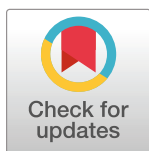


RESEARCH ARTICLE

Bayesian function registration with random truncation

Yi Lu¹, Radu Herbei², Sebastian Kurtek²***1** Mathematics and Computer Science Department, Drew University, Madison, New Jersey, United States of America, **2** Department of Statistics, The Ohio State University, Columbus, Ohio, United States of America

* These authors contributed equally to this work.

* kurtek.1@stat.osu.edu

Abstract

In this work, we develop a new set of Bayesian models to perform registration of real-valued functions. A Gaussian process prior is assigned to the parameter space of time warping functions, and a Markov chain Monte Carlo (MCMC) algorithm is utilized to explore the posterior distribution. While the proposed model can be defined on the infinite-dimensional function space in theory, dimension reduction is needed in practice because one cannot store an infinite-dimensional function on the computer. Existing Bayesian models often rely on some pre-specified, fixed truncation rule to achieve dimension reduction, either by fixing the grid size or the number of basis functions used to represent a functional object. In comparison, the new models in this paper randomize the truncation rule. Benefits of the new models include the ability to make inference on the smoothness of the functional parameters, a data-informative feature of the truncation rule, and the flexibility to control the amount of shape-alteration in the registration process. For instance, using both simulated and real data, we show that when the observed functions exhibit more local features, the posterior distribution on the warping functions automatically concentrates on a larger number of basis functions. Supporting materials including code and data to perform registration and reproduce some of the results presented herein are available online.

OPEN ACCESS

Citation: Lu Y, Herbei R, Kurtek S (2023) Bayesian function registration with random truncation. PLoS ONE 18(7): e0287734. <https://doi.org/10.1371/journal.pone.0287734>

Editor: Junyuan Yang, Shanxi University, CHINA

Received: December 16, 2022

Accepted: June 12, 2023

Published: July 7, 2023

Copyright: © 2023 Lu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting information](#) files.

Funding: SK's work was in part supported by the National Science Foundation (NSF) projects CCF-1740761, CCF-1839252 and DMS-2015226, and the National Institutes of Health (NIH) project R37-CA214955. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Advances in data collection technology have made functional data prevalent in various applied domains including biology, biometrics, medicine, computer vision, bioinformatics, and many others. This, in turn, has prompted rapid development of functional data analysis (FDA) methods for estimation, alignment, summarization, and statistical modeling (and inference) for such data. In this work, we specifically focus on the problem of Bayesian model-based alignment of two or more functions, termed pairwise and multiple registration, respectively. In particular, we elucidate the challenges arising from nonlinearity and infinite-dimensionality of the representation spaces on which observation and prior models must be defined.

We consider the task of registration of real-valued functions defined on a subinterval of the real line. The goal of registration is to separate two sources of variability in functional data termed amplitude (y -axis variation) and phase (x -axis variation or time warping), and our aim

is to temporally align (or warp) a set of functions, such that the amplitude variation in the observed data is comparable (i.e., amplitude features like local extrema occur at the same time along the x -axis, across all functions). The phase variation is then captured by a set of time warping functions that achieve the alignment. Formal definitions of amplitude, phase and the registration problem that are considered in this work are provided in subsequent sections.

Real data examples of the types of functions we consider are plotted in Fig 1. To motivate the problem of function registration, we consider the widely-used Berkeley growth study dataset [1]. The data is comprised of height measurements for boys and girls recorded from age 1 to age 18. To discover patterns of growth such as growth spurts, it is often preferable to analyze growth rate functions, i.e., the time derivative of the height measurement functions, since periods of fast/slow growth result in local extrema. The growth rate functions for 54 girls are displayed in the third panel, top row of Fig 1. It is clear that the functions have very similar shapes and exhibit one peak near the middle of the domain, corresponding to the pubertal growth spurt. However, the pubertal growth spurt does not occur at the same time across all growth rate functions since different children go through puberty at different times. Thus, it becomes necessary to register the growth rate functions prior to statistical analysis such that the amplitude variability (magnitude of pubertal growth spurts) and phase variability (timing of pubertal growth spurts) are separated.

To enhance statistical analysis, registration of functional data has been utilized in a wide range of applications including biomechanical data [2–4], handwriting samples [5, 6], gene expression and proteomics data [7, 8], neural spike trains [9], and gait data [10, 11]. Traditionally, function registration is formulated as an optimization problem under a specific optimality criterion, preferably a metric. We refer readers to standard FDA textbooks (e.g., [12, 13]) for an overview of the many approaches that have been proposed. Recently, model-based or probabilistic frameworks have become popular in formulating the registration problem. Specifically, the Bayesian modeling paradigm provides considerable flexibility as it allows the user to

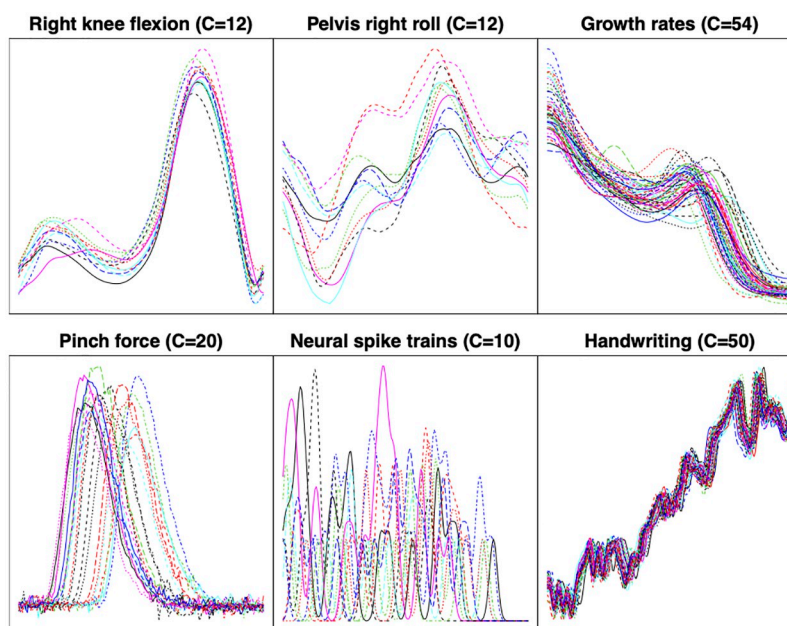


Fig 1. Examples of the types of functions we consider. The value C represents the number of functions in each dataset.

<https://doi.org/10.1371/journal.pone.0287734.g001>

specify a prior distribution over the phase parameter space. Additionally, it yields a principled approach for a more comprehensive exploration of the phase parameter space and provides quantified uncertainty measures via the posterior distribution.

The literature includes multiple different Bayesian formulations of function registration. Telesca and Inoue [7] use the B-spline basis to model functional parameters; Claeskens et al. [14] decompose time warping functions into so-called warping component functions or *warp-plets*. Cheng et al. [15], Bharath and Kurtek [16], and Matuk et al. [17] use Dirichlet priors for the time (increments of) a warping function. Finally, Earls and Hooker [18], Kurtek [19], and Lu et al. [11] use Gaussian process priors on a transformed warping parameter space. We refer the reader to Matuk et al. [20] for a general overview of Bayesian registration methods. As evident, the main differences among the aforementioned methods are in the specification of the observation model, the prior model on phase, and the algorithms used for parameter space exploration; we discuss benefits and drawbacks of the different choices later as we introduce the proposed framework. Some of the methods (e.g., [16]) are additionally able to incorporate information on landmarks (predetermined, user-defined points of interest) into the registration problem.

This work extends and improves the Bayesian models proposed in [11] to enable full exploration of the functional parameter space. In [11], the authors specify a Gaussian process prior over the infinite-dimensional phase parameter space and represent time warping functions using a sequence of basis functions. However, at the implementation stage, the dimension of the parameter space is reduced by choosing a *fixed* number of basis functions. Thus, the resulting model is not truly infinite-dimensional and is able to explore only a small subset of the underlying parameter space. In practice, the main disadvantage of such a formulation is that the dimension reduction is performed a priori and is thus not informed by the data. Furthermore, it is generally not obvious how many basis functions are needed to achieve satisfactory registration results.

