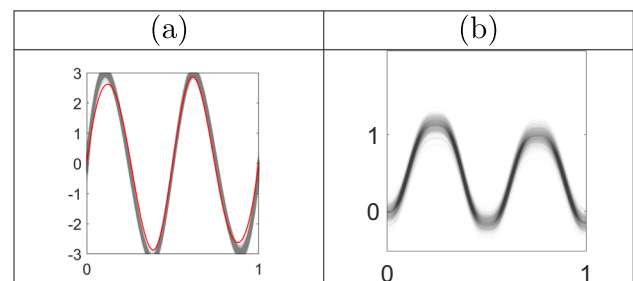


Fig. 7 Visualization of the marginal posterior for the template for each ENSO state: El Niño in red, neutral in blue, La Niña in yellow. (a) Posterior mean templates in the SRVF space are shown as solid curves. The band around each posterior mean represents the posterior uncertainty: ± 2 pointwise marginal posterior standard deviations. (b) Marginal posterior mean for the template for each ENSO state in the original function space

5.6 Real data analysis 3: segmented PQRST complexes

The electrocardiogram (ECG) is routinely used to assess heart function and diagnose various medical conditions, e.g., myocardial infarction. The data recorded via ECG is a long signal composed of a periodic sequence of a pattern known as the PQRST complex: the P wave corresponds to the first

small peak, the QRS wave is composed of a sharp valley followed by a sharp peak followed by another sharp valley, and the T wave represents the last high peak. Prior to statistical analysis, it is beneficial to segment the long ECG signal into its PQRST complexes (Kurtek et al. 2013). Our focus here is not on the segmentation problem, but rather on registration of PQRST complexes as the long signal is recorded and segmented sequentially. While there is natural variability among the PQRST complexes along an ECG signal, abnormalities in estimates of the underlying template and variation in timing of PQRST patterns (phase) with respect to that template are beneficial for the aforementioned diagnostic purposes. For example, duration of the QT interval is useful in drug development and approval (Zhou and Sedransk 2009).

We consider 50 PQRST complexes, denoted by $f_{1:50}$, segmented from a long ECG signal (Kurtek et al. 2013); six functions from this set are shown in Figure 8(a). We initialize the sequential registration algorithm using 10,000 MCMC samples based on the data $f_{1:30}$ after burn-in. PQRST complexes tend to have “sharper” features than the data we encountered in the previous two examples. Thus, we model the template function using a linear combination of 13 B-spline basis functions ($B = 13$); we set $\Sigma_c = 20I_B$ as previously. The concentration parameter in the prior over increments of the phase components is $\kappa = 5$ and the size of the uniform partition is $M_\gamma = 15$. Finally, the hyperparameters for the error variance are $\alpha_\sigma = 10$ and $\beta_\sigma = 0.01$.

Results based on the last posterior distribution in the sequence, $\pi(\cdot \mid f_{1:50})$, are shown in Figure 8. Panel (b) shows the weighted samples of the phase components for the functions in (a), and panel (c) visualizes the weighted samples of the template SRVF with transparency reflecting the magnitude of weights. The marginal posterior means of the phase components are shown in panel (d). In general, they capture the relative acceleration or delay of the corresponding PQRST complex with respect to the template. Pointwise uncertainty for each phase component during the QRS time interval (sharp valley followed by a sharp peak followed by another sharp valley) tends to be very small compared to other time points along the domain. This is intuitive, as the QRS shape features in a PQRST complex are very prominent resulting in very small uncertainty in their timing relative to the template. The registered data, obtained by applying the marginal posterior mean phase components in (d) to the corresponding PQRST complexes, is shown in panel (e). The proposed method achieves very good horizontal synchronization of all features of the PQRST complexes. The marginal posterior mean of the template function in the original function space is illustrated with a thick black curve in panel (e). The estimated template function contains all features of the PQRST complex and is representative of the shape in the data. The uncertainty in the shape of the template is further shown using principal component analysis in panel (f): the

black function is the posterior mean template; the blue and red functions show the positive and negative direction along the first principal component with $\alpha = 3$, respectively, as described in Section 5.1. Here, we note that most variation occurs in the magnitude of the RST features with very little variation in the magnitude of the PQ features.

