



Robust curve registration using the t distribution

Mariko Takagishi¹ · Hiroshi Yadohisa²

Received: 22 October 2018 / Accepted: 12 February 2019 / Published online: 22 February 2019
© The Author(s) 2019

Abstract

In functional data analysis, it is often of interest to discover a general common pattern, or shape, of the function. When the subject-specific amplitude and phase variation of data are not of interest, curve registration can be used to separate the variation from the data. Shape-invariant models (SIM), one of the registration methods, aim to estimate the unknown shared-shape function. However, the use of SIM and of general registration methods assumes that all curves have the shared-shape in common and does not consider the existence of outliers, such as a curve, whose shape is inconsistent with the remainder of the data. Therefore, we propose using the t distribution to robustify SIMs, allowing outliers of amplitude, phase, and other errors. Our SIM can identify and classify the three types of outliers mentioned above. We use simulation and an empirical data set to evaluate the performance of our robust SIM.

Keywords Outliers · Functional data analysis · Registration · Mixed-effect model · Robust estimation

1 Introduction

Data that are observed repeatedly from subjects at different times can be assumed to be generated from underlying smooth functions. Functional data analysis (Ramsay and Silverman 2005) is concerned with the underlying curves.

If we are interested in discovering a general pattern or shape exhibited by the function that underlies the functions of all the subjects, then the aim of data analysis

Communicated by Hidetoshi Matsui.

✉ Mariko Takagishi
takagishi@sigmath.es.osaka-u.ac.jp

¹ Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka, Japan

² Department of Culture and Information Science, Doshisha University, 1-3 Tatara Miyakodani, Kyotanabe, Kyoto, Japan

becomes the estimation of the unknown shared-shape function from the observed data. A naive approach to achieving that objective is simply to take the cross-sectional mean of smoothed observations. However, repeatedly observed data often contain variation along the horizontal and vertical axes, making the estimation of the shared-shape difficult. The variations along the horizontal and vertical axes are often referred to as the amplitude and phase variation, respectively.

Curve registration, also called curve alignment in biology, is a method for separating the amplitude and phase variation from observed data when they are not of interest in data analysis. One possible objective of registration is to compare all of the curves without the uninteresting variation. In this case, registration can be considered as a pre-processing step of data analysis, in the sense that curves are transformed to enable all transformed curves to be as similar as possible, with those transformed (registered) curves then used for the subsequent data analysis. Another possible objective of registration is to estimate the unknown shared-shape function with the uninformative variation removed. In this case, the registration itself can be considered as a statistical analysis, in the sense that we are interested in a “true” shared-shape function and in estimating it from observed data. Though these objectives can be overlapping and many registration methods can achieve both objectives, we focus in this paper on the second objective for registration.

There is substantial literature on registration. Examples include landmark registration (Kneip and Gasser 1992; Kneip and Engel 1995), continuous monotone registration (Ramsay and Li 1998), registration by a self-modeling warping function (Gervini and Gasser 2004), and Fisher-Rao metric-based registration (Srivastava et al. 2011). There is also other work in this area such as Sakoe and Chiba (1978), Gervini and Gasser (2005), James (2007), Telesca and Inoue (2008), Raket et al. (2014) and Earls and Hooker (2017).

Self-modeling nonlinear regression, commonly known as the use of a shape-invariant model (SIM), is also considered to be a curve registration method (Lawton et al. 1972; Kneip and Gasser 1988), and can be used to achieve the second objective above. There is also substantial previous work on SIMs, including a method using kernel density estimation (Kneip and Engel 1995), nonparametric maximum likelihood estimation (Rønn 2001), considering a more flexible time transformation (Brumback and Lindstrom 2004), and simultaneously registering and clustering curves (Gaffney and Smyth 2005; Liu and Yang 2009; Zhang and Telesca 2014; Szczesniak et al. 2014). In addition, many studies such as Lindstrom (1995), Ke and Wang (2001), Rønn (2001) and Gaffney (2004) define a SIM model as a nonlinear mixed-effect model, for which the shared-shape function is considered as fixed effects and other subject-specific variation as random effects.

Though the use of SIMs and other registration methods allows the presence of subject-specific amplitude and phase variation in data, they assume that all relevant curves have the shared shape. In other words, they do not consider the existence of “outliers” in data. Outliers can be generally defined as observation values, whose data are inconsistent with the others (Welsh and Richardson 1997). In this paper as outliers, we consider a curve, whose shape is different from the remainder of the data, or one having considerably large amplitude or phase variation. If the data contain such outliers, those outliers will negatively affect the estimation of the

shared-shape function. Especially, Gaussian distributions, which are quite sensitive to outliers, are often used for SIMs.

Consequently, in this paper, we address this issue by proposing a new kind of SIM that is more robust to outliers. Specifically, we use a mixed-effect model type SIM and assume a t distribution for the subject variation such as amplitude, phase, and other errors including the shape, allowing the existence of outliers for each component. The characteristic feature of our method is the formulation of our SIM to enable it to be related easily to a linear mixed-effect model (LMM, e.g., Verbeke 1997), with the result that we can utilize the advantages of an LMM using a t distribution. Specifically, our SIM can provide useful information for identification and classification of three types of outliers, amplitude, phase, and other errors. In addition, using the hierarchical structure of a t distribution, we derive an efficient expectation maximization (EM)-type algorithm to estimate parameters (Dempster et al. 1977; Meng and Rubin 1993; Liu and Rubin 1994), which can be considered as a simple iteratively reweighted least squares, and also, allows to efficiently estimate the robust tuning parameter along with other parameters.

The remainder of this paper is organized as follows. In Sect. 2, we introduce our proposed model and derive the EM algorithm for estimating parameters. In addition, we show how our model can be used to separate amplitude, phase, and other errors, which can be used to identify outlier types. In Sects. 3 and 4, we use simulation and an empirical data set, respectively, to evaluate the performance of our model.

2 Robust shape-invariant model

In this section, we propose a new method, robust SIMs, for estimating the shared-shape function without being affected by outliers. Various means of formulating an SIM have been proposed (Lawton et al. 1972; Rønn 2001; Gaffney 2004). In this paper, we choose the SIM defined in Liu and Yang (2009) as the base model to robustify, because using that SIM enables us to relate our model to LMMs more easily. As a result of this relationship, we can derive some properties that are useful for parameter estimation and identification of outlier types.

2.1 Definition of robust shape-invariant model

Suppose that we observed n pairs (t_{im}, y_{im}) that t_{im} is the m th timepoint ($i = 1, \dots, n; m = 1, \dots, M_i$), and that y_{im} is the observed value of subject i at the m th timepoint. We propose the following model:

$$y_{im} = \alpha_i + \sum_{j=1}^p \eta_j (b_j(t_{im}) + \beta_i b'_j(t_{im})) + \varepsilon_{im}, \quad (1)$$

$$\alpha_i \stackrel{i.i.d.}{\sim} t(0, \sigma_1^2, h_i), \quad \beta_i \stackrel{i.i.d.}{\sim} t(0, \sigma_2^2, h_i), \quad \varepsilon_{im} \stackrel{i.i.d.}{\sim} t(0, \sigma_3^2, h_i).$$

Here, b_j ($j = 1, \dots, p$) denotes the given basis functions, and b'_j is the first derivative of b_j . Note that we can use any basis functions. In this paper, we use a B-spline basis (De Boor 1978), because it allows flexible modeling of curves. As different arguments (i.e., timepoints t_{im}) can be specified in b_j , subject-varying timepoints can be handled in our model. η_j is an unknown coefficient for the j th basis. ε_{im} indicate an observation error at the m th timepoint of subject i . α_i and β_i represent the shift variations along the horizontal and vertical axes, respectively. Similar to Liu and Yang (2009), we refer to α_i and β_i as the amplitude and phase variation, respectively. $t(\mu, \sigma^2, h_i)$ indicates t distribution, where μ , σ^2 , and h_i indicate the mean, variance, and degrees of freedom of the t distribution, respectively. Assuming that all random variables α_i , β_i and ε_{im} are t -distributed, the model (1) can be robust to outliers in α_i , β_i and ε_{im} (Lange et al. 1989; Welsh and Richardson 1997). In this paper, outliers in α_i , β_i and ε_{im} are called α_i outliers, β_i outliers and ε_{im} outliers, respectively.

In robust SIM, the parameters η and σ_r ($r = 1, 2, 3$) are estimated given (t_{im}, y_{im}) ($i = 1, \dots, n$; $m = 1, \dots, M_i$). Especially, in robust SIM, the number of basis function p must be pre-specified.