To remedy these issues, instead of using a fixed truncation, we allow the truncation to be *random*. This is done by randomizing either (i) the *number* of basis functions or (ii) *which* basis functions are used. We incorporate this random truncation as a separate parameter, leading to nonparametric, infinite-dimensional models in the sense that the prior distributions are assigned on the entire functional phase parameter space. The proposed models with random truncation have three advantages. First, the level of smoothness of the functional phase parameter is informed by the data, thus avoiding potential mis-specifications of the number of basis functions. As we will show, under-specification of the number of basis functions can lead to poor registration results. Second, one can flexibly incorporate prior beliefs or desired constraints on the shape of the functional phase parameter. For instance, how much shape-alteration occurs in the observed functions can be controlled by the prior on the number of basis functions. Third, our model allows one to make inference on the random truncation parameter, which can provide additional information about the shapes of the functions in the data. For instance, the posterior tends to keep a larger number of basis functions for the phase parameter when the observed functions exhibit a lot of local features that must be registered. Following ideas from [11, 21, 22], we develop algorithms that allow efficient sampling from the posterior distribution of both the functional parameter and the random truncation parameter. A key designing principle for our model is to treat the phase parameter space as infinite-dimensional and to allow the data to dictate the amount of dimension reduction that is needed.

The rest of the paper is organized as follows. We first introduce the statistical problem of function registration, focusing on relevant function spaces, in Section Problem Formulation and Function Spaces of Interest. The proposed Bayesian registration models are formally

specified in Sections Pairwise Bayesian Registration Model with Random Basis Truncation and Multiple Function Bayesian Registration Model with Random Basis Truncation. In Sections Simulation Study and Applications, we demonstrate the proposed method on a pairwise simulation study and several real datasets.

Problem formulation and function spaces of interest

We formulate the task of function registration as a statistical problem by defining (i) the data and the observation space, and (ii) the parameter(s) and their corresponding representation spaces. As will be seen, it is essential to transform both the observation space and the parameter space. The transformations we adopt here are developed in [23, 24] in the context of function registration and have been utilized for registration models in a host of recent manuscripts, including [11, 17–19, 25, 26]. A comprehensive discussion of these transformations can be found in [13]. Here, for brevity, we will only introduce the relevant notation and briefly state the transformations. We refer the reader to the aforementioned references for more details.

We start with the simpler case of *pairwise* registration, where two real valued functions, f_1 and f_2 , are observed. Without loss of generality, we assume that the domain on which the functions are observed is $[0, 1]$. In this scenario, these two functions are regarded as *data* with the corresponding *observation space* $\mathcal{F} = \{f : [0, 1] \mapsto \mathbb{R} \mid f \text{ is absolutely continuous}\}$. Suppose our goal is to register f_2 to f_1 . This is achieved by finding a *warping function* γ such that $f_2 \circ \gamma$ and f_1 are aligned. The role of γ is to warp the domain of f_2 so that the amplitude variation of f_2 is retained, but its phase variation is altered (ideally to match that of f_1). The amount of alteration, which quantifies the difference in phase variation between f_1 and f_2 , is captured by γ . The warping function γ is regarded as the *parameter* with the corresponding *parameter space* $\Gamma = \{\gamma : [0, 1] \mapsto [0, 1] \mid \gamma(0) = 0, \gamma(1) = 1, 0 < \gamma' < \infty\}$. In the next two paragraphs, we separately describe the transformations carried out on the (1) observation (data) space, and (2) parameter space. More details on these transformations can be found in the [S1 File](#).

Observation (data) space. For $f \in \mathcal{F}$, we use the *square-root velocity* transformation

$$Q(f)(t) = \text{sign}(f'(t))\sqrt{|f'(t)|} \equiv q(t), \quad (1)$$

where f' is the derivative of f . The resulting function, denoted by q for simplicity and referred to as the square-root velocity function (SRVF), is an element of the transformed observation space $\mathcal{Q}$, which is a subset of $\mathbb{L}^2([0, 1])$ [23]. The mapping Q is bijective up to a translation and $f \in \mathcal{F}$ can be recovered from $q \in \mathcal{Q}$ using $Q^{-1}(q)(t) = f(0) + \int_0^t q(s)|q(s)| \, ds$. The SRVF of a time warped function, $f \circ \gamma \in \mathcal{F}$, is given by

$$Q(f \circ \gamma)(t) = Q(f)(\gamma(t))\sqrt{\gamma'(t)} \equiv (q, \gamma)(t). \quad (2)$$

Note that this is not the same as function composition $q \circ \gamma$, because of the additional term $\sqrt{\gamma'}$; for brevity, we denote this quantity by (q, γ) .

Parameter space. For $\gamma \in \Gamma$, we apply two transformations:

$$Q(\gamma)(t) = \text{sign}(\gamma'(t))\sqrt{|\gamma'(t)|} = \sqrt{\gamma'(t)} \equiv \psi(t); \quad (3)$$

$$\exp_1^{-1}(\psi) = \frac{\theta}{\sin(\theta)}(\psi - \cos(\theta)) \equiv g, \text{ where } \theta = \cos^{-1}\left(\int_0^1 \psi(t)dt\right). \quad (4)$$

The first transformation is the square-root velocity transformation in [Eq \(1\)](#) (note that $\gamma'(t) > 0 \forall t$). The second transformation is the inverse exponential map for a unit sphere that allows us to linearize the space Ψ , which is a transformed representation space of the warping

functions. The resulting function, denoted by g , belongs to a subset of a *linear space*, defined as $A \equiv \{g \in T_1(\Psi) | \exp_1(g) > 0\}$ (the notation $T_1(\Psi)$ refers to the tangent space of Ψ at the function 1; see [S1 File](#) for details). We can transform g back to γ via

$$\gamma(t) = Q^{-1}(\psi)(t) = \int_0^t \psi^2(s) ds = \int_0^t \exp_1^2(g)(s) ds,$$

where $\exp_1(g) = \cos(\|g\|) + \frac{\sin(\|g\|)}{\|g\|} g$ ($\|\cdot\|$ is the $\mathbb{L}^2$ norm). The warping of an SRVF, (q, γ) , can now be written in terms of the function g , which lies in a linear space, via

$$(q, \gamma)(t) = q(\gamma(t)) \sqrt{\gamma'(t)} = q\left(\int_0^t \exp_1^2(g)(s) ds\right) \exp_1(g)(t). \quad (5)$$

In summary, our approach is to perform statistical inference on the parameter $g \in A$, using the (SRVFs of) the observed data $q_1, q_2 \in \mathcal{Q}$. Generalization to multiple function registration is straightforward. Suppose we observe $C > 2$ functions, denoted by $f_1, \dots, f_C$. The goal is to register them simultaneously to a *template* function, f^* . In some applications, we can pre-specify a *known* function as the template (common choices include one of the observed functions or their point-wise mean). Alternatively, we can treat f^* as another *unknown* parameter to be estimated. Registration is achieved via estimation of the warping functions, $\gamma_i, i = 1, \dots, C$, corresponding to each observed function. After applying the same transformations, we treat $q_1, \dots, q_C \in \mathcal{Q}$ as data and $g_1, \dots, g_C \in A$ (the warping functions) and $Q(f^*) = q^* \in \mathcal{Q}$ (the template function) as parameters.

Pairwise Bayesian registration model with random basis truncation

In the case of pairwise registration, the data consists of two functions f_1 and f_2 , represented via their SRVFs q_1 and q_2 , observed on a finite grid of size N denoted by $[t] = \{t_1, \dots, t_N\}$. We use the notation $[t]$ to denote discretization of the domain $[0, 1]$ throughout the rest of the paper. Thus, $f([t])$ denotes evaluations of the function f at the domain points $[t]$; similarly, $\int_0^{[t]} h(s) ds$ denotes the N -dimensional vector $\left(\int_0^{t_1} h(s) ds, \dots, \int_0^{t_N} h(s) ds\right)$. We model the difference between $q_1([t])$ and $(q_2, \gamma)([t])$ by a zero-mean N -dimensional Gaussian distribution. The main parameter of interest is the warping function $\gamma \in \Gamma$ represented via $g \in A$. At the implementation stage, dimension reduction is necessary, and this is achieved by an auxiliary variable T . Specifically, we first represent g by an infinite sum using basis functions. Then, we use the random variable T to truncate the infinite sum to a finite sum, which can be evaluated on a computer. The *truncated* version of g will be henceforth denoted by $\tilde{g}$.

The pair (g, T) fully specifies $\tilde{g}$ and we specify a prior distribution for $\tilde{g}$ by assigning a joint prior probability model for the pair (g, T) . To that end, we use a Gaussian process to model g and a general distribution τ_T that does not depend on g to model T . We further ensure that the pair (g, T) results in a valid warping function by restricting the joint prior to the domain $B \equiv \{(g, T) : \exp_1(\tilde{g}) > 0\}$. The full model is given below.

Model 1.

$$q_1([t]) - q_2\left(\int_0^{[t]} \exp_1^2(\tilde{g})(s) ds\right) \cdot \exp_1(\tilde{g})([t]) \mid g, T, \sigma_1^2 \sim N(0_N, \sigma_1^2 I_N),$$

$$g, T \sim \{\text{Gaussian}(0, \mathcal{C}_g) \cdot \tau_T\}_B,$$

$$\sigma_1^2 \sim \text{IG}(\text{shape} = a, \text{scale} = b),$$

where $\tilde{g}$ is a function of g and T , C_g is a covariance operator for the Gaussian process prior, τ_T is a prior distribution for T , $\{\cdot\}_B$ denotes the truncation of the joint prior distribution to the set B , $IG(\cdot, \cdot)$ is the inverse-gamma distribution, and a and b are fixed constants.