6 Discussion and future work

We propose a novel sequential Bayesian registration method for functional data that addresses two common challenges: confounded sources of variation across functions, and availability of an increasing number of functional observations over time. In particular, we propose a Bayesian registration model and an SMC algorithm to update inference when a new function is observed, exploiting an approximate weight updating approach for MH kernels without a closed form. As illustrated in the simulation studies and real data examples, the proposed sequential learning algorithm is computationally more efficient, and more accurate in terms of posterior estimates, compared to the batch MCMC algorithm. It efficiently explores multimodal posterior distributions, which often pose a challenge for many MCMC techniques. Application of this framework to real data studies improves our understanding of the underlying structure of the data, and provides quantitative evidence to study associated research problems.

An important consideration is that the proposed method leads to approximate posterior inference due to the need to circumvent computation of the MH kernels in the SMC algorithm. The quality of the approximation depends on the assumption that marginals of adjacent posterior distributions are similar. In general, this holds when a new function is assimilated with a relatively large number of functions that are already registered. If the number of previously registered functions is small (< 5), and a function with a very different shape arrives, the difference in the marginal posteriors may lead to non-representative weights. In such cases, we suggest adding annealing steps between adjacent target distributions. With this adjustment, the proposed algorithm is robust to outlying functions that have potentially different shape relative to other functions already in the sample.

Particle degeneracy often arises in SMC sampling for a sequence of distributions with increasing state space dimension, and can lead to unreliable posterior estimates. We partly resolve this issue by initializing new components of each particle such that the updated particles lie in a region of high posterior probability. When phase components are added to the state space as new data arrives, this initialization is performed by drawing from a distribution centered at a phase function obtained by aligning the new functional observation to existing template particles. Furthermore, we monitor