In addition, in the t distribution, the degrees of freedom h_i can be considered as a tuning parameter for robustness: with smaller h_i , the degree of downweighting of subjects having large distances increases (Lange et al. 1989). In practical situations, h_i is not allowed to be different for each subject, because it requires to estimate a large number of parameters, and thus would result in an unstable result. Therefore, in this paper, we specify h_i in three ways. The first approach [case (i)] assumes a common pre-specified $h_i = h$ for all $i = 1, \dots, n$. The second approach [case (ii)] estimates a common h from the data. In the third approach [case (iii)], we pre-specify a small number of groups of subjects (say V groups), and estimate V degrees of freedom, $\bar{h}_1, \dots, \bar{h}_V$, for each group. In this case, using estimated $\bar{h}_1, \dots, \bar{h}_V$, each h_i is specified as follows:

$$h_i = \sum_{v=1}^V g_{iv} \bar{h}_v, \quad (2)$$

where $g_{iv} = 1$ if subject i belongs to the v th group, and the other elements are 0. We call g_{iv} ($v = 1, \dots, V$; $i = 1, \dots, n$) group indicators. Setting $V = 1$ and $g_{iV} = 1$ for all i in expression (2) yields case (ii) above. The method for estimating the degrees of freedom as well as other parameters are explained in Sect. 2.3, and the performances of all cases are investigated in Sect. 3.

To understand the model (1), consider the following SIM introduced in Liu and Yang (2009):

$$\begin{aligned} y_{im} &= \alpha_i + \phi(t_{im} + \beta_i) + \varepsilon_{im} \\ &= \alpha_i + \sum_{j=1}^p \eta_j b_j(t_{im} + \beta_i) + \varepsilon_{im}, \end{aligned} \quad (3)$$

where $\phi(t) = \sum_{j=1}^p \eta_j b_j(t)$ is an unknown shared-shape function modeled using B-spline basis functions. To handle both α_i and β_i in the same space as y_{im} , similar to Liu and Yang (2009), the first order Taylor expansion is applied to each B-spline function as follows:

$$b_j(t + \beta_i) \approx b_j(t) + \beta_i b'_j(t), \quad (j = 1, \dots, p).. \tag{4}$$

Replacing $b_j(t_{im} + \beta_i)$ in (3) with the right-hand side of (4), with the t distribution assumption for all random variables, we obtain the robust SIM in (1). This approximation is reasonable, because mostly, we can assume that the shift in phase tends to be small compared to the time scale considered, see Liu and Yang (2009) for the details.

Thus, the model (1) can be interpreted as follows. We assume that the observed data y_{im} ($m = 1, \dots, M_i$) consist of an unknown shared-shape function $\phi(t)$ and that there is also other subject variation in the data, viz., the amplitude variation α_i , phase variation β_i , and other errors ϵ_{im} . This ϵ_{im} can be considered to contain not only observation errors, but also the shape difference from the shared-shape that cannot be explained by the simple shift variation α_i and β_i .

Moreover, applying the Taylor expansion has the advantage that the resulting model (1) includes all random variables α_i , β_i and ϵ_{im} in the separate terms linearly related to y_{im} . This makes it easier to relate our model to LMM, as explained in later subsections.

2.2 Relationship to the linear mixed-effect model

Assume that observations of n subjects are given, and define $(M_i \times 1)$ vector $\mathbf{y}_i = (y_{im})$ and $\boldsymbol{\epsilon}_i = (\epsilon_{im})$, and $(M_i \times p)$ matrix $\mathbf{B}_i = (b_j(t_{im}))$ and $\mathbf{D}_i = (b'_j(t_{im}))$, and $(p \times 1)$ vector $\boldsymbol{\eta} = (\eta_j)$ ($i = 1, \dots, n; m = 1, \dots, M_i; j = 1, \dots, p$). Then, robust SIM (1) can be rewritten as

$$\mathbf{y}_i = (\mathbf{B}_i + \beta_i \mathbf{D}_i) \boldsymbol{\eta} + \alpha_i \mathbf{1}_{M_i} + \boldsymbol{\epsilon}_i. \tag{5}$$

Here, $\mathbf{1}_{M_i}$ is a $M_i \times 1$ vector with all elements 1. If $\beta_i = 0$, (5) can be considered as an LMM using the t distribution (Welsh and Richardson 1997; Pinheiro et al. 2001; Song et al. 2007), with $\boldsymbol{\eta}$ and α_i as fixed and random effects. However, owing to the existence of β_i , robust SIM is not a direct extension of LMM, in the sense that there is the cross term of fixed effect $\boldsymbol{\eta}$ and random effect β_i . Nevertheless, as a result of the Taylor expansion, we can still utilize the key property of LMM: we can derive an efficient EM-type algorithm and we can easily decompose the subject variation, α_i , β_i and ϵ_{im} , into separate components, which can be used to identify outlier types. These will be explained in detail in Sects. 2.3 and 2.4, respectively.

2.3 Parameter estimation

In robust SIM, parameters $\theta = (\boldsymbol{\eta}, \sigma_1^2, \sigma_2^2, \sigma_3^2, \mathbf{h})$, where $\mathbf{h} = (h_1, \dots, h_n)$ are estimated given \mathbf{y}_i ($i = 1, \dots, n$). From (1), we can say

$$\mathbf{y}_i \sim t(\mathbf{B}_i \boldsymbol{\eta}, \mathbf{V}_i, h_i)$$

$$\text{where } \mathbf{V}_i = \sigma_1^2 \mathbf{1}_{M_i} \mathbf{1}_{M_i}^T + \sigma_2^2 \mathbf{D}_i \boldsymbol{\eta} \boldsymbol{\eta}^T \mathbf{D}_i^T + \sigma_3^2 \mathbf{I}_{M_i}.$$

That is, \mathbf{y}_i has the following probability density function $f(\mathbf{y}_i)$:

$$f(\mathbf{y}_i) = \frac{\Gamma\left(\frac{h_i + M_i}{2}\right) |\mathbf{V}_i|^{-1/2}}{(\pi h_i)^{M_i/2} \Gamma(h_i/2) (1 + d_i^2/h_i)^{(h_i + M_i)/2}}$$

$$\text{where } \Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt$$

$$d_i^2 = (\mathbf{y}_i - \mathbf{B}_i \boldsymbol{\eta})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{B}_i \boldsymbol{\eta}).$$

The corresponding log-likelihood function, an objective function for estimating parameters in robust SIM is defined as

$$\log L(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^n \left[-\frac{M_i}{2} \log \pi h_i - \log \Gamma\left(\frac{h_i}{2}\right) + \log \Gamma\left(\frac{h_i + M_i}{2}\right) - \frac{1}{2} \log |\mathbf{V}_i| + \frac{h_i + M_i}{2} \log h_i - \frac{h_i + M_i}{2} \log(h_i + d_i^2) \right]. \quad (6)$$

Since this objective function cannot be solved analytically, we use the EM algorithm (Dempster et al. 1977), an iterative algorithm for obtaining maximum likelihood estimators. Specifically, the EM algorithm consists of two steps:

E-step given the k th current estimator $\boldsymbol{\theta}^{(k)}$, the auxiliary function of (6) which we call the Q -function, denoted by $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})$ is calculated.

M-step The parameters $\boldsymbol{\theta}^{(k)}$ are updated by maximizing the Q -function.

When the degrees of freedom are unknown, the EM algorithm converges very slowly (e.g., Liu and Rubin 1994). Therefore, we suggest adopting the two extensions of EM algorithm shown in Liu and Rubin (1995), namely, a multi-cycle version of the ECM algorithm, and the ECME algorithm. The ECM algorithm (Meng and Rubin 1993) replaces the M-step of the EM algorithms by several computationally simpler conditional maximization (CM) steps, whereas multi-cycle ECM computes the E-step before each CM-step. The ECME algorithm (Liu and Rubin 1994) maximizes either the

Q -function or the actual log-likelihood function in each CM-step, whereas the ECM algorithm only maximizes the Q -function in the CM-step. When estimating the degrees of freedom in t distributions, the log-likelihood is maximized instead of the Q -function. Both the multi-cycle ECM and ECME algorithms can be computationally faster than EM, and to estimate parameters in t distribution, ECME is reportedly faster than ECM (Liu and Rubin 1995), because (unlike ECM) it allows a closed form of the log-likelihood function to be maximized.

To estimate parameters using the ECM or ECME algorithm, we need to derive Q -function. As the Q -function, the conditional expectation of the log-likelihood function for complete data is used. “The complete data” indicates that augmented data of the observed data and “missing data” that are not actually observed but assumed to be observed. In our case, y_i are observed data, while α_i and β_i are considered as missing data. However, since all data are t -distributed, the resulting log-likelihood would still be complicated. Therefore, we use the gamma-normal hierarchical structure of the t distribution to derive the Q -function (e.g., Lange et al. 1989; Liu and Rubin 1995).