The likelihood function is then given by

$$L(g, T, \sigma_1^2 \mid q_1, q_2) = \left(\frac{1}{\sqrt{2\pi}}\right)^N \left(\frac{1}{\sigma_1^2}\right)^{N/2} \exp\left\{-\frac{1}{2\sigma_1^2} \cdot \text{SSE}(\tilde{g})\right\}, \quad (6)$$

where

$$\text{SSE}(\tilde{g}) = \sum_{i=1}^N \left(q_1(t_i) - q_2 \left(\int_0^{t_i} \exp_1^2(\tilde{g})(s) ds \right) \exp_1(\tilde{g})(t_i) \right)^2. \quad (7)$$

This likelihood is identical to that of [11, 19], except that the truncated parameter $\tilde{g}$ replaces the parameter g .

Random truncation mechanisms

We now discuss the following two mechanisms for the prior distribution of the random truncation T , which is key to the proposed approach.

(1) Random Number of Basis Functions. In the first scenario, consider $T \equiv M$, where M is the number of basis functions used to represent the parameter g . In this case, $\pi_T \equiv \pi_M$ is a prior distribution on the set of positive integers and $\tilde{g}(t) = \sum_{i=1}^M \xi_i b_i(t)$, where $\{b_i\}_{i \geq 1}$ forms an orthonormal basis for $T_1(\Psi)$.

(2) Random Indicators. In the second scenario, we can randomly switch a basis function on and off (called a sieve prior in [22]). This is done via the random sequence $\{\chi_i\}_{i=1, \dots, \infty}$, $\chi_i \in \{0, 1\}$; we refer to this sequence simply as χ . This sequence controls which basis functions are kept in the basis expansion of g . In other words, $T \equiv \chi$ and $\tilde{g}$ is calculated as $\tilde{g}(t) = \sum_{i=1}^{\infty} \chi_i \cdot \xi_i b_i(t)$. Let $M_{\max}$ be the maximum number of basis functions stored at the implementation stage, and let M_{on} be the number of active basis functions (i.e., $M_{\text{on}} = \sum_{i=1}^{M_{\max}} \chi_i$). The domain of χ , denoted $\mathcal{X}$, is the collection of vectors of the form $(\chi_1 \in \{0, 1\}, \dots, \chi_{M_{\max}} \in \{0, 1\})$, and $\pi_T \equiv \pi_{\chi}$ is a prior on $\mathcal{X}$.

Posterior distribution and sampling via Markov chain Monte Carlo

The posterior distribution is a probability measure μ on the product space $T_1(\Psi) \times (\text{domain of } T) \times \mathbb{R}^+$ and is dominated by the prior measure μ_0 . The Radon-Nikodym derivative is given by Bayes' formula: $\frac{d\mu}{d\mu_0}(g, T, \sigma_1^2) \propto L(g, T, \sigma_1^2 \mid q_1, q_2)$. We use a Metropolis-within-Gibbs algorithm to sample from the posterior distribution of (g, T, σ_1^2) . To update the posterior at each step, the algorithm iteratively draws from (i) the full conditional distribution of (g, T) , and (ii) the full conditional distribution of σ_1^2 . Details are given as follows.

Sampling of (g, T) . We first update g . For this purpose, we use a Z -mixture p CN proposal (see [11, 22] for details) by setting $g' = g\sqrt{(1 - \beta_z^2) + \beta_z \xi}$, where g is the current value, ξ is a draw from the Gaussian process prior, and $\beta_z \in (0, 1)$ is a tuning parameter drawn with probability p_z satisfying $\sum_{z=1}^Z p_z = 1$. As an example, in the case of a 2-mixture p CN proposal, we can draw $\beta_1 = 0.5$ with probability 0.8 and $\beta_2 = 0.001$ with probability 0.2, resulting in “big jump proposals” approximately 80% of the time and “very small jump proposals” approximately 20% of the time. Sampling a new function ξ from $\text{Gaussian}(0, C_g)$ is done via the Karhunen-Loève expansion. We specify the covariance operator C_g by its eigenpairs $\{(b_i, \lambda_i^2)\}_{i \geq 1}$,

where $\{b_i(\cdot)\}$ forms an orthonormal basis for $T_1(\Psi)$ and the sequence of coefficients satisfy $\sum \lambda_i^2 < \infty$. While theoretically $i = 1, \dots, \infty$, at the implementation stage, we store a large number M_{max} of basis functions. We thus sample independent random variables $\xi_i \sim N(0, \lambda_i^2)$, $i = 1, \dots, M_{max}$. Then, $\xi(t) = \sum_{i=1}^M \xi_i b_i(t)$ if truncation mechanism (1) is used, or $\xi(t) = \sum_{i=1}^{M_{max}} \chi_i \cdot \xi_i b_i(t)$ if truncation mechanism (2) is used.

We then independently generate a proposal, T' , according to a density $\mathbb{Q}_T(\cdot|\cdot)$, which depends on the truncation mechanism. The pair (g', T') is accepted with probability $1 \wedge \rho$, where

$$\rho = \frac{L(g', T', \sigma_1^2) \cdot \pi_T(T') \cdot \mathbb{Q}_T(T|T')}{L(g, T, \sigma_1^2) \cdot \pi_T(T) \cdot \mathbb{Q}_T(T'|T)} \cdot I\{(g', T') \in B\}, \quad (8)$$

and π_T is the density function (with respect to the Lebesgue measure) of τ_T . This acceptance ratio is intuitive if distributions of g have densities with respect to the Lebesgue measure. In that case, the form of ρ follows directly from the fact that $\text{Gaussian}(g'; 0, \mathcal{C}_g) \cdot \mathbb{Q}_{pCN}(g|g')$ is symmetric in g and g' ($\mathbb{Q}_{pCN}$ is the pCN proposal described earlier). Since a dominating Lebesgue measure does not exist on $T_1(\Psi)$, we can derive the acceptance ratio formally using the dominating prior Gaussian measure (given in the [S1 File](#)).

If truncation mechanism (1) is used, $\mathbb{Q}_T(T|T') = \mathbb{Q}_M(M|M')$, and we can use a K -step random walk proposal of the form

$$M' = \begin{cases} M & \text{w.p. } p_0 \\ M \pm 1 & \text{w.p. } p_1 \\ M \pm 2 & \text{w.p. } p_2 \\ \vdots & \\ M \pm K & \text{w.p. } p_K \end{cases}$$

where $p_0 + \sum_{k=1}^K 2p_k = 1$. In other words, M' can either stay at the current value, with probability p_0 , or move forward or backward up to K steps. This is a symmetric proposal and the Metropolis Hastings (MH) acceptance ratio in [Eq \(8\)](#) simplifies to

$$\rho = \frac{L(g', M', \sigma_1^2) \cdot \pi_M(M')}{L(g, M, \sigma_1^2) \cdot \pi_M(M)} \cdot I\{(g', M') \in B\}. \quad (9)$$

If the truncation mechanism (2) is used, then $\mathbb{Q}_T(T|T') = \mathbb{Q}_\chi(\chi|\chi')$, and we can use an *on-off* proposal, where χ' is proposed by either switching on a nonactive basis function, with probability 0.5, or switching off an active basis function, again with probability 0.5. If all of the basis functions are currently on, i.e., $M_{on} = M_{max}$, we switch one of them off with probability 1. On the other hand, if only one basis function is on, i.e., $M_{on} = 1$, we switch on another basis function with probability 1. This is the form of proposal suggested in [\[22\]](#). It is not a symmetric proposal and the MH acceptance ratio in [Eq \(8\)](#) can be written as

$$\rho = \frac{L(g', \chi', \sigma_1^2) \cdot \pi_\chi(\chi')}{L(g, \chi, \sigma_1^2) \cdot \pi_\chi(\chi)} \cdot a_{\chi, \chi'} \cdot I\{(g', \chi') \in B\}, \quad (10)$$

where

$$a_{\chi, \chi'} = \begin{cases} (M_{\max} - M_{\text{on}})/(M_{\text{on}} + 1) & \text{if } M_{\text{on}} \neq 1, M_{\text{on}} \neq M_{\max}, M'_{\text{on}} = M_{\text{on}} + 1 \\ M_{\text{on}}/(M_{\max} - M_{\text{on}} + 1) & \text{if } M_{\text{on}} \neq 1, M_{\text{on}} \neq M_{\max}, M'_{\text{on}} = M_{\text{on}} - 1 \\ (M_{\max} - 1)/4 & \text{if } M_{\text{on}} = 1 \\ M_{\max}/2 & \text{if } M_{\text{on}} = M_{\max}. \end{cases}$$

The derivation of $a_{\chi, \chi'}$ is given in the [S1 File](#).