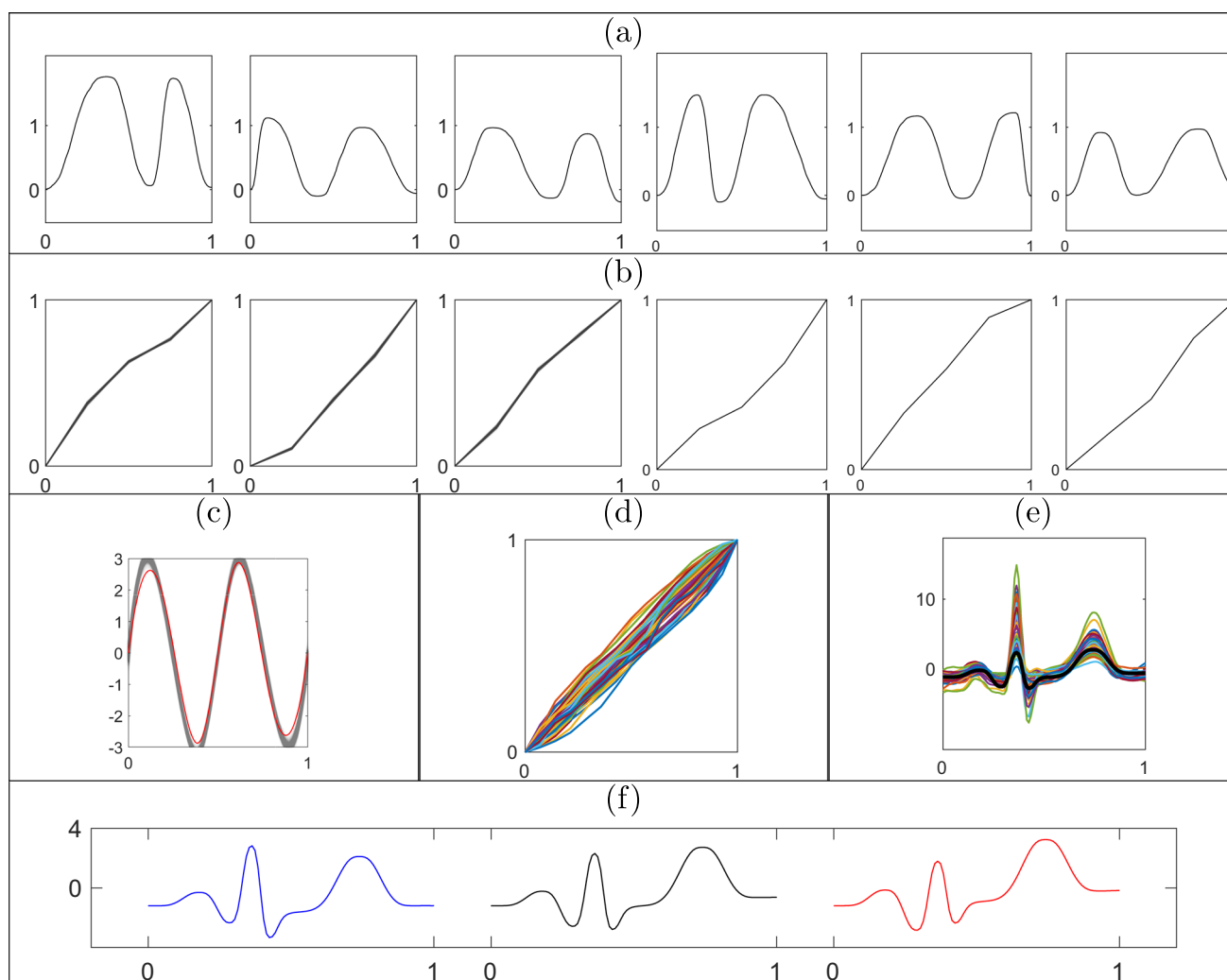


Fig. 8 Sequential Bayesian registration of PQRST complexes extracted from an ECG signal. (a) Subset of 50 PQRST complexes. (b) Weighted posterior samples of relative phases for the functions in (a). (c) Weighted posterior samples of the template SRVF. Transparency in (b) and (c) reflects the magnitude of weights. (d) Posterior mean phases for the 50

functions. (e) Registered PQRST complexes (solid colors) using phase components in (d) and posterior mean template in the original function space (thick black line). (f) Posterior mean template (black), and negative (red) and positive (blue) directions along the first principal component of the template in the original function space

ESS as small values of ESS can indicate evidence of particle degeneracy. Degeneracy may be remedied by duplicating existing particles, thereby increasing the number of weighted samples. This was not necessary in our studies, as the ESS was large and no significant decrease in ESS was observed as we added more functions to the data.

The proposed Bayesian registration model may be modified depending on the application of interest. The key assumption that makes the model computationally tractable is that phase components are piecewise linear functions with Dirichlet increments. This reduces the state space dimension relative to similar models in the literature and enables faster posterior inference. Another consideration is that, for the sake of isometry and simple derivations, we build the

Bayesian hierarchical model on the SRVF space. This limits our model to functions for which derivatives exist almost everywhere. The proposed method is therefore applicable to various real data studies where interest lies in smooth evolutions of a variable.

In this work, we assume that the functions are independent given the template and phase components. Such an assumption limits forecasting ability beyond the use of the posterior distribution based on historical data for prediction. Thus, a natural extension of the proposed framework is to account for temporal dependence between sequentially observed functions, which is a reasonable assumption for applications such as climate monitoring. This requires a re-formulation of the Bayesian registration model as a functional time series model

(Bosq 2000; Kokoszka 2012). Another natural extension is to incorporate time-varying covariates via a functional linear model. We plan to explore these two directions in future work.

The presented Bayesian registration model can be extended to the case of multivariate functional data, i.e., functional data with multiple dependent or independent components. Under the assumption of zero cross-component phase variation within each observation, the main modification to the presented model requires the definition of a prior distribution over the multiple components of the template. However, in the presence of cross-component and cross-observation phase variation, one would have to construct more complicated prior models for the phase component, and ensure that cross-component and cross-observation phase are identifiable. This extension would be especially relevant in the context of ECG signals. While in this work we used segmented PQRST complexes from an ECG signal recorded by a single lead, the standard ECG has 12 leads, each recording the heart’s electrical activity based on a different placement of an electrode. Harnessing information recorded across all 12 leads could provide more information about the heart’s function for diagnostic purposes.