Specifically, letting $\boldsymbol{\gamma}_i = (\alpha_i, \beta_i)^T$ and combining all distribution assumptions for y_i , and $\boldsymbol{\gamma}_i$ in (1), it can be shown that

$$y_i | \boldsymbol{\gamma}_i, \tau_i \stackrel{i.i.d.}{\sim} N\left(\mathbf{B}_i + \beta_i \mathbf{D}_i \boldsymbol{\eta} + \alpha_i \mathbf{1}_{M_i}, \frac{\sigma_3^2}{\tau_i} \mathbf{I}_{M_i}\right)$$

$$\boldsymbol{\gamma}_i | \tau_i \stackrel{i.i.d.}{\sim} N\left(\mathbf{0}, \frac{1}{\tau_i} \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right), \quad \tau_i \stackrel{i.i.d.}{\sim} \text{gamma}\left(\frac{h}{2}, \frac{h}{2}\right),$$

where $\boldsymbol{\gamma}_i | \tau_i$ is independent of $\epsilon_i | \tau_i$. Here, $X \sim \text{gamma}(a, b)$ indicates that the random variable X is distributed as a gamma distribution with parameters a, b , defined by the density function for $a > 0, b > 0$ and $x > 0$:

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx).$$

In addition, this decomposition indicates that introducing a latent random variable τ_i results in y_i and $\boldsymbol{\gamma}_i$ having a Gaussian distribution. Then, $(y_i, \boldsymbol{\gamma}_i, \tau_i)$ are called the complete data. Using this hierarchical structure, the joint density for the complete data can be decomposed as

$$f(y_i, \boldsymbol{\gamma}_i, \tau_i) = f(y_i | \boldsymbol{\gamma}_i, \tau_i) f(\boldsymbol{\gamma}_i | \tau_i) f(\tau_i).$$

Consequently, the conditional expectation of the log-likelihood for the complete data will be

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)})$$

$$= \sum_{i=1}^n E_{\boldsymbol{\theta}^{(k)}}[\log L(\boldsymbol{\theta}; y_i, \alpha_i, \beta_i, \tau_i) | y_i]$$

$$= \sum_{i=1}^n E_{\boldsymbol{\theta}^{(k)}}[\log L(\boldsymbol{\eta}, \sigma_3^2; y_i, \alpha_i, \beta_i, \tau_i)$$

$$+ \log L(\sigma_1^2; \alpha_i, \tau_i) + \log L(\sigma_2^2; \beta_i, \tau_i) + \log L(h_i; \tau_i) | y_i] + \text{const.}$$
(7)

where

$$2 \log L(\boldsymbol{\eta}, \sigma_3^2; \mathbf{y}_i, \alpha_i, \beta_i, \tau_i) = -M_i \log \sigma_3^2 - \frac{\tau_i}{\sigma_3^2} (\mathbf{y}_i - \alpha_i \mathbf{1}_{M_i} - \mathbf{X}_i \boldsymbol{\eta})^\top (\mathbf{y}_i - \alpha_i \mathbf{1}_{M_i} - \mathbf{X}_i \boldsymbol{\eta}) \tag{8}$$

$$2 \log L(\sigma_1^2; \alpha_i, \tau_i) = -\log \sigma_1^2 - \frac{\tau_i \alpha_i^2}{\sigma_1^2} \tag{9}$$

$$2 \log L(\sigma_2^2; \beta_i, \tau_i) = -\log \sigma_2^2 - \frac{\tau_i \beta_i^2}{\sigma_2^2} \tag{10}$$

$$2 \log L(h_i; \tau_i) = \sum_{i=1}^n \left\{ \frac{h_i}{2} \left[\log \left(\frac{h_i}{2} \right) + \log \tau_i - \tau_i \right] - \log \tau_i - \log \left[\Gamma \left(\frac{h_i}{2} \right) \right] \right\} \tag{11}$$

and $\mathbf{X}_i = \mathbf{B}_i + \beta_i \mathbf{D}_i$.

To simplify the notation, let $\mathbf{u}_{1i} = \mathbf{y}_i - \alpha_i \mathbf{1}_{M_i} - \mathbf{X}_i \boldsymbol{\eta}$, $\mathbf{u}_{2i} = \alpha_i$, $\mathbf{u}_{3i} = \beta_i$, $\mathbf{Z}_{1i} = \mathbf{I}_{M_i}$, $\mathbf{Z}_{2i} = \mathbf{1}_{M_i}$, $\mathbf{Z}_{3i} = \mathbf{D}_i \boldsymbol{\eta}$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_1 = (\boldsymbol{\eta}, \sigma_1^2, \sigma_2^2, \sigma_3^2)$ and $\boldsymbol{\theta}_2 = (\mathbf{h})$. Using these notations and the Q function in (7), we now describe the multi-cycle version ECM and ECME algorithms. Note that as the E-steps and CM steps for updating $\boldsymbol{\theta}_1$ are identical in both algorithms, the CM steps of the multi-cycle ECM and ECME algorithms differ only in their $\boldsymbol{\theta}_2$ updates. In addition, in the following algorithm, the degrees of freedom are assumed to be estimated, while the extension to the case, where h_i is fixed [i.e., case (i)] is straightforward.

Initialization: Given \mathbf{y}_i and $\mathbf{t}_i = (t_{i1}, \dots, t_{iM_i})$ ($i = 1, \dots, n$), determine basis functions b_j ($j = 1, \dots, p$) and p . When estimating \mathbf{h} , the group indicator g_{i1}, \dots, g_{iV} should also be given. Then randomly generate initial values for parameters $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\eta}^{(0)}, \sigma_1^{(0)2}, \sigma_2^{(0)2}, \sigma_3^{(0)2}, \mathbf{h}^{(0)})$, and set the number of iteration to $k = 0$ and convergence criterion ϵ .

E-step 1: Given the current estimate $\boldsymbol{\theta}^{(k)}$ at the k th iteration, the Q function is obtained by computing

$$\hat{\tau}_i^{(k)} = E_{\boldsymbol{\theta}^{(k)}}[\tau_i | \mathbf{y}_i] = \frac{h_i^{(k)} + M_i}{h_i^{(k)} + d_i^{(k)2}} \tag{12}$$

where $d_i^{(k)2} = (\mathbf{y}_i - \mathbf{B}_i \boldsymbol{\eta}^{(k)})^\top \mathbf{V}_i^{(k)-1} (\mathbf{y}_i - \mathbf{B}_i \boldsymbol{\eta}^{(k)})$

$$V_i^{(k)} = \sigma_1^{(k)2} \mathbf{1}_{M_i} \mathbf{1}_{M_i}^T + \sigma_2^{(k)2} \mathbf{D}_i \boldsymbol{\eta}^{(k)} \boldsymbol{\eta}^{(k)T} \mathbf{D}_i^T + \sigma_3^{(k)2} \mathbf{I}_{M_i} \tag{13}$$

and for $r = 1, 2, 3$

$$\begin{aligned} E_{\boldsymbol{\theta}^{(k)}}[\mathbf{u}_{ri} | \mathbf{y}_i, \tau_i] &= \sigma_r^{(k)2} \mathbf{Z}_{ri}^T V_i^{(k)-1} (\mathbf{y}_i - \mathbf{B}_i \boldsymbol{\eta}^{(k)}) \\ V_{\boldsymbol{\theta}^{(k)}}[\mathbf{u}_{ri} | \mathbf{y}_i, \tau_i] &= \sigma_r^{(k)2} \mathbf{I}_{M_i} - \sigma_r^{(k)4} \mathbf{Z}_i^T V_i^{(k)-1} \mathbf{Z}_i \\ E_{\boldsymbol{\theta}^{(k)}}[\mathbf{u}_{ri}^T \mathbf{u}_{ri} | \mathbf{y}_i, \tau_i] &= \text{tr}(V_{\boldsymbol{\theta}^{(k)}}[\mathbf{u}_{ri} | \mathbf{y}_i, \tau_i]) + E_{\boldsymbol{\theta}^{(k)}}[\mathbf{u}_{ri} | \mathbf{y}_i, \tau_i]^T E_{\boldsymbol{\theta}^{(k)}}[\mathbf{u}_{ri} | \mathbf{y}_i, \tau_i]. \end{aligned} \tag{14}$$

CM-step 1: Fix $\boldsymbol{\theta}_2^{(k)}$, and update $\boldsymbol{\theta}_1^{(k+1)}$ by maximizing conditional expectation of (8),..., (10), which leads to

$$\begin{aligned} \boldsymbol{\eta}^{(k+1)} &= \left(\frac{\hat{\tau}_i^{(k)}}{\sigma_3^{(k)2}} E_{\boldsymbol{\theta}^{(k)}}[\mathbf{X}_i^T \mathbf{X}_i \boldsymbol{\eta} | \mathbf{y}_i, \tau_i] \right)^{-1} \frac{\hat{\tau}_i^{(k)}}{\sigma_3^{(k)2}} E_{\boldsymbol{\theta}^{(k)}}[\mathbf{X}_i^T (\mathbf{y}_i - \alpha_i \mathbf{1}_{M_i}) | \mathbf{y}_i, \tau_i] \\ \sigma_r^{2(k+1)} &= \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}^{(k)}}[\mathbf{u}_{ri}^T \mathbf{u}_{ri} | \mathbf{y}_i, \tau_i], \quad (r = 1, 2, 3). \end{aligned} \tag{15}$$

E-step 2: Given the current estimate, define $\boldsymbol{\theta}^{(k+1/2)} = (\boldsymbol{\eta}^{(k+1)}, \sigma_1^{(k+1)2}, \sigma_2^{(k+1)2}, \sigma_3^{(k+1)2}, \mathbf{h}^{(k)})$. Then, calculating (12) using $\boldsymbol{\theta}^{(k+1/2)}$ instead of $\boldsymbol{\theta}^{(k)}$.