Alternatively, one can consider a symmetric, *choose-k* proposal, where χ' is proposed via the following steps: (i) sample $k \in \{1, 2, \dots, K \leq M_{\max}\}$ (for simplicity, we set the probability for each value of k to be the same, but this is not required by the algorithm), (ii) randomly select k entries from the vector $\chi = (\chi_1 \in \{0, 1\}, \dots, \chi_{M_{\max}} \in \{0, 1\})$, and (iii) propose χ' by switching (on to off, off to on) all of the selected entries. This proposal is symmetric. For example, if $\chi = (0, 0, 1, 1)$ and $\chi' = (0, 1, 0, 1)$, then

$\mathbb{Q}_{\chi}(\chi'|\chi) = \mathbb{Q}_{\chi'}(\chi|\chi') = P(\text{select 2nd and 3rd entries}|k=2)P(k=2)$. As a result, the MH acceptance ratio for this proposal is the same as the one given in [Eq \(10\)](#) with $a_{\chi, \chi'} = 1$.

Sampling of σ_1^2 . To update σ_1^2 , we draw directly from the conjugate inverse-gamma distribution with shape parameter $\frac{N}{2} + a$ and scale parameter $\frac{1}{2} \text{SSE}(\tilde{g}) + b$.

Multiple function Bayesian registration model with random basis truncation

The model for multiple function registration is a direct extension of Model (1). Based on the observed functions $f_1, \dots, f_C$, represented via the SRVFs $q_1, \dots, q_C$, we aim to make inference on $\tilde{g}_1, \dots, \tilde{g}_C$, which are determined by the pairs $(g_1, T_1), \dots, (g_C, T_C)$. In addition, we treat the template function q^* as a parameter and consider the same truncation mechanisms as described in the pairwise case. The truncated template function is denoted by $\tilde{q}^*$ and is determined by the pair (q^*, T_q) . We assign a Gaussian process prior to q^* and a prior τ_T^* to the random truncation T_q .

Model 2.

$$\begin{aligned} q_1 \left(\int_0^{[t]} \exp_1^2(\tilde{g}_1)(s) ds \right) \exp_1(\tilde{g}_1)([t]) - \tilde{q}^*([t]) & \mid q^*, T_q, g_1, T_1, \sigma_1^2 \sim N(0_N, \sigma_1^2 I_N), \\ & \vdots \\ q_C \left(\int_0^{[t]} \exp_1^2(\tilde{g}_C)(s) ds \right) \exp_1(\tilde{g}_C)([t]) - \tilde{q}^*([t]) & \mid q^*, T_q, g_C, T_C, \sigma_1^2 \sim N(0_N, \sigma_1^2 I_N), \\ (g_1, T_1) & \sim \{ \text{Gaussian}(0, \mathcal{C}_g) \cdot \tau_T \}_B, \\ & \vdots \\ (g_C, T_C) & \sim \{ \text{Gaussian}(0, \mathcal{C}_g) \cdot \tau_T \}_B, \\ (q^*, T_q) & \sim \text{Gaussian}(0, \mathcal{C}_q) \cdot \tau_T^*, \\ \sigma_1^2 & \sim \text{IG}(\text{shape} = a, \text{scale} = b). \end{aligned}$$

The likelihood function is then given by

$$L(g_1, T_1, \dots, g_C, T_C, q^*, T_q, \sigma_1^2) = \prod_{i=1}^C \left\{ \left(\frac{1}{\sqrt{2\pi}} \right)^N \left(\frac{1}{\sigma_1^2} \right)^{N/2} \exp \left(-\frac{1}{2\sigma_1^2} \text{SSE}(\tilde{g}_i, \tilde{q}^*) \right) \right\}, \quad (11)$$

where

$$\text{SSE}(\tilde{g}_i, \tilde{q}^*) = \sum_{j=1}^N \left(\tilde{q}^*(t_j) - q_i \left(\int_0^{t_j} \exp_1^2(\tilde{g}_i)(s) ds \right) \exp_1(\tilde{g}_i)(t_j) \right)^2. \quad (12)$$

Sampling from the posterior distribution is performed in the same fashion as in the pairwise case using a Metropolis-within-Gibbs algorithm. At each step, the algorithm first updates each of the truncated warping parameters (g_b, T_i) , $i = 1, \dots, C$ sequentially via a Metropolis step. The acceptance ratio for updating any of the pairs (g_b, T_i) takes the same form and, for (g_1, T_1) , is given by

$$\begin{aligned} & 1 \wedge \frac{L(g'_1, T'_1, \dots, g_C, T_C, q^*, T_q, \sigma_1^2) \cdot \pi_T(T'_1) \cdot \mathbb{Q}_T(T_1|T'_1)}{L(g_1, T_1, \dots, g_C, T_C, q^*, T_q, \sigma_1^2) \cdot \pi_T(T_1) \cdot \mathbb{Q}_T(T'_1|T_1)} \cdot I\{(g'_1, T'_1) \in B\} \\ &= 1 \wedge \frac{\exp\left\{-\frac{1}{2\sigma_1^2} \text{SSE}(\tilde{g}'_1, \tilde{q}^*)\right\} \cdot \pi_T(T'_1) \cdot \mathbb{Q}_T(T_1|T'_1)}{\exp\left\{-\frac{1}{2\sigma_1^2} \text{SSE}(\tilde{g}_1, \tilde{q}^*)\right\} \cdot \pi_T(T_1) \cdot \mathbb{Q}_T(T'_1|T_1)} \cdot I\{(g'_1, T'_1) \in B\}. \end{aligned}$$

The algorithm then updates the template (q^*, T_q) with the acceptance ratio

$$\begin{aligned} & 1 \wedge \frac{L(g_1, T_1, \dots, g_C, T_C, q', T'_q, \sigma_1^2) \cdot \pi_T(T'_q) \cdot \mathbb{Q}_T(T_q|T'_q)}{L(g_1, T_1, \dots, g_C, T_C, q^*, T_q, \sigma_1^2) \cdot \pi_T(T_q) \cdot \mathbb{Q}_T(T'_q|T_q)} \\ &= 1 \wedge \frac{\exp\left\{-\frac{1}{2\sigma_1^2} \sum_i \text{SSE}(\tilde{g}_i, \tilde{q}^*)\right\} \cdot \pi_T(T'_q) \cdot \mathbb{Q}_T(T_q|T'_q)}{\exp\left\{-\frac{1}{2\sigma_1^2} \sum_i \text{SSE}(\tilde{g}_i, \tilde{q}^*)\right\} \cdot \pi_T(T_q) \cdot \mathbb{Q}_T(T'_q|T_q)}. \end{aligned}$$

Note that, for simplicity, we use the same prior π_T and proposal $\mathbb{Q}_T$ for each T_b , $i = 1, \dots, C$ and T_q , but this is not required. Lastly, the algorithm updates σ_1^2 by drawing directly from an inverse-gamma distribution with shape parameter $\frac{1}{2}C \cdot N + a$ and scale parameter

$$\frac{1}{2} \sum_{i=1}^C \text{SSE}(\tilde{g}_i, \tilde{q}^*) + b.$$

Simulation study

We first assess the performance of the proposed pairwise Bayesian registration model via a simulation study. We simulate two observed functions, f_1 and f_2 , both of which are warped versions of a template function, $f(t) = \sin(4\pi t^2)$ [2]. The two observed functions are constructed such that $f_1 = f_2 \circ \gamma_{\text{true}}$ where the true warping γ_{true} is randomly generated. We then apply our model to register f_2 to f_1 to obtain γ_{est} , an estimate of the true warping, and compare γ_{est} to γ_{true} . We use five sets of true warping functions, which are shown in Fig 2. Each set contains ten warpings and has varying degrees of local features: set (2) shown in the second panel in the top row contains warping functions that are very smooth with no “small wiggles,” whereas set (5) (fifth panel in the top row) contains warping functions that are “very wiggly.” Set (1) contains piecewise linear warping functions. We compare the proposed method to the model in [11], with the recommended setting of using 20 basis functions (we refer to this model as M20).