Acknowledgements This project was supported in part by NIH R37-CA214955, NSF CCF-1740761, NSF CCF-1839252, NSF DMS-2015226 and NSF DMS-2413747 (to SK), and NASA-80NSSC18K1322 (to OC). The authors thank Dr. Frederick Bingham and Susannah Brodnitz for their valuable comments on the sea surface salinity case study and extracting the monthly EN4 data. The authors also thank Dr. Nancy Grulke for her thoughtful comments on the drought intensity case study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics. Springer, New York, NY (2006)

Lajos, H., Kokoszka, P.: Inference for Functional Data with Applications. Springer Series in Statistics. Springer, New York, NY (2012)

Ramsay, J.O., Dalzell, C.J.: Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B* **53**(3), 539–561 (1991)

Ramsay, J., Silverman, B.W.: Functional Data Analysis. Springer Series in Statistics. Springer, New York, NY (2005)

Hoi, S.C.H., Sahoo, D., Lu, J., Zhao, P.: Online learning: A comprehensive survey. *Neurocomputing* **459**, 249–289 (2021)

Andrieu, C., Freitas, N., Doucet, A.: Sequential MCMC for Bayesian model selection. In: *IEEE Signal Processing Workshop on Higher-Order Statistics*, pp. 130–134 (1999)

Storvik, G.: Particle filters for state space models with the presence of static parameters. *IEEE Trans. on Signal Processing* **50**, 281–289 (2002)

Lopes, H., Tsay, R.: Particle filters and Bayesian inference in financial econometrics. *Journal of Forecasting* **30**, 168–209 (2011)

Schweizer, N.: Non-asymptotic error bounds for sequential MCMC methods in multimodal settings. [arXiv: 1205.6733](https://arxiv.org/abs/1205.6733) (2012)

Paulin, D., Jasra, A., Thiery, A.H.: Error bounds for sequential Monte Carlo samplers for multimodal distributions. *Bernoulli* **25**(1), 310–340 (2019)

Dai, C., Heng, J., Jacob, P.E., Whiteley, N.: An invitation to sequential Monte Carlo samplers. *Journal of the American Statistical Association* **117**(539), 1587–1600 (2022)

Gilks, W.R., Berzuini, C.: Following a moving target—Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society, Series B* **63**(1), 127–146 (2001)

Chopin, N.: A sequential particle filter method for static models. *Biometrika* **89**(3), 539–552 (2002)

Del Moral, P., Doucet, A., Jasra, A.: Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society, Series B* **68**, 411–436 (2006)

Kantas, N., Beskos, A., Jasra, A.: Sequential Monte Carlo methods for high-dimensional inverse problems: A case study for the Navier–Stokes equations. *SIAM/ASA Journal on Uncertainty Quantification* **2** (2013)

Beskos, A., Jasra, A., Muzaffer, E.A., Stuart, A.M.: Sequential Monte Carlo methods for Bayesian elliptic inverse problems. *Statistics and Computing* **25**, 727–737 (2015)

Liu, J., West, M.A.: Combined parameter and state estimation in simulation-based filtering. In: *Sequential Monte Carlo Methods in Practice* (2001)

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

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

Tang, R., Müller, H.: Pairwise curve synchronization for functional data. *Biometrika* **95**(4), 875–889 (2008)

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., Marron, J.S.: Registration of functional data using Fisher-Rao metric. [arXiv: 1103.3817v2](https://arxiv.org/abs/1103.3817v2) (2011)

Kneip, A., Li, X., MacGibbon, K.B., Ramsay, J.O.: Curve registration by local regression. *The Canadian Journal of Statistics* **28**(1), 19–29 (2000)

Telesca, D., Inoue, L.Y.T.: Bayesian hierarchical curve registration. *Journal of the American Statistical Association* **103**(481), 328–339 (2008)

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

Horton, W.Z., Page, G.L., Reese, C.S., Lepley, L.K., White, M.: Template priors in Bayesian curve registration. *Technometrics* **63**(4), 487–499 (2021)

Cheng, W., Dryden, I., Huang, X.: Bayesian registration of functions and curves. *Bayesian Analysis* **11**(2), 447–475 (2016)

Bharath, K., Kurtek, S.: Distribution on warp maps for alignment of open and closed curves. *Journal of the American Statistical Association* **115**(531), 1378–1392 (2020)

Matuk, J., Bharath, K., Chkrebti, O., Kurtek, S.: Bayesian framework for simultaneous registration and estimation of noisy, sparse, and