CM-step 2: Fix $\boldsymbol{\theta}_1^{(k+1)}$, and update $\boldsymbol{\theta}_2^{(k+1)} = \mathbf{h}^{(k+1)}$ by maximizing either function below:

(multi-cycle version of ECM) Obtain $\mathbf{h}^{(k+1)}$ by maximizing conditional expectation of (11), that is

$$\begin{aligned} \bar{h}_v^{(k+1)} &= \arg \max_h \sum_{i=1}^n g_{iv} \left\{ \frac{h}{2} \left(\log \frac{h}{2} + E_{\boldsymbol{\theta}^{(k+1/2)}}[\log \tau_i | \mathbf{y}_i] - \hat{\tau}_i^{(k+1/2)} \right) \right. \\ &\quad \left. - \log \Gamma\left(\frac{h}{2}\right) \right\}, \quad (v = 1, \dots, V). \end{aligned}$$

(ECME) Obtain $\mathbf{h}^{(k+1)}$ by maximizing (6), which leads to

$$\begin{aligned} \bar{h}_v^{(k+1)} &= \arg \max_h \sum_{i=1}^n g_{iv} \left\{ -\log \Gamma\left(\frac{h}{2}\right) + \log \Gamma\left(\frac{h + M_i}{2}\right) + \frac{h}{2} \log h \right. \\ &\quad \left. - \frac{h + M_i}{2} \log \left(h + d_i^{(k+1/2)2} \right) \right\}. \end{aligned}$$

Convergence test Compute $\log L(\theta^{(k+1)})$, the value of the objective function (6) using updated parameters and, for $k > 1$, if $\log L(\theta^{(k+1)}) - \log L(\theta^{(k)}) < \epsilon$, terminate; otherwise, let $k = k + 1$ and return to E-step 1.

Note that CM-step 2 is simpler in ECME than in multi-cycle ECM, because $E_{\theta^{(k+1/2)}}[\log \tau_i | y_i]$ in CM-step 2 of multi-cycle ECM does not have closed form. In addition, “Appendix” provides a more detailed expression of (15).

From these updating formulas, this algorithm can be considered as iteratively reweighted least squares (Lange et al. 1989), and $\hat{\tau}_i$ can be interpreted as a weight for the subject i on estimation. This weight is determined by the Maharanobis distance d_i^2 in (13); for a subject, whose distance is larger, that is, whose observations are more distant from the mean, the weight of the subject on estimation decreases.

2.4 Decomposition of amplitude, phase and other errors

Though the principal objective of robust SIM is to estimate the shared-shape function without being affected by outliers, we can also use our robust SIM to identify subjects having outlying observations: subjects having very large Maharanobis distances d_i calculated using (13) compared to others could be considered as outliers.

Furthermore, using the relationship with LMM explained in Sect. 2.2, we can also identify which outlier type, α_i , β_i or ϵ_i the subject has. Specifically, as we allow a randomness in the α_i , β_i and ϵ_i component, a subject identified as having outliers can include outliers in either one component or several. Then, there could be a situation in which we want to identify subjects having only ϵ_i outliers, that is, subjects whose shape is inconsistent with others, or only α_i outliers, that is, those having amplitude outliers. In that case, decomposing d_i , where η and α_i , β_i are replaced with those current estimates, enables us to see the distance from the mean for each component separately.

Specifically, letting the current estimates as $\hat{\alpha}_i = E_{\theta^{(k)}}[\alpha_i | y_i, \tau_i]$, $\hat{\beta}_i = E_{\theta^{(k)}}[\beta_i | y_i, \tau_i]$ and $\hat{\epsilon}_i = E_{\theta^{(k)}}[\epsilon_i | y_i, \tau_i]$, the Maharanobis distance can be decomposed as follows (Pinheiro et al. 2001; Matos et al. 2013):

$$\begin{aligned} \hat{d}_i^2 &= (y_i - B_i \eta)^T \hat{V}_i^{-1} (y_i - B_i \eta) \\ &= \frac{\hat{\alpha}_i^2}{\sigma_1^2} + \frac{\hat{\beta}_i^2}{\sigma_2^2} + \frac{\hat{\epsilon}_i^T \hat{\epsilon}_i}{\sigma_3^2} \\ &= d_{i1}^2 + d_{i2}^2 + d_{i3}^2. \end{aligned} \tag{16}$$

This decomposition indicates that the subject’s difference from the shared-shape function can be separated into three components, amplitude, phase, and other errors. Through this decomposition, we can check how distant each component is from the

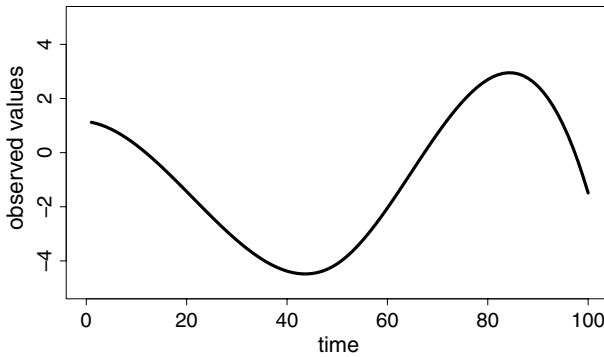


Fig. 1 Artificial true shared-shape function in the simulation study, which is generated using $p = 5$ B-spline basis functions

mean separately, and consequently, can identify subjects having outliers in a particular component.

Note that this type of decomposition had originally been performed in the LMM framework (Pinheiro et al. 2001), in which the distance is decomposed into the random effect γ_i and other error ϵ_i components. However, in this paper, this idea is extended to the further decomposition of random effects into amplitude and phase components, enabling us to handle the shift variation along the horizontal and vertical axes separately.

3 Simulation study

In this section, we describe a simulation we conducted to evaluate the performance of robust SIM compared to the existing methods.

3.1 Data generation

The model in (3) was used to generate artificial data. Specifically, we first generated simple “true shared-shape data”, as shown in Fig. 1, using $p = 5$ and order 4 B-spline basis functions with an interior knot between 0 and T (where T was set to 100). Next, we decided the timepoints t_{im} ($m = 1, \dots, M_i; i = 1, \dots, n$) for each $i = 1, \dots, n$, by randomly selecting M_i from $\{5, \dots, 15\}$, and t_{i1}, \dots, t_{iM_i} from $\{1, \dots, T\}$. Then, we added subject variation $\gamma_i = (\alpha_i, \beta_i)$ and ϵ_i , using the following mixture Gaussian distribution:

$$\begin{aligned} \gamma_i &\sim (1 - c)N(\mathbf{0}, \Phi) + cN(\mathbf{0}, s\Phi) \quad \text{where } \gamma_i = (\alpha_i, \beta_i), \Phi = \text{diag}(\sigma_1^2, \sigma_2^2) \\ \epsilon_i &\sim (1 - c)N(\mathbf{0}, \sigma_3^2 \mathbf{I}_{M_i}) + cN(\mathbf{0}, s\sigma_3^2 \mathbf{I}_{M_i}), \quad (i = 1, \dots, n), \end{aligned}$$

where c denotes the proportion of subjects having outlying observations, and s denotes the scale of the variance in these subjects. That is, a proportion c of the subjects were assumed to follow a distribution with a large variance, so would likely include outlying observations in α_i , β_i and ϵ_i , that are distant from the mean. In addition, we fixed variances to $\sigma_1^2 = 5$, $\sigma_2^2 = 10$ and $\sigma_3^2 = 5$.

3.2 Setting and evaluation

In this simulation, we considered a full factorial design with $c = 0.2, 0.5$, $s = 5, 10$ and $n = 30, 60$. We also considered a “non-noisy situation”, with no outlying observations (i.e., $c = 0$ and $s = 1$). For each combination of all $2 \times 2 \times 2 + 2$ cases (+2 is for non-noisy situation in $n = 30$ and 60), we randomly generated 50 different data sets. For each data set, we applied the method and evaluated.