Prior Specification

We now discuss different choices for the prior distributions for the pairwise registration Model (1). We use an inverse-gamma distribution with $a = 0.1$ and $b = 0.1$ for σ_1^2 . For the

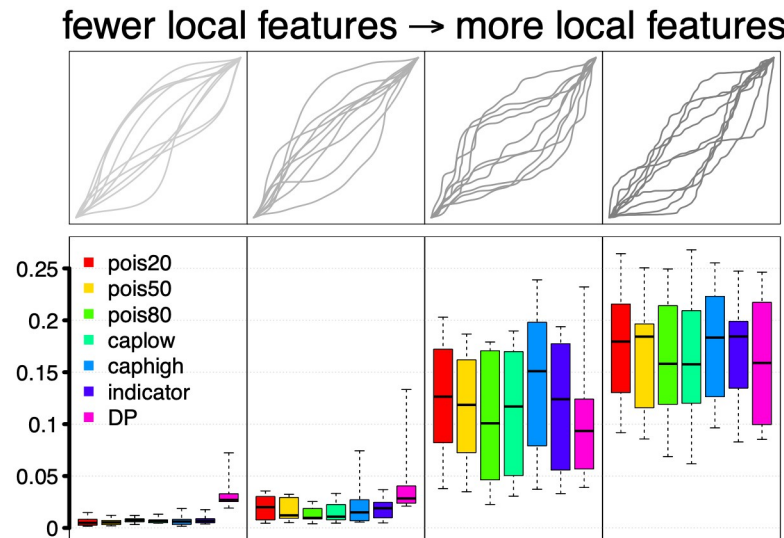


Fig 2. Pairwise simulation results. Top: Five sets of ten true warping functions with different amount of local features. Bottom: FR distance-based performance of different models for estimating the true warping functions in each set; smaller distance indicates better performance.

<https://doi.org/10.1371/journal.pone.0287734.g002>

Gaussian process prior of the warping parameter g , we must specify the covariance operator $\mathcal{C}_g$ by its eigenpairs $\{(b_i, \lambda_i^2)\}_{i \geq 1}$. We use the Fourier basis functions $b_i(t) \in \{\sqrt{2} \sin(2\pi i t), \sqrt{2} \cos(2\pi i t)\}_{i \geq 1}$ and set $\lambda_i^2 = 4^2 \cdot i^{-1.2}$. The constant 1.2 in this expression controls the decay rate of the Fourier series. If it is desirable to put more prior weight on warping functions with many local features, one can choose a smaller constant so that the higher frequency Fourier basis elements are weighted more (this constant should be greater than 1 to satisfy $\sum \lambda_i^2 < \infty$). For the random truncation T , we use a few different priors based on which truncation mechanism is used, as described below.

Truncation mechanism (1): There are multiple possible prior choices for the number of basis functions. First, we consider three Poisson priors, truncated to the domain $[1, M_{\max} = 200]$, with means equal to 20 (*pois20*), 50 (*pois50*) and 80 (*pois80*). These priors reflect beliefs about the smoothness of the warping functions. For instance, if the prior belief is that the warping function is relatively smooth and the phase variation should only relate to the general shape of the observed functions, a prior with a smaller mean should be chosen. Second, we consider two discrete uniform prior distributions on $[30, M_{\max} = 200]$ (*caplow*) and $[1, 50]$ (*caphigh*), respectively. These two priors enforce a maximum (minimum) level of smoothness of the warping functions via the restriction that $M \geq 30$ ($M \leq 50$), but are non-informative in the sense that any M between 30 and $M_{\max} = 200$ (1 and 50) is equally likely.

Truncation mechanism (2): We choose a uniform prior distribution on the sequence of on-off switches for the basis functions (*indicator*). This prior is non-informative on the shape of the warping functions since any of the basis elements in the sequence are equally likely to be switched on or off.

The flexibility of prior choices for the random truncation is an important benefit of the proposed model, in comparison to existing models with fixed truncation. In practice, if one wishes to register the general shapes of the functions without altering small, local features, a prior that puts more weight on smoother functions (i.e., truncation mechanism (1) with Poisson with a small mean or a uniform with a small upper bound) should be chosen. If one has no strong

prior opinion and wants the posterior to be mostly informed by the data, a less informative prior (i.e., truncation mechanism (2) with uniform for the on-off indicators) should be chosen. At the same time, the model is robust to the prior choices of the decay rate of the Fourier basis and the model variance σ_1^2 (see the [S1 File](#) for a sensitivity analysis).

Implementation details

At the implementation stage, both the observed functions and the parameters are stored on a grid of size 200, which is the same as M_{max} , the number of stored basis functions. We first perform pairwise registration for each set of warping functions using the deterministic Dynamic Programming (DP) algorithm, which is implemented in the R package *fdasrvf* [27]. We use the DP estimate to initialize the MCMC sampling algorithm for the proposed Bayesian model. We note that, while DP offers a good starting point and allows the chain to mix faster, the performance of the MCMC algorithm is independent of the starting point in the long run. Examples with different starting points are included in the [S1 File](#). For the warping parameter g , we use a 3-mixture *pCN* proposal with jump sizes $\beta_z = (0.5, 0.05, 0.0001)$ and corresponding proposal probabilities $p_z = (0.3, 0.3, 0.4)$. For truncation mechanism (1), the proposal for M is a 10-step random walk where M' stays at the current value or moves forward or backward by up to 10 steps (with the probability for each move $(p_0, p_1, \dots, p_{10}) \propto (1, 0.5, 0.445, \dots, 0.001)$, where $0.5, 0.445, \dots, 0.001$ is an equally spaced decreasing sequence of length 10). For truncation mechanism (2), the proposed state χ' is generated by switching up to five randomly selected indicators. We use the *choose-k* proposal because we notice that it tends to have faster convergence than the *on-or-off* alternative.

Results

To evaluate performance, we calculate the Fisher-Rao distance between γ_{true} and γ_{est} [24]:

$$d_{FR}(\gamma_{true}, \gamma_{est}) = \cos^{-1}(\langle \psi_{true}, \psi_{est} \rangle) = \cos^{-1} \left(\int_0^1 \psi_{true}(t) \psi_{est}(t) dt \right), \quad (13)$$

where the posterior mean is used to construct γ_{est} for the Bayesian models. The results are shown in [Fig 2](#). Importantly, we see that the *M20* model proposed in [11] does not perform well compared to the proposed random truncation or random indicator models when the true warping functions have many local features. While M can be fixed to a larger value, our model has the advantage that the value of M does not need to be decided a priori and, instead, is informed by the data. Comparing to DP, we notice that when γ_{true} has fewer local features (sets (2) and (3)), the proposed Bayesian models perform better irrespective of the prior distribution for the random truncation T . When γ_{true} has more local features (sets (4) and (5)), models *pois20* and *caphigh* perform worse, as expected, since they impose strong prior constraints on the maximum smoothness of the warping function. For the set of linear functions, DP performs better than the Bayesian models. This is not surprising since DP performs registration by solving a minimization problem in a piecewise linear fashion. On the other hand, the Fourier basis used for the Bayesian models are not piecewise linear. We also note that, while we compare DP with the Bayesian methods numerically using the FR distance, they are fundamentally different approaches. DP is optimization-based and the algorithm is very fast, but it has some known limitations: (1) it is not easy to enforce restrictions on the parameter space, (2) while regularization can be imposed by adding a term to the cost function, the choice of the regularization parameter is difficult in practice, (3) performance of the algorithm heavily depends on the discretization and neighborhood size settings, and (4) there is no prescribed

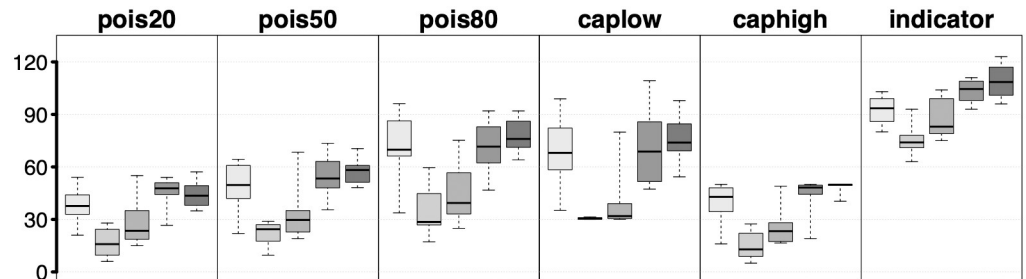


Fig 3. Posterior mean of the number of (active) basis functions for different registration models in the pairwise simulation study. The five boxplots in each panel correspond to the five sets of true warping functions (top row of Fig 2).

<https://doi.org/10.1371/journal.pone.0287734.g003>

approach for uncertainty quantification. In comparison, model-based approaches are more flexible and provide a more principled exploration of the parameter space.