- fragmented functional data. *Journal of the American Statistical Association* **117**(540), 1964–1980 (2021)
- Srivastava, A., Klassen, E.P.: *Functional and Shape Data Analysis*. Springer Series in Statistics. Springer, New York, NY (2016)
- Lang, S., Brezger, A.: Bayesian P-splines. *Journal of Computational and Graphical Statistics* **13**(1), 183–212 (2004)
- Gordon, N.J., Salmond, D.J., Smith, A.F.M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings F (Radar and Signal Processing)* **140**(2), 107–113 (1993)
- Liu, J.S., Chen, R.: Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association* **93**(443), 1032–1044 (1998)
- Bertsekas, D.P.: *Dynamic Programming and Optimal Control*, 2nd edn. Athena Scientific, Nashua, NH (2000)
- Zang, X.: *Clustering functional data based on amplitude-phase separation*. PhD thesis, The Ohio State University (2021)
- Mann, M.E., Gleick, P.H.: Climate change and California drought in the 21st century. *Proceedings of the National Academy of Sciences* **112**(13), 3858–3859 (2015)
- Diffenbaugh, N.S., Swain, D.L., Touma, D.: Anthropogenic warming has increased drought risk in California. *Proceedings of the National Academy of Sciences* **112**(13), 3931–3936 (2015)
- Son, B., Im, J., Park, S., Lee, J.: Satellite-based drought forecasting: Research trends, challenges, and future directions. *Korean Journal of Remote Sensing* **37**, 815–831 (2021)
- Dikshit, A., Pradhan, B., Alamri, A.M.: Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model. *Science of the Total Environment* **755**, 142638 (2021)
- Mohammed, S., Alsafadi, K., Al-Awadhi, T., Sherief, Y., Harsanyie, E., El Kenawy, A.: Space and time variability of meteorological drought in Syria. *Acta Geophysica Polonica* **68**(6), 1877–1898 (2020)
- Ferijal, T., Batelaan, O., Shanafeld, M.: Spatial and temporal variation in rainy season droughts in the Indonesian Maritime Continent. *Journal of Hydrology* **603**, 126999 (2021)
- Mishra, A.K., Singh, V.P.: A review of drought concepts. *Journal of Hydrology* **391**(1), 202–216 (2010)
- Mishra, A.K., Singh, V.P.: Drought modeling - a review. *Journal of Hydrology* **403**(1), 157–175 (2011)
- Kim, Y., Grulke, N.E.: 80-year meteorological record and drought indices for Sequoia National Park and Sequoia National Monument. CA, Forest Service Research Data Archive (2022)
- Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I.: A multi-scalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate* **23**(7), 1696–1718 (2010)
- Durack, P.J.: Ocean salinity and the global water cycle. *Oceanography* **28**(1), 20–31 (2015)
- Bingham, F.M., Foltz, G.R., McPhaden, M.J.: Characteristics of the seasonal cycle of surface layer salinity in the global ocean. *Ocean Science* **8**(5), 915–929 (2012)
- Bingham, F.M., Brodnitz, S., Yu, L.: Sea surface salinity seasonal variability in the tropics from satellites, gridded in situ products and mooring observations. *Remote Sensing* **13**(1) (2021)
- Zhu, J., Huang, B., Zhang, R., Hu, Z., Kumar, A., Balmaseda, M., Marx, L., Kinter, J.: Salinity anomaly as a trigger for ENSO events. *Scientific Reports* **4**, 6821 (2014) <https://doi.org/10.1038/srep06821>
- Hackert, E., Kovach, R.M., Molod, A., Vernieres, G., Borovikov, A., Marshak, J., Chang, Y.: Satellite sea surface salinity observations impact on El Niño/southern oscillation predictions: Case studies from the NASA GEOS seasonal forecast system. *Journal of Geophysical Research: Oceans* **125**(4) (2020)
- British Crown Copyright Met Office: EN.4.2.2 data (2021). <https://www.metoffice.gov.uk/hadobs/en4/>
- Good, S.A., Martin, M.J., Rayner, N.A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans* **118**(12), 6704–6716 (2013)
- Kurtek, S., Wu, W., Christensen, G.E., Srivastava, A.: Segmentation, alignment and statistical analysis of biosignals with application to disease classification. *Journal of Applied Statistics* **40**(6), 1270–1288 (2013)
- Zhou, Y., Sedransk, N.: Functional data analytic approach of modeling ECG T-wave shape to measure cardiovascular behavior. *Annals of Applied Statistics* **3**(4), 1382–1402 (2009)
- Bosq, D.: *Linear Processes in Function Spaces: Theory and Applications*. Lecture Notes in Statistics. Springer, New York, NY (2000)
- Kokoszka, P.: Dependent functional data. *Probability and Statistics* **2012** (2012)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.