To estimate a robust SIM, we need to pre-specify the applied basis functions and the way of handling the degrees of freedom h_i . As basic functions, we used the $p = 5$ B-splines shown in Fig. 1. To estimate h_i , we considered all cases (i), (ii), and (iii) mentioned in Sect. 2.1. Specifically, in case (i), we fixed $h_i = h$ for all $i = 1, \dots, n$ varying h as 1, 5, 10, or 20. In case (ii), h was estimated by a data-driven approach. In case (iii), where h_i is estimated in groups, we allocated a proportion $1 - c$ of the subjects generated from a distribution with a smaller variance to group 1, and the remaining c proportion of the subjects to group 2. Note that as no outlying observations are generated in $s = 1$ and $c = 0$ situation, robust SIM (ii) and (iii) are equivalent for this situation, so only (ii) is applied.

To evaluate the performance of our robust SIM, we examine the accuracy of estimation for the true shared-shape function and compare it with several existing registration methods. In this simulation, we additionally applied following methods: (iv) SIM with a Gaussian assumption, that is, the model (1) in which a Gaussian distribution is assumed for all α_i , β_i and ϵ_{im} (Gaussian SIM), (v) landmark registration (Kneip and Gasser 1992), (vi) continuous monotone registration (Lange et al. 1989), (vii) registration by a self-modeling warping function (Gervini and Gasser 2004), and (viii) Fisher–Rao metric-based registration (Srivastava et al. 2011). Methods (v) and (vi) were applied using the R package “fda”, method (vii) was implemented in Matlab code obtained from the author’s webpage¹, and method (viii) was applied using the R package “fdasrvf”. To handle the subject-varying timepoints in these software packages, we first smoothed the observed data using the true B-spline basis functions, and applied the values of the smoothed function at the $1, \dots, T$ timepoints in each method.

Landmarks must be specified in advance to apply (v). Here, two curving points in Fig. 1 are specified as landmarks, obtained as follows. Each observed datum y_{im} is smoothed using the same B-spline basis functions, as shown in Fig. 1, and from the resulting smoothed functions f_i ($i = 1, \dots, n$), the first landmark is set to be the timepoint $\ell_{i1} = \arg \min_{t \in \{1, \dots, T\}} \{f'_i(t) = 0 \wedge f''_i(t) > 0\}$, and the second is the timepoint

¹ <https://people.uwm.edu/gervini/>

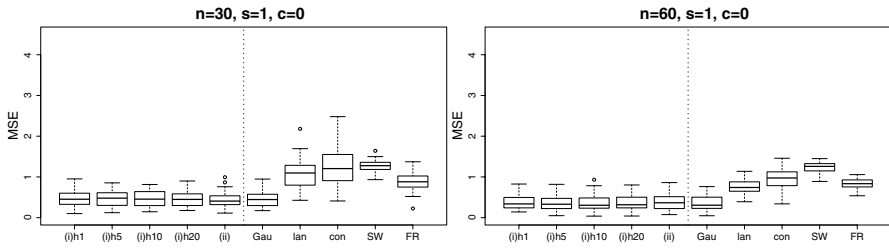


Fig. 2 Boxplot of MSE results of 50 data sets for non-noisy situation ($c = 0, s = 1$). From the left axis, labels indicate robust SIM with $h = 1, 5, 10, 20$ in (i), robust SIM with (ii) case, Gaussian SIM (Gau), landmark registration (lan), continuous monotone registration (con), registration by a self-modeling warping function (SW), Fisher–Rao-based registration (FR)

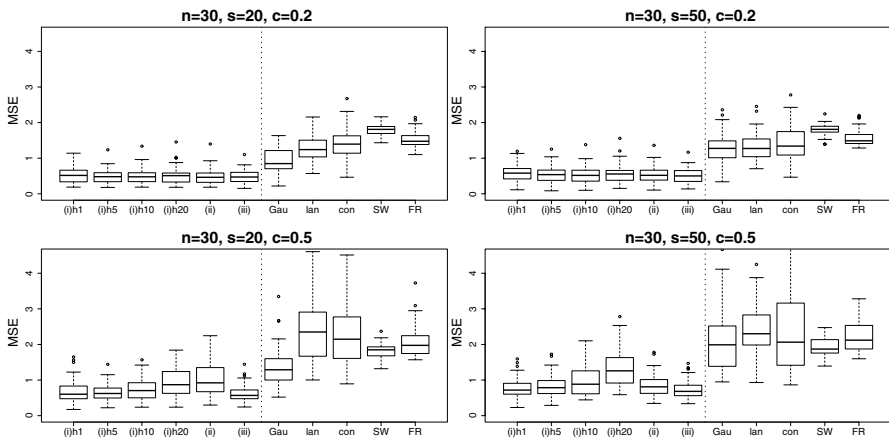


Fig. 3 Boxplot of MSE results of 50 data sets for $n = 30$. The 6th label from the left, (iii), indicates robust SIM with (iii) case

$\ell_{i2} = \arg \min_{t \in \{1, \dots, T\}} \{f'(t) = 0 \wedge f''(t) < 0\}$. Note that we assume that $\ell_{i1} < \ell_{i2}$ for all i .

For (vi) and (vii), we must specify a “target function” to which all functions are aligned in registration. Here, we use a median function as a target. That is, we take the median of coefficients of smoothed data f_1, \dots, f_n , and use that median as the coefficient for the B-spline basis. For the hyperparameters in (vii), we specify the number of components to $q_1 = 2$ and the number of basis functions as $q_2 = 6$, because Gervini and Gasser (2004) recommends $q_2 = 3q_1$. for methods that require initial values for estimation, we used 50 initial random starts.

To evaluate the performance, we calculate the mean squared error (MSE). That is, letting $\hat{\phi}$ be the estimated shared-shape function, MSE is defined as $[\sum_{m=1}^T (\hat{\phi}(t_m) - \phi(t_m))^2 / T]^{1/2}$, where $t_m = 1, 2, \dots, T$. The $\hat{\phi}$ in each method is constructed as follows. In (i) and (ii), we use $\hat{\phi}(t_m) = \sum_{j=1}^p \hat{\eta}_j B_j(t_m)$ ($m = 1, \dots, T$), where $\hat{\eta}_j$ ($j = 1, \dots, p$) are estimated coefficients. For other methods, we use

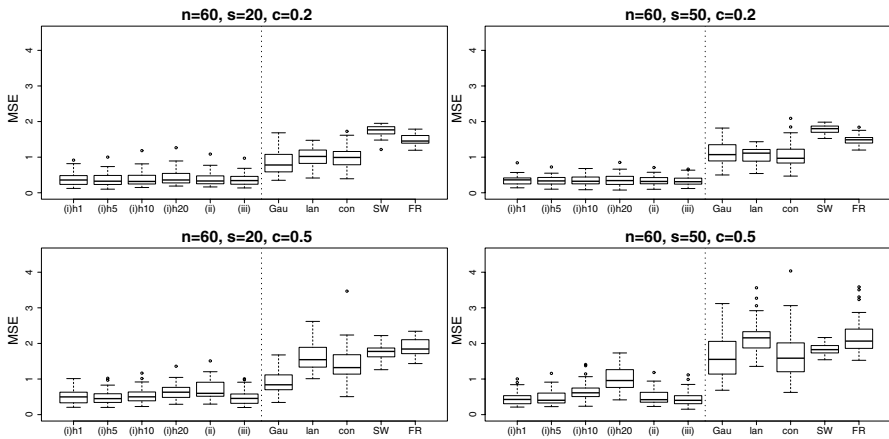


Fig. 4 Boxplot of MSE results of 50 data sets for $n = 60$

Table 1 Estimated degrees of freedom by robust SIM with (ii) and (iii) case

n	c	s	(ii)		(iii) group 1		(iii) group 2	
			Mean	SD	Mean	SD	Mean	SD
30	0	1	761.41	54.60	–	–	–	–
	0.2	20	13.63	2.56	127.83	19.33	3.38	0.55
		50	8.66	0.97	132.52	18.03	2.43	0.37
		0.5	20	47.10	33.39	130.02	17.85	4.24
	50	5.91	0.81	125.45	19.50	2.60	0.31	
60	0	1	1337.75	420.25	–	–	–	–
	0.2	20	13.60	1.84	164.99	24.26	3.44	0.41
		50	8.56	0.79	164.49	26.00	2.42	0.26
		0.5	20	27.32	22.54	161.48	13.13	4.06
	50	5.81	0.49	155.80	14.95	2.59	0.20	

$\hat{\phi}(t_m) = \text{median}(\hat{f}_1(t_m), \dots, \hat{f}_n(t_m))$ ($m = 1, \dots, T$), where \hat{f}_i ($i = 1, \dots, n$) are registered functions. Here, we take the median of all registered functions rather than mean, because if we were to take the mean to define $\hat{\phi}$, the resulting MSE of some methods would be too large to compare with the other methods, possibly because there might be some severely mis-registered functions due to outliers.