We also examine the posterior means of the number of active basis functions in Fig 3. Based on this result, we highlight two additional advantages of the proposed method compared to the model in [11]. First, inference on the level of smoothness of the warping function is clearly informed by the underlying data. We see that when the true warping functions, and consequently the observed functions, have more local features, the (posterior mean) number of active basis functions is larger regardless of the prior chosen for the random truncation T . Note that the set of linear functions requires more basis functions due to the sharp turns at the break points. Second, one can flexibly incorporate constraints on the level of smoothness via this prior. For example, the model *pois20* generates fewer active basis functions than the model *pois80*, regardless of the level of smoothness in the observed functions. The model *caplow* results in estimated warping functions based on at least 30 basis elements even when the true warping function is very smooth. On the other hand, the model *caphigh* results in estimated warping functions based on at most 50 basis elements even when the true warping function has many local features. Fig 4 shows four examples that compare registration performance between the Bayesian models *caplow* and *caphigh*. In each example, we display f_2 (grey), $f_1 = f_2 \circ \gamma_{true}$ (black), $f_2 \circ \gamma_{est, caplow}$ (green) and $f_2 \circ \gamma_{est, caphigh}$ (blue). The estimated warping functions are obtained using the posterior means in each case. It is clear that the model *caplow* allows for registration of more local features than the model *caphigh*.

Applications

In this section, we apply the proposed multiple function Bayesian registration model to the six datasets displayed in Fig 1. They arise in five different application domains:

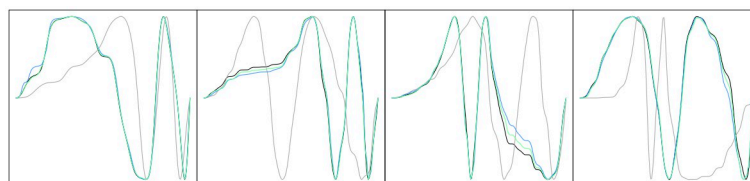


Fig 4. Four examples of pairwise Bayesian registration with priors *caplow* and *caphigh*. In each panel, we show f_1 (black), f_2 (grey), $f_2 \circ \gamma_{est, caplow}$ (green) and $f_2 \circ \gamma_{est, caphigh}$ (blue). The estimated warping functions correspond to posterior means.

<https://doi.org/10.1371/journal.pone.0287734.g004>

1. **Right knee flexion and pelvis right roll.** This data is comprised of two gait variables (measurements taken by markers as participants walk) for $C = 12$ participants. We obtained the data from the online supplementary material of [28]. Each gait cycle is linearly scaled such that they are observed on the same time interval, i.e., the x -axis can be interpreted as percentage of one gait cycle. For more information on the role of registration in gait cycle analysis see, e.g., [10, 11, 29]
2. **Growth rates.** We use the growth rate functions of $C = 54$ girls from the Berkeley growth study [1], available in the R package *fda* [30]. This dataset is widely used to assess registration performance.
3. **Pinch force.** Each function records the pinch force exerted by the thumb and forefingers during a brief squeeze. These measurements were collected for $C = 20$ test subjects [2, 12, 31] and the full dataset is available in the R package *fda*. The starting time of the pinch as well as the time spent to reach the maximum force are different across test subjects, necessitating a registration step to account for this temporal variation prior to further analysis.
4. **Neural spike trains (sequence of electrical pulses sent by the neurons to the brain).** The dataset is comprised of $C = 10$ smoothed neural spike train functions; see [9] for a detailed description. This dataset is analyzed in multiple papers, including [24, 32–36], with a particular interest in function registration.
5. **Handwriting samples.** We use the x -coordinates of $C = 50$ replicates (generated by a single person) of handwritten Chinese characters for ‘statistical science’ [6]), available in the R package *fda*. This dataset is used in [5] as an application of function registration.

Prior specification

The covariance operator in the Gaussian process prior of the warping functions, $\mathcal{C}_g$, and the inverse-gamma prior for the model variance, σ_1^2 , are identical to the priors specified in the pairwise registration simulation study. The covariance operator in the Gaussian process prior for the template function, $\mathcal{C}_q$, is specified by the Fourier basis with corresponding eigenvalues $\lambda_i^2 = \sigma_q^2 \cdot i^{-1.2}$, where σ_q^2 is fixed based on the scale of the observed functions. Using the Fourier basis to represent the template function is especially suitable when the observed functions are periodic on $[0, 1]$, i.e., the gait cycle variables. For other datasets, we set the Fourier period to 2. As shown in the simulation study, these prior choices yield good registration results across different shapes of observed functions; on the other hand, the model is robust to alternative prior choices for these model parameters.

The random truncation is specified by truncation mechanism (1). For both the warping functions and the template, we use a discrete uniform prior on $[5, M_{\max} = 200]$ for the number of basis functions M . This prior is non-informative and allows us to evaluate registration performance that is primarily driven by the data. Intuitively, the chosen range, $[5, M_{\max} = 200]$, ensures that the warping functions have a minimum level of complexity (as captured by the first five Fourier basis elements), but are allowed to have as many local features as possible (since the observed functions are evaluated on a discretized grid of size 200). In practice, this is a suitable approach when one does not have strong prior information for the warping parameter and does not wish to restrict the level of shape-alteration during the registration process. For the proposals, we want to have a variety of small and intermediate jump sizes to explore the parameter space thoroughly. To that end, we use a p CN proposal with β_z values equally spaced between 0.001 and 0.0001 for the functional parameters and use a 1-step random walk

proposal for the number of basis functions. Convergence is visually monitored by checking the trace plots of the log-likelihood. Trace plots for two of the datasets are provided in the [S1 File](#).

Results

For comparison, we also perform registration with the *M20* model [11]. Registration results for the six different datasets under consideration are shown in [Fig 5](#). In each panel, we show the registered data (left) with respect to the estimated template (middle). For the estimated template, we also visualize the level of uncertainty (as measured by the posterior pointwise standard deviations, standardized by the scale of the original data; blue = small standard deviation, red = large standard deviation). We see that, when the observed functions have very different shapes (e.g., neural spike trains), the estimated template tends to have more uncertainty. In the right panel, we plot the template function estimated using the *M20* model. We see that the proposed method is better at producing a template function that resembles the shape of the original functions. For instance, the template estimated by the *M20* model does not have the small wiggles at both ends of the pinch force curves, and the *M20* model cannot recover a good template for the handwriting curves. This, again, shows the limitation of the model proposed in [11], where the number of basis functions can be mis-specified. In comparison, the proposed model uses 178 basis functions (posterior average) to estimate the template function of the pinch force dataset and 196 basis functions for the handwriting dataset. When the observed functions are relatively smooth, the posterior of the template function reflects that by using fewer basis functions (e.g., 55 and 33 for the two gait datasets). This shows the data-informative feature of the proposed method.

For a quantitative assessment of registration results, we report the inverse of pairwise correlation (IPC) [11, 15], calculated using $IPC = \frac{\sum_{i \neq j} r(f_i, f_j)}{\sum_{i \neq j} r(\hat{f}_i, \hat{f}_j)}$, where $r(\cdot, \cdot)$ is the pairwise Pearson's

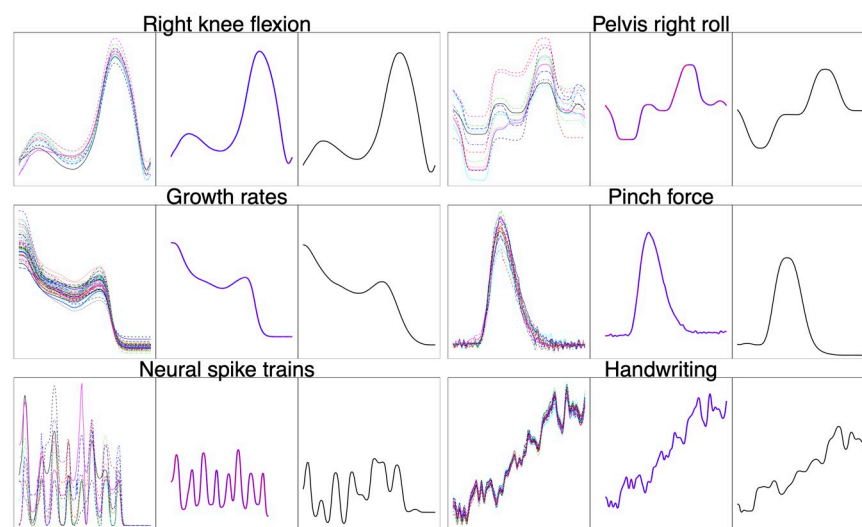


Fig 5. Registration results for six real datasets. The original observed functions for each dataset can be found in [Fig 1](#). Here, we display the registered functions (left panel), the estimated template function using the proposed random truncation model (middle panel), with the color corresponding to the pointwise standard deviation (red—larger standard deviation and more uncertainty; blue—smaller standard deviation and less uncertainty), and the estimated template function using the *M20* model (right panel). The warping and template functions used to perform registration are estimated using the posterior means.