3.3 Results

Figures 2 (left) and 3 show the results for $n = 30$ in non-noisy and noisy situation, respectively. While the MSE results of robust and Gaussian SIM are similar in non-noisy situation, the differences between them are apparent in noisy situation, especially for $c = 0.5$. This indicates that SIM assuming a Gaussian distribution is sensitive to outliers.

In addition, the performance of existing methods worsens as c increases, while that of robust SIM appears not to be affected much by an increase in c . Comparing the situation between $s = 5$ and 10, robust SIM appears to perform similarly for $s = 10$ than for $s = 5$. This indicates that robust SIM can be robust especially for distant outliers.

Though the general tendencies of MSE results by (i), (ii), and (iii) in robust SIM are similar, it seems that robust SIM with a smaller h in (i) or (iii) have better results in accuracy. Specifically, for $h = 10, 20$ in (i) and (ii), the increase in c does affect the accuracy, while with $h = 1, 5$ and (iii), the accuracy seems stable for all situations.

Figures 2 (right) and 4 show the results for $n = 60$. For all methods the variance of MSE is reduced and the median of MSE is improved, compared to $n = 30$. However, for $c = 0.5$, the variance of MSE is increased for $h = 10, 20$, (ii), and existing methods (iv), ..., (viii).

Table 1 shows the degrees of freedom estimated by the robust SIM in case (ii) and (iii). h_i estimated by (ii) and (iii) for group 2 (having outlying observations) tends to be smaller than the one by (iii) for group 1. This indicates that the algorithm efficiently provides robustness to the model. In particular, in (iii) case, h_i is adjusted during each iteration, thereby conferring robustness to a subset of the data rather than the whole data set.

In summary, robust SIM performs better than the other methods in the sense that it shows a lower variance of MSE for all cases, and a lower median of MSE especially for $c = 0.5$. This indicates that estimating the shared-shape function using robust SIM is more stable than for the other methods, and consequently, we can expect a better accuracy for the estimation. Among cases (i), (ii) and (iii), case (iii) and $h = 1, 5$ in (i) apparently performed well in all scenarios. This indicates that when we can specify groups for h_i , that is, specify g_{iv} ($v = 1, \dots, V$), approach (iii) is a reasonable choice, because it yields a low and stable MSE result even without pre-specifying the h_i .

4 Application

4.1 Data and setting

In this section, we use an empirical electrocardiogram (ECG) data set to illustrate the performance of robust SIM. The ECG data set was obtained from the UCR Time Series Classification and Clustering Archive² which was analyzed in Olszewski (2001). The ECG database contains 200 data sets (corresponding to subjects in this paper), each consisting of measurements at 96 timepoints during one heartbeat. Each measurement was recorded by an electrode placed on the body. The first 100 subjects were defined as the training data, while the latter 100 were the test data. Each of the 200 data sets were labeled as “normal” or “abnormal.”

The purpose of this data analysis was twofold; first, it estimated the shared-shape functions from both the training and test data; second, it identified the outlier types

² http://www.cs.ucr.edu/~eamonn/time_series_data/

among the subjects in test data, using the estimated shared-shape function as a reference curve. Therefore, to apply the data in this analysis, we picked up only “normal” subjects from the training data, and both “normal” and “abnormal” subjects from the test data. To reduce the computational burden, we used the data of 40 subjects, and 80% from the training data, and 20% from the test data with an equal number of normal and abnormal subjects. All subjects were randomly chosen.

In addition, we estimated the degrees of freedom in each group [corresponding to case (iii) in Sect. 2.1]. We pre-specified g_{iv} ($v = 1, \dots, V; i = 1, \dots, n$) with $V = 2$, and set the subjects from the training and test data as groups 1 and 2, respectively. Note that we did not define groups by “normal” and “abnormal”, because the label of the test data was assumed to be unknown in this analysis.

Robust SIM requires prespecification of p . To determine p , we use Akaike Information Criteria (AIC) as Liu and Yang (2009) suggested, which is defined as $AIC = -2 \log \hat{L} + 2K$, where $\log \hat{L}$ is the maximized log likelihood and K is the total number of free parameters, which is $p + 3 + V$. $p = 8, 9, 10, 11$ are used as the candidate values. Same as the simulation study, we use order 4 and the $p - 4$ equidistant interior knots. Finally, we ran the algorithm with ten different initial values.

4.2 Result

Table 2 shows AIC values for different p . This result indicates that robust SIM with $p = 10$ is the best fitted model according to AIC. Figure 5 shows the shared-shape function estimated by our robust SIM. The degrees of freedom in groups 1 and 2 were estimated as 1258.06 and 74.66, respectively, indicating that group 2 may include outlying observations.

Furthermore, using the decomposition in Sect. 2.4, we identify outlier types for each subject. Figure 6 shows each separated distance of amplitude, phase, and other errors, d_{1i}^2, d_{2i}^2 and d_{3i}^2 in (16). We can see that the subject $i = 34, 36, 40$ appear to have large d_{1i}^2 , while several subjects such as $i = 37, 38, 39$ have large d_{13}^2 . At the same time, Figure 7 depicted the observations of subjects having the three largest d_{12}^2 and d_{13}^2 values using solid and dashed lines. This indicates that the shapes of observations having large d_{12}^2 with solid lines in Fig. 7 (left) appear similar to the shared shape, though these have the large phase difference from the shared one. On the other hand, the shape of observations having large d_{13}^2 in Figure 7 (right) with a dashed line appears to be inconsistent with that of the other observations. Furthermore, those with solid lines in right figure have shapes that are more similar to the shared one than the dashed line; however, the difference in shape from the shared one appears to be more than a simple shift variation.

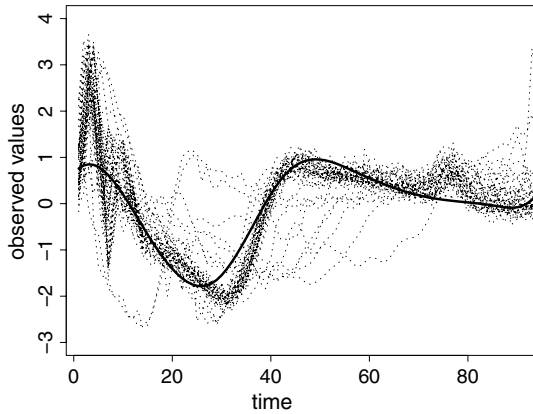
5 Discussion

In this paper, we developed robust SIM using the t distribution. Our main objective was to estimate the shared unknown shape function without being affected by outliers. In addition, we showed that robust SIM can provide useful information for identifying outlier types.

Table 2 Akaike information criterion (AIC) results for different p

p	8	9	10	11
AIC	-443.22	-634.91	-811.03	-491.63

Fig. 5 ECG data. Dot lines indicate data of each subject, while solid line indicates the estimated shared-shape function for both train and test data, using $p = 10$ basis functions



Specifically, in our robust SIM, by applying the first order Taylor expansion on SIM, we separated α_i , β_i and ϵ_i , making it easier to relate our model to LMM using the t distribution to utilize the properties of that model: we can decompose the distance into separate components, which can be used to identify outlier types. In addition, utilizing the hierarchical structure of t distribution, we could obtain efficient multi-cycle ECM and ECME algorithms which can be considered as a iteratively reweighted least squares and allow to estimate the degrees of freedom.

In the simulation, we showed that the estimation of our robust SIM is more stable than the other existing methods, especially when $c = 0.5$. In an empirical data application, we illustrated the performance of our robust SIM with case (iii), showing how the decomposition of distance can be utilized to identify outlier types.

Finally, we mention three possible directions for future work. First, our simulation study suggests that estimating h_i by group improves the accuracy of the result. However, pre-specifying the groups is a complicated task in many practical situations. One can heuristically estimate g_{iv} from the data by a heuristic algorithm such as the k -means algorithm (e.g., MacQueen 1967). Specifically, when h_i is fixed, CM-step obtains v^* which contributes the increase of objective function the most from $v = 1, \dots, V$, and sets $g_{iv^*}^{(k+1)} = 1$ and all other elements to zero, and repeat the iteration. However, this approach increases the computational burden and slows the convergence. Therefore, an efficient and practical means of calculating h_i is highly desired.

Second, we might wish to vary h_i for the various components α_i , β_i and ϵ_i . For example, in the ECG data, a larger h_i for α_i than β_i and ϵ_i might be beneficial.

Finally, robust SIM explicitly parameterizes only the shift difference along the horizontal and vertical axes, not the scale difference. However, there could be a

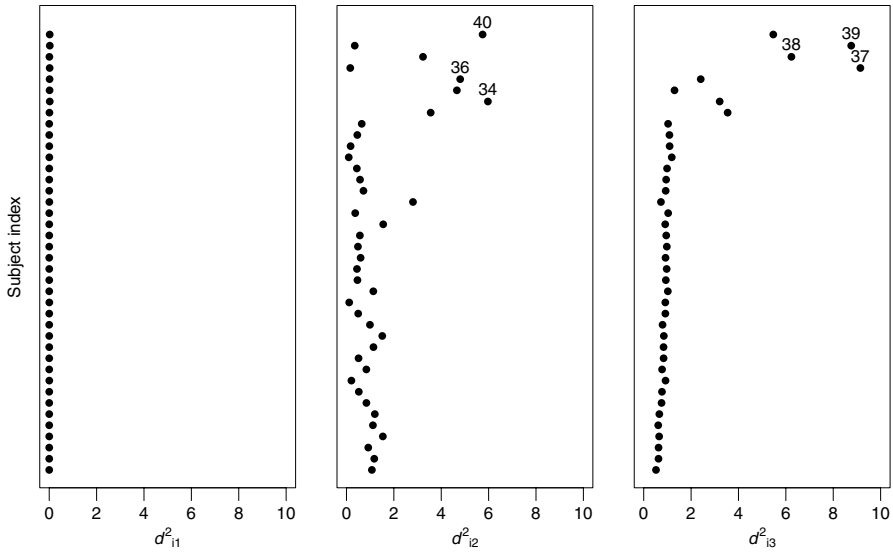


Fig. 6 Distance in each component d^2_{i1} , d^2_{i2} and d^2_{i3} , calculated according to (16). The number in middle and right plots indicates the index of subject having three largest d^2_{i2} , d^2_{i3} , respectively

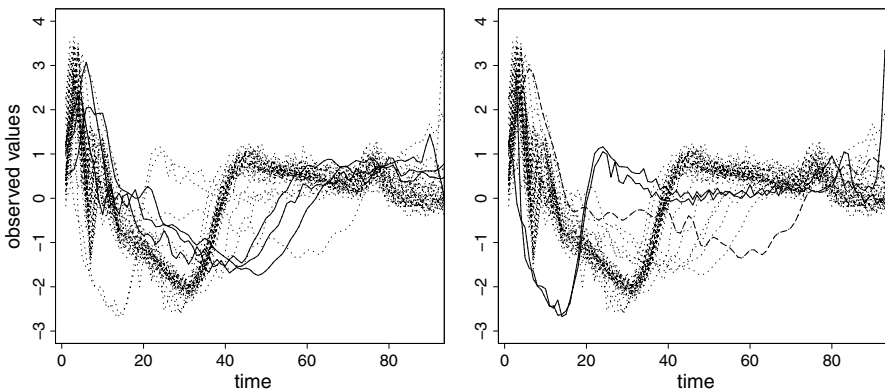


Fig. 7 ECG data with the subjects having the three largest distance of β_i (left) and ϵ_i (right) in Fig. 6. Specifically, while left and right figures show the same data set, observations having large distance are depicted using as solid line ($i = 34, 36, 40$ in left, and $i = 37, 39$ in right figures) and dashed line ($i = 38$ in right)

situation, where this is not enough. For example, we could say that the dashed line in Fig. 7 (left) might exhibit the scale difference from the shared shape, and therefore, it could be beneficial to extend robust SIM to distinguish such outliers in scale difference.

Note that these possible extensions of robust SIM are not straightforward, because the computations of the resulting model would become considerably

more complicated. Therefore, we must consider efficient algorithms that guarantee a high level of estimation accuracy.

Acknowledgements We are grateful to the Reviewers for their comments which have helped us to improve the manuscript.

OpenAccess This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Proof

The updating formula of $\boldsymbol{\eta}$ in (15) can be obtained by calculating the conditional expectation. To avoid calculating the integral directly, an algebraic expression of the formula is more convenient. Therefore, I derive here a more detailed expression of the updating formula.

Proposition A1 Consider the updating formula in (15)

$$\boldsymbol{\eta}^{(k+1)} = \left(\frac{\hat{\tau}_i^{(k)}}{\sigma_3^{(k)2}} \mathbb{E}_{\theta^{(k)}} [X_i^T X_i \boldsymbol{\eta} | y_i, \tau_i] \right)^{-1} \frac{\hat{\tau}_i^{(k)}}{\sigma_3^{(k)2}} \mathbb{E}_{\theta^{(k)}} [X_i^T (y_i - \alpha_i \mathbf{1}_{M_i}) | y_i, \tau_i].$$

Each expectation can be derived in more detail as

$$\begin{aligned} \mathbb{E}_{\theta^{(k)}} [X_i^T X_i \boldsymbol{\eta} | y_i, \tau_i] &= \mathbb{E}_{\theta^{(k)}} [\tau_i | y_i] \mathbf{B}_i^T \mathbf{B}_i \\ &\quad - 2\mathbb{E}_{\theta^{(k)}} [\tau_i | y_i] \mathbb{E}_{\theta^{(k)}} [\beta_i | y_i] \mathbf{B}_i^T \mathbf{D}_i + \mathbb{E}_{\theta^{(k)}} [\tau_i | y_i] \mathbb{E}_{\theta^{(k)}} [\beta_i^2 | y_i] \mathbf{D}_i^T \mathbf{D}_i \end{aligned} \tag{17}$$

and

$$\begin{aligned} \mathbb{E}_{\theta^{(k)}} [X_i^T (y_i - \alpha_i \mathbf{1}_{M_i}) | y_i, \tau_i] &= \mathbb{E}_{\theta^{(k)}} [\tau_i | y_i] \mathbf{B}_i^T y_i \\ &\quad + \mathbb{E}_{\theta^{(k)}} [\tau_i | y_i] \mathbb{E}_{\theta^{(k)}} [\beta_i | y_i, \tau_i] \mathbf{D}_i^T y_i - \mathbb{E}_{\theta^{(k)}} [\tau_i | y_i] \mathbb{E}_{\theta^{(k)}} [\alpha_i | y_i, \tau_i] \mathbf{B}_i^T \mathbf{1}_{M_i} \\ &\quad - \mathbb{E}_{\theta^{(k)}} [\tau_i | y_i] \mathbb{E}_{\theta^{(k)}} [\alpha_i \beta_i | y_i, \tau_i] \mathbf{D}_i^T \mathbf{1}_{M_i}, \end{aligned} \tag{18}$$

where

$$\mathbb{E}_{\theta^{(k)}} [\alpha_i \beta_i | y_i, \tau_i] = \mathbb{E}_{\theta^{(k)}} [\beta_i | y_i, \tau_i] \mathbf{c}_1^T (y_i - \mathbf{B}_i \boldsymbol{\eta}) + c_2 \mathbb{E}_{\theta^{(k)}} [\beta_i^2 | y_i, \tau_i], \tag{19}$$

and the remaining expectation is obtained using (14).

Proof Differentiating (7) with respect to $\boldsymbol{\eta}$, and equating it to 0 yields

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\eta}} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(k)}) &= 0 \\ \iff \frac{1}{\sigma_3^2} E_{\boldsymbol{\theta}^{(k)}} [\boldsymbol{\tau}_i \mathbf{X}_i^T \mathbf{X}_i \boldsymbol{\eta} | \mathbf{y}] &= \frac{1}{\sigma_2^2} E_{\boldsymbol{\theta}^{(k)}} [\boldsymbol{\tau}_i (\mathbf{X}_i^T \mathbf{y}_i - \boldsymbol{\alpha}_i \mathbf{X}_i^T \mathbf{1}_{M_i}) | \mathbf{y}] \end{aligned}$$

Then using

$$\begin{aligned} f(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, \boldsymbol{\tau}_i | \mathbf{y}_i; \boldsymbol{\theta}^{(k)}) &= f(\boldsymbol{\tau}_i | \mathbf{y}_i; \boldsymbol{\theta}^{(k)}) f(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i | \mathbf{y}_i, \boldsymbol{\tau}_i; \boldsymbol{\theta}^{(k)}) \\ &= f(\boldsymbol{\tau}_i | \mathbf{y}_i; \boldsymbol{\theta}^{(k)}) f(\boldsymbol{\beta}_i | \mathbf{y}_i, \boldsymbol{\tau}_i; \boldsymbol{\theta}^{(k)}) f(\boldsymbol{\alpha}_i | \mathbf{y}_i, \boldsymbol{\beta}_i, \boldsymbol{\tau}_i; \boldsymbol{\theta}^{(k)}), \end{aligned}$$

we obtain (17) and (18). To calculate (19), we use the property of a Gaussian distribution, that is, omitting (k) , $E_{\boldsymbol{\theta}^{(k)}}[\boldsymbol{\alpha}_i | \mathbf{y}_i, \boldsymbol{\beta}_i, \boldsymbol{\tau}_i]$ can be rewritten as