<https://doi.org/10.1371/journal.pone.0287734.g005>

correlation, f are the observed functions, and $\tilde{f}$ are the registered functions. The IPC values corresponding to registration performance of different models are shown in Fig 6. Overall results are comparable across different models when the observed functions are relatively smooth. The *M20* model notably does not perform as well when the observed functions have many local features (e.g. pinch force, spike trains, and handwriting data) due to an under-specification of the number of basis functions. Compared to DP, the proposed Bayesian model achieves similar performance for all datasets (in the case of multiple function registration, DP aligns each function to an estimated template function in a pairwise manner; see [24] for details). Another numerical criterion (Sync) based on the $\mathbb{L}^2$ distance between the registered functions shows similar results and is discussed in the S1 File.

We note that the IPC (and Sync) values are based only on the correlation or the distance between the registered functions and do not take into account how well the shapes of the functions are preserved. It also only accounts for the quality of the template indirectly via the registered functions. As we have shown in Fig 5, the *M20* model does not recover the shape of the template function as well as the proposed method. On the other hand, DP sometimes does not preserve the original shapes of the observed functions after registration. For instance, we highlight two neural spike trains in the top row of Fig 7. The original observed curves are shown in the left panel; we see that one of the curves (blue) exhibits more local features than the other (red). Despite a good registration result (middle panel), DP has smoothed out the unmatched local features in the blue curve. In contrast (right panel), the proposed Bayesian model preserves the original shapes of the two curves better. In fact, when the observed functions have different numbers of features, e.g., local extrema, the random truncation component of the Bayesian model enables one to detect this pattern. The two highlighted neural spike trains were identified, because their corresponding warping parameters have the largest posterior means for the number of basis functions. An estimated warping function represented by a large number of basis elements signals that the observed function has been altered a lot in its local features after registration, likely due to some unmatched features being “squeezed,” as evident in this example.

In some applications, it is not desirable to alter the shapes of the observed functions too much. The proposed Bayesian model offers a flexible way to control how much shape

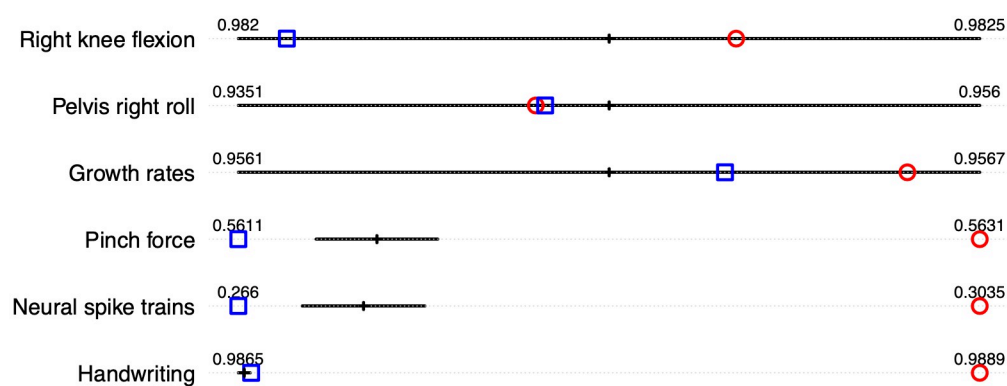


Fig 6. IPC values for six real datasets after registration. Smaller value means better alignment. The black line corresponds to the proposed approach (specifically, this is a 95% credible interval of the IPC values constructed using 200 randomly sampled posterior draws); the red round dot corresponds to the *M20* model; the blue square corresponds to DP. Since the IPC value for each dataset is plotted on a different scale (and the values are not directly comparable across different datasets), we display the numeric values of the end points as a reference.

<https://doi.org/10.1371/journal.pone.0287734.g006>

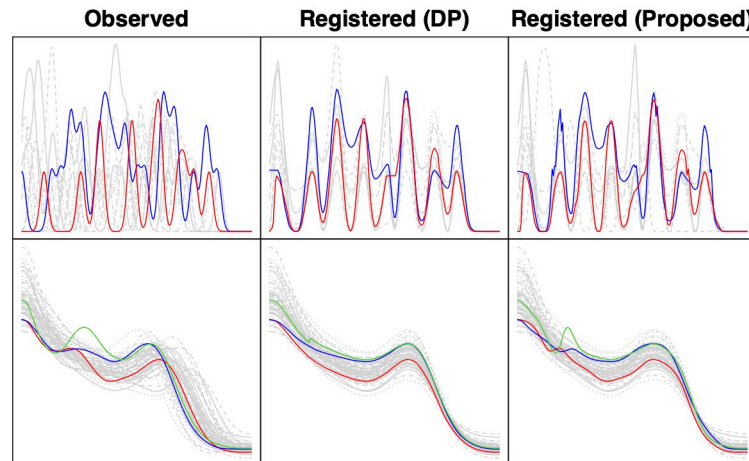


Fig 7. Registration results highlighting two neural spike trains (top) and three growth rate functions (bottom).

<https://doi.org/10.1371/journal.pone.0287734.g007>

alteration is allowed via the prior distribution of the random truncation parameter. Specifically, shape-alteration during the registration process will be limited if the prior puts very small or zero probability on large values of the number of basis functions. We show an example of this for the growth rate functions in the bottom row of Fig 7. We again highlight three observed functions in the left panel. They all exhibit two modes while most of the other functions in the data have only one mode. After registration using DP (middle panel), the bimodal pattern in the highlighted functions is no longer obvious. As a result, one might overlook the fact that these individuals have two growth spurts rather than the more common pattern of one pubertal growth spurt. To limit the level of shape alteration, we apply the proposed Bayesian registration model with a restrictive discrete uniform prior on $[1, 10]$ for random truncation for the warping parameters. In this case, the prior limits the number of basis functions to be at most 10. The corresponding results (right panel) show that the bimodal feature of the highlighted growth rate functions is much better preserved after registration.

Summary

We develop Bayesian models for pairwise and multiple function registration. These models build on existing Bayesian registration techniques which assign Gaussian process priors to the warping function, after a sequence of function space transformations. When building the registration models, the functional parameter is represented via an infinite sequence of basis functions, but at the implementation stage, it is necessary to truncate this sequence. Our main contribution lies in the randomization of this truncation process. This is done by introducing a new random truncation parameter that controls how many or which basis functions are used to represent the functional parameters. The resulting Bayesian models can then explore the full parameter space instead of a small subset of a truncated parameter space.

In practice, there are three main benefits of the proposed method compared to models where the truncation mechanism is fixed. First, the posterior distribution on the truncation parameter is informed by both the data and the prior, and one does not have to choose the truncation a priori, thus avoiding possible mis-specifications of the truncation parameter. Second, one can put restrictions on the registration process by using a restrictive prior for the truncation parameter, which controls how much shape alteration is allowed. For instance, as we have shown in the growth rate example, by limiting the number of basis functions to be at

most 10 for the warping function parameter, we retain important features (one or two growth spurts) in the registered growth rate functions. Third, the new models also allow us to make inference on how much truncation has occurred. This can help detect when an observed function has undergone a significant shape change during registration, especially in its local features. We demonstrate the aforementioned advantages of the proposed approach through a simulation study and multiple real datasets.

In addition, the proposed modeling framework is very flexible. In the case of multiple function registration, one can use a different prior for each of the warping functions, corresponding to each of the observed functions, and for the template function, e.g., one can use non-informative priors for the warping functions and a restrictive prior for the template function that would constrain the template to have a smooth shape with limited local features. This enables fine-tuning of the models based on the application of interest.

The Metropolis-within-Gibbs algorithm we use to sample from the posterior distribution is efficient in the sense that the computational cost is largely unaffected by the size of the grid on which the functions are observed. The proposals for the functional and random truncation parameters further allow much flexibility in the jump sizes for exploring the parameter space. On the other hand, a drawback of the algorithm is that it is not informed by the likelihood. As a result, in some applications, convergence can be slow; for example, based on trace plots, the real datasets considered in the Applications section require at least 2×10^5 iterations and can take more than 10^6 iterations (trace plots for two datasets are given as examples in the [S1 File](#); computation time for registering 10 functions is about 100 minutes per 10^5 updates). While this is not surprising, because the algorithm is exploring a very large parameter space, a possible future research direction is to design algorithms with faster convergence rates, by using likelihood-informed proposals or adaptive proposals with jump sizes automatically tuned by acceptance rates.

Supporting information

S1 File. Supplementary material. This pdf file serves as an appendix to the main manuscript and includes: 1) additional derivations; 2) trace plots for two real data examples; 3) Sync values for assessing registration performance for the six real datasets in Section; and 4) a discussion of model sensitivity.
(PDF)

S2 File. Code and data. The code folder includes R code to perform pairwise and multiple function registration. Datasets used in the manuscript are also included as .RData files. A readme file is included to provide instructions.
(ZIP)

Author Contributions

Conceptualization: Yi Lu, Radu Herbei, Sebastian Kurtek.