$$\begin{aligned} E_{\boldsymbol{\theta}^{(k)}}[\boldsymbol{\alpha}_i | \mathbf{y}_i, \boldsymbol{\beta}_i, \boldsymbol{\tau}_i] &= (\text{Cov}_{\boldsymbol{\theta}^{(k)}}[\boldsymbol{\alpha}_i, \mathbf{y}_i | \boldsymbol{\tau}_i], \text{Cov}_{\boldsymbol{\theta}^{(k)}}[\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i | \boldsymbol{\tau}_i]) \mathbf{V}_{\boldsymbol{\theta}^{(k)}}[(\mathbf{y}_i, \boldsymbol{\beta}_i) | \boldsymbol{\tau}_i]^{-1} \\ &\quad \begin{pmatrix} \mathbf{y}_i - E_{\boldsymbol{\theta}^{(k)}}[\mathbf{y}_i | \boldsymbol{\tau}_i] \\ \boldsymbol{\beta}_i - E_{\boldsymbol{\theta}^{(k)}}[\boldsymbol{\beta}_i | \boldsymbol{\tau}_i] \end{pmatrix} + E_{\boldsymbol{\theta}^{(k)}}[\boldsymbol{\alpha}_i | \boldsymbol{\tau}_i] \\ &= (\sigma_1^2 \mathbf{1}_{M_i}^T, 0) \begin{pmatrix} \mathbf{V}_i & \sigma_2^2 (\mathbf{D}_i \boldsymbol{\eta})^T \\ \sigma_2^2 \mathbf{D}_i \boldsymbol{\eta} & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y}_i - \mathbf{B}_i \boldsymbol{\eta} \\ \boldsymbol{\beta}_i \end{pmatrix} \\ &= (\mathbf{c}_1^T, c_2) \begin{pmatrix} \mathbf{y}_i - \mathbf{B}_i \boldsymbol{\eta} \\ \boldsymbol{\beta}_i \end{pmatrix} \end{aligned} \tag{20}$$

where \mathbf{c}_1 and c_2 are the first M_i elements and the last $(M_i + 1)$ th element of $(\text{Cov}_{\boldsymbol{\theta}^{(k)}}[\boldsymbol{\alpha}_i, \mathbf{y}_i | \boldsymbol{\tau}_i], 0) \mathbf{V}_{\boldsymbol{\theta}^{(k)}}[(\mathbf{y}_i, \boldsymbol{\beta}_i) | \boldsymbol{\tau}_i]^{-1}$, respectively. Then, inserting (20) in (19), we have

$$\begin{aligned} E_{\boldsymbol{\theta}^{(k)}}[\boldsymbol{\alpha}_i \boldsymbol{\beta}_i | \mathbf{y}_i] &= \int \boldsymbol{\beta}_i (\mathbf{c}_1^T (\mathbf{y}_i - \mathbf{B}_i \boldsymbol{\eta}) + \boldsymbol{\beta}_i c_2) f(\boldsymbol{\beta}_i | \mathbf{y}_i, \boldsymbol{\tau}_i; \boldsymbol{\theta}^{(k)}) d\boldsymbol{\beta}_i \\ &= \int \boldsymbol{\beta}_i \mathbf{c}_1^T (\mathbf{y}_i - \mathbf{B}_i \boldsymbol{\eta}) f(\boldsymbol{\beta}_i | \mathbf{y}_i, \boldsymbol{\tau}_i; \boldsymbol{\theta}^{(k)}) d\boldsymbol{\beta}_i + \int \boldsymbol{\beta}_i^2 c_2 f(\boldsymbol{\beta}_i | \mathbf{y}_i, \boldsymbol{\tau}_i; \boldsymbol{\theta}^{(k)}) d\boldsymbol{\beta}_i \\ &= E_{\boldsymbol{\theta}^{(k)}}[\boldsymbol{\beta}_i | \mathbf{y}_i, \boldsymbol{\tau}_i] \mathbf{c}_1^T (\mathbf{y}_i - \mathbf{B}_i \boldsymbol{\eta}) + c_2 E_{\boldsymbol{\theta}^{(k)}}[\boldsymbol{\beta}_i^2 | \mathbf{y}_i, \boldsymbol{\tau}_i] \end{aligned}$$

Thus we obtain the explicit update formula for $\boldsymbol{\eta}$. □

References

- Brumback LC, Lindstrom MJ (2004) Self modeling with flexible, random time transformations. *Biometrics* 60(2):461–470
- De Boor C (1978) *A practical guide to splines*. Springer, New York
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B (Statistical Methodology)* 39(1):1–38
- Earls C, Hooker G (2017) Variational bayes for functional data registration, smoothing, and prediction. *Bayesian Anal* 12(2):557–582
- Gaffney S (2004) Probabilistic curve-aligned clustering and prediction with regression mixture models. Ph.D. thesis, University of California, Irvine
- Gaffney SJ, Smyth P (2005) Joint probabilistic curve clustering and alignment. In: *Advances in neural information processing systems*, pp 473–480
- Gervini D, Gasser T (2004) Self-modelling warping functions. *J R Stat Soc Ser B (Statistical Methodology)* 66(4):959–971
- Gervini D, Gasser T (2005) Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. *Biometrika* 92(4):801–820
- James GM (2007) Curve alignment by moments. *Ann Appl Stat* 1(2):480–501
- Ke C, Wang Y (2001) Semiparametric nonlinear mixed-effects models and their applications. *J Am Stat Assoc* 96(456):1272–1298
- Kneip A, Engel J (1995) Model estimation in nonlinear regression under shape invariance. *Ann Stat* 23(2):551–570
- Kneip A, Gasser T (1988) Convergence and consistency results for self-modeling nonlinear regression. *Ann Stat* 16(1):82–112
- Kneip A, Gasser T (1992) Statistical tools to analyze data representing a sample of curves. *Ann Stat* 20(3):1266–1305
- Lange KL, Little RJ, Taylor JM (1989) Robust statistical modeling using the t distribution. *J Am Stat Assoc* 84(408):881–896
- Lawton W, Sylvestre E, Maggio M (1972) Self modeling nonlinear regression. *Technometrics* 14(3):513–532
- Lindstrom MJ (1995) Self-modelling with random shift and scale parameters and a free-knot spline shape function. *Stat Med* 14(18):2009–2021
- Liu C, Rubin DB (1994) The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* 81(4):633–648
- Liu C, Rubin DB (1995) ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Stat Sin* 5(1):19–39
- Liu X, Yang MC (2009) Simultaneous curve registration and clustering for functional data. *Comput Stat Data Anal* 53(4):1361–1376
- MacQueen (1967) Some methods for classification and analysis of multivariate observations. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, University of California Press, Berkeley, CA, vol 1, pp 281–297
- Matos LA, Prates MO, Chen MH, Lachos VH (2013) Likelihood-based inference for mixed-effects models with censored response using the multivariate- t distribution. *Stat Sin* 23:1323–1345
- Meng XL, Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 80(2):267–278
- Olszewski RT (2001) Generalized feature extraction for structural pattern recognition in time-series data. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA
- Pinheiro JC, Liu C, Wu YN (2001) Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *J Comput Graph Stat* 10(2):249–276
- Raket LL, Sommer S, Markussen B (2014) A nonlinear mixed-effects model for simultaneous smoothing and registration of functional data. *Pattern Recognit Lett* 38:1–7
- Ramsay JO, Li X (1998) Curve registration. *J R Stat Soc Ser B (Statistical Methodology)* 60(2):351–363
- Ramsay JO, Silverman BW (2005) *Functional data analysis*. Springer, New York
- Rønn BB (2001) Nonparametric maximum likelihood estimation for shifted curves. *J R Stat Soc Ser B (Statistical Methodology)* 63(2):243–259
- Sakoe H, Chiba S (1978) Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans Acoust Speech Signal Process* 26(1):43–49

- Song PXX, Zhang P, Qu A (2007) Maximum likelihood inference in robust linear mixed-effects models using multivariate t distributions. *Stat Sin* 17:929–943
- Srivastava A, Wu W, Kurtek S, Klassen E, Marron JS (2011) Registration of functional data using Fisher-Rao metric. arXiv preprint [arXiv:11033817](https://arxiv.org/abs/11033817)
- Szczesniak RD, Viele K, Cooper RL (2014) Mixtures of self-modelling regressions. *J Biom Biostat* 5(4):1–8
- Telesca D, Inoue LYT (2008) Bayesian hierarchical curve registration. *J Am Stat Assoc* 103(481):328–339
- Verbeke G (1997) *Linear mixed models for longitudinal data*. Springer, New York
- Welsh A, Richardson A (1997) Approaches to the robust estimation of mixed models. In: Maddala GS, Rao CR (eds) *Handbook of statistics*. Elsevier, Amsterdam, pp 343–384
- Zhang Y, Telesca D (2014) Joint clustering and registration of functional data. arXiv preprint [arXiv:14037134](https://arxiv.org/abs/14037134)

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