Formal analysis: Yi Lu, Radu Herbei, Sebastian Kurtek.

Methodology: Yi Lu, Radu Herbei, Sebastian Kurtek.

Visualization: Yi Lu.

Writing – original draft: Yi Lu.

Writing – review & editing: Yi Lu, Radu Herbei, Sebastian Kurtek.

References

1. Tuddenham RD, Snyder MM. Physical Growth of California Boys and Girls from Birth to Eighteen Years. Publications in Child Development at University of California, Berkeley. 1954; 1(2):183. PMID: [13217130](https://doi.org/10.13217130)
2. Ramsay JO, Li X. Curve Registration. *Journal of the Royal Statistical Society Series B*. 1998; 60:351–363. <https://doi.org/10.1111/1467-9868.00129>
3. Ramsay JO, Gribble P, Kurtek S. Analysis of juggling data: Landmark and continuous registration of juggling trajectories. *Electronic Journal of Statistics*. 2014; 8(2):1835–1841.
4. Ramsay JO, Gribble P, Kurtek S. Description and Processing of Functional Data Arising from Juggling Trajectories. *Electronic Journal of Statistics*. 2014; 8(2):1811–1816.
5. Kneip A, Li X, MacGibbon KB, Ramsay JO. Curve Registration by Local Regression. *The Canadian Journal of Statistics*. 2000; p. 19–29. <https://doi.org/10.2307/3315251.n>
6. Ramsay JO. Functional Components of Variation in Handwriting. *Journal of the American Statistical Association*. 2000; 95(449):9–15. <https://doi.org/10.1080/01621459.2000.10473894>
7. Telesca D, Inoue LYT. Bayesian Hierarchical Curve Registration. *Journal of the American Statistical Association*. 2008; 103(481):328–339. <https://doi.org/10.1198/016214507000001139>
8. Koch I, Hoffmann P, Marron JS. Proteomics Profiles from Mass Spectrometry. *Electronic Journal of Statistics*. 2014; 8(2):1703–1713.
9. Wu W, Hatsopoulos NG, Srivastava A. Introduction to neural spike train data for phase-amplitude analysis. *Electronic Journal of Statistics*. 2014; 8(2):1759–1768. <https://doi.org/10.1214/14-EJS865>
10. Helwig NE, Hong S, Hsiao-Weckler ET, Polk JD. Methods to Temporally Align Gait Cycle Data. *Journal of Biomechanics*. 2011; 44(3):561–566. <https://doi.org/10.1016/j.jbiomech.2010.09.015> PMID: [20887992](https://pubmed.ncbi.nlm.nih.gov/20887992/)
11. Lu Y, Herbei R, Kurtek S. Bayesian Registration of Functions with a Gaussian Process Prior. *Journal of Computational and Graphical Statistics*. 2017; 26(4):894–904. <https://doi.org/10.1080/10618600.2017.1336444>
12. Ramsay JO, Silverman BW. *Functional Data Analysis*. 2nd ed. New York: Springer-Verlag; 2005.
13. Srivastava A, Klassen EP. *Functional and Shape Data Analysis*. vol. 1. New York: Springer; 2016.
14. Claeskens G, Silverman BW, Slaets L. A Multiresolution Approach to Time Warping Achieved by a Bayesian Prior–posterior Transfer Fitting Strategy. *Journal of the Royal Statistical Society*. 2010; 72(5):673–694. <https://doi.org/10.1111/j.1467-9868.2010.00752.x>
15. Cheng W, Dryden IL, Huang X. Bayesian Registration of Functions and Curves. *Bayesian Anal*. 2016; 11(2):447–475. <https://doi.org/10.1214/15-BA957>
16. Bharath K, Kurtek S. Distribution on warp maps for alignment of open and closed curves. *Journal of the American Statistical Association*. 2020; 115(531):1378–1392. <https://doi.org/10.1080/01621459.2019.1632066> PMID: [34413553](https://pubmed.ncbi.nlm.nih.gov/34413553/)
17. Matuk J, Bharath K, Chkrebtii O, Kurtek S. Bayesian Framework for Simultaneous Registration and Estimation of Noisy, Sparse, and Fragmented Functional Data. *Journal of the American Statistical Association*. 2021; p. 1–17. <https://doi.org/10.1080/01621459.2021.1893179> PMID: [36945325](https://pubmed.ncbi.nlm.nih.gov/36945325/)
18. Earls C, Hooker G. Variational Bayes for Functional Data Registration, Smoothing, and Prediction. *Bayesian Analysis*. 2017; 12(2):557–582. <https://doi.org/10.1214/16-BA1013>
19. Kurtek S. A Geometric Approach to Pairwise Bayesian Alignment of Functional Data Using Importance Sampling. *Electronic Journal of Statistics*. 2017; 11(1):502–531. <https://doi.org/10.1214/17-EJS1243>
20. Matuk J, Herbei R, Kurtek S. Bayesian Registration of Functions. *Wiley StatsRef: Statistics Reference Online*. 2021; p. 1–15. <https://doi.org/10.1002/9781118445112.stat08265>
21. Stuart AM. Inverse Problems: a Bayesian Perspective. *Acta Numerica*. 2010; 19:451–559. <https://doi.org/10.1017/S0962492910000061>
22. Cotter SL, Roberts GO, Stuart AM, White D, et al. MCMC Methods for Functions: Modifying Old Algorithms to Make them Faster. *Statistical Science*. 2013; 28(3):424–446. <https://doi.org/10.1214/13-STS421>
23. Srivastava A, Klassen E, Joshi SH, Jermyn IH. Shape Analysis of Elastic Curves in Euclidean Spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2011a;33(7):1415–1428.
24. Srivastava A, Wu W, Kurtek S, Klassen E, Marron JS. Registration of Functional Data Using Fisher-Rao Metric. *arXiv preprint*. 2011b;arXiv:1103.3817.
25. Kurtek S, Srivastava A, Klassen E, Ding Z. Statistical Modeling of Curves Using Shapes and Related Features. *Journal of the American Statistical Association*. 2012; 107(499):1152–1165. <https://doi.org/10.1080/01621459.2012.699770>

26. Srivastava A, Jermyn I, Joshi S; IEEE. Riemannian Analysis of Probability Density Functions with Applications in Vision. IEEE Conference on Computer Vision and Pattern Recognition. 2007; p. 1–8.
27. Tucker JD. Elastic Functional Data Analysis; 2020. Available from: <https://cran.r-project.org/web/packages/fdasrvf/fdasrvf.pdf>.
28. van den Bogert AJ, Geijtenbeek T, Even-Zohar O, Steenbrink F, Hardin EC. A Real-time System for Biomechanical Analysis of Human Movement and Muscle Function. Medical & biological engineering & computing. 2013; 51(10):1069–1077. <https://doi.org/10.1007/s11517-013-1076-z> PMID: 23884905
29. Sadeghi H, Allard P, Shafie K, Mathieu PA, Sadeghi S, Prince F, et al. Reduction of Gait Data Variability Using Curve Registration. Gait & posture. 2000; 12(3):257–264. [https://doi.org/10.1016/S0966-6362\(00\)00085-0](https://doi.org/10.1016/S0966-6362(00)00085-0) PMID: 11154937
30. Ramsay JO, Wickham H, Graves S, Hooker G. Functional Data Analysis; 2014. Available from: <https://cran.r-project.org/web/packages/fda/fda.pdf>.
31. Ramsay JO, Wang X, Flanagan R. A Functional Data Analysis of the Pinch Force of Human Fingers. Applied Statistics. 1995; p. 17–30. <https://doi.org/10.2307/2986192>
32. Patriarca M, Sangalli LM, Secchi P, Vantini S. Analysis of spike train data: An application of *k*-mean alignment. Electronic Journal of Statistics. 2014; 8(2):1769–1775. <https://doi.org/10.1214/14-EJS865A>
33. Wu W, Srivastava A. Analysis of spike train data: Alignment and comparisons using the extended Fisher-Rao metric. Electronic Journal of Statistics. 2014; 8(2):1776–1785. <https://doi.org/10.1214/14-EJS865B>
34. Cheng W, Dryden IL, Hitchcock DB, Le H. Analysis of spike train data: Classification and Bayesian alignment. Electronic Journal of Statistics. 2014; 8(2):1786–1792. <https://doi.org/10.1214/14-EJS900C>
35. Lu X, Marron JS. Analysis of spike train data: Comparison between the real and the simulated data. Electronic Journal of Statistics. 2014; 8(2):1793–1796. <https://doi.org/10.1214/14-EJS865D>
36. Hadjipantelis PZ, Aston JAD, Müller HG, Moriarty J. Analysis of spike train data: A multivariate mixed effects model for phase and amplitude. Electronic Journal of Statistics. 2014; 8(2):1797–1807. <https://doi.org/10.1214/14-EJS865E>