SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Article

Statistical Methods in Medical Research
0(0) 1–13

DOI: 10.1177/0962280217710050
journals.sagepub.com/home/smm
SAGE

Dissimilarity for functional data clustering based on smoothing parameter commutation

ShengLi Tzeng,[1] Christian Hennig,[2] Yu-Fen Li[1] and Chien-Ju Lin[3]

Abstract

Many studies measure the same type of information longitudinally on the same subject at multiple time points, and clustering of such functional data has many important applications. We propose a novel and easy method to implement dissimilarity measure for functional data clustering based on smoothing splines and smoothing parameter commutation. This method handles data observed at regular or irregular time points in the same way. We measure the dissimilarity between subjects based on varying curve estimates with pairwise commutation of smoothing parameters. The intuition is that smoothing parameters of smoothing splines reflect the inverse of the signal-to-noise ratios and that when applying an identical smoothing parameter the smoothed curves for two similar subjects are expected to be close. Our method takes into account the estimation uncertainty using smoothing parameter commutation and is not strongly affected by outliers. It can also be used for outlier detection. The effectiveness of our proposal is shown by simulations comparing it to other dissimilarity measures and by a real application to methadone dosage maintenance levels.

Keywords

Clustering, irregular longitudinal data, functional data, smoothing splines, dissimilarity, outliers

1 Introduction

Clustering sets out to find groups of subjects based on several different characteristics, where subjects within a cluster are considered to be similar based on the given characteristics. The degree of similarity and dissimilarity can be defined in many ways, and there are many clustering methods, including hierarchical clustering, k-means, DBSCAN, etc. Berkhin[1] gives an overview of both partition-based and hierarchical clustering methods, Bouveyron and Brunet-Saumard[2] review popular partition-based methods for high-dimensional data, and Murtagh and Contreras[3] review several hierarchical clustering algorithms, see also Hennig et al.[4] A bottom-up hierarchical method does not require any statistical model assumption, rather only a linkage method by which two clusters are merged at each step of the hierarchical agglomerative process. This process can be displayed by a dendrogram from which clusters are obtained. However, once merged, clusters cannot be separated at the next step. Partition-based methods, in contrast, partition subjects into a desired number of clusters, which needs to be specified. The cluster assignment of subjects for one number of clusters for these methods does not restrict cluster assignments for other numbers of clusters.

In many situations, the same type of information on the same subject is measured at multiple time points. To cluster such data, one should take into account the data format and the temporal order structure. Data of each subject are often collected at unequally spaced time points. As a result of aligning records into a conventional 'variable-by-variable' format, there are many 'empty' records at regular time points. Those empty records can be regarded as a kind of 'missing' value. In addition, even if all subjects were observed at the same time points,

[1]Department of Public Health, China Medical University, Taiwan
[2]Department of Statistical Science, University College London, UK
[3]MRC Biostatistics Unit, University of Cambridge, UK

Corresponding author:
Chien-Ju Lin, MRC Biostatistics Unit, Cambridge Institute of Public Health, Forvie Site, Robinson Way, Cambridge Biomedical Campus, Cambridge CB2 0SR, UK.
Email: chienju@mrc-bsu.cam.ac.uk

conventional clustering fails to take into account the temporal order structure of variables, on which adjacent measurements for the same subject are expected to have similar values. Functional data clustering exists for grouping such data. Three major categories for functional data clustering are dissimilarity-based methods, decomposition-based methods and model-based methods.

Dissimilarity-based methods, using pointwise dissimilarities between pairs of subjects, are the most straightforward approach.[5–7] These methods often take care of data format and time coordinates by certain curve smoothing or imputation techniques, and subsequently dissimilarities between subjects are computed to which the conventional dissimilarity-based methods can be applied. Little attention, however, has been paid to the uncertainty of smoothing or imputation. To the best of our knowledge, the only two exceptions are (a) a prediction-based approach of Alonso et al.[8] that was modified by Vilar et al.,[9] and (b) a hypotheses-testing-like approach of Maharaj.[10] The former is computationally intensive and the latter is designed for invertible ARMA processes, which restricts their application.

Decomposition-based methods overcome the issue of smoothing and sequential order by transforming the observed data into a finite series of common features. These procedures deal with the uncertainty of smoothing implicitly. For example, Abraham et al.[11] used spline basis functions, James et al.[12] used functional principal component analysis, and Warren Liao[13] reviewed more sophisticated 'feature-extraction' algorithms. These approaches define common features for all groups and then assign weights to features by which groups are identified. Each group has different weights on those features, and each group can be interpreted according to its lower-dimensional projection on features. Features extracted from certain transformations of data are also popular, such as spectral densities,[14] periodograms[15,16] and permutation distributions.[17] Nonetheless, in reality, not all groups share the same number of features, and it is not easy to determine an appropriate number of dimensions.

In light of the difficulties encountered by the first two methods, many researchers suggest the third alternative: various model-based frameworks. They estimate individual underlying curves and cluster subjects simultaneously, and then statistical inference can be made based on the working models for clusters, such as measuring the uncertainty for cluster assignment and 'within-cluster' variation. Unfortunately, these approaches encounter other challenges. Purely parametric functional forms as used in Jones and Nagin[18] may not be realistic, and its assumption of subjects sharing the same 'underlying' curve within a group can be too restrictive. Applying semi- or non-parametric methods usually requires some dimension reduction within each group (e.g. FCM,[19] funHDDC,[20] Funclust[21] and K-centre[22]), which encounters a similar problem as decomposition-based methods. A pure likelihood-based framework (without dimension reduction) called longclust is proposed by McNicholas and Murphy.[23] This method is limited to short time series and breaks down easily due to the curse of dimensionality. Even worse, the notion of distribution for random functions is not well-defined as curves could have infinite dimensions.[24]

We have reviewed the strengths and weaknesses of the existing functional data clustering methods. Moreover, it is worth mentioning the dilemma resulting from curve variability. Clustering curves can be a difficult 'chicken-and-egg' problem between (a) how to determine the within-cluster variations before identifying subgroups, and (b) how to separate subgroups when within-cluster variations are unknown. This dilemma is related directly to the smoothing uncertainty problem in dissimilarity-based approaches. Decomposition- and model-based approaches estimate such variability with necessity, but the magnitude of the estimate is often distorted when outliers occur. A two-step strategy exploiting relative merits of different methods seems reasonable: initially separate potential outliers based on an 'outlier-invariant' pairwise dissimilarity, and then form main clusters with another appropriate clustering method. For such a strategy, a dissimilarity measure concerning the variability of curve estimation or feature selection is crucial.

In this article, we develop an easily implementable and practically advantageous dissimilarity measure between subjects. The curve smoothing used here is based on the technique of smoothing splines, which is completely determined by the chosen smoothing parameter. With an infinite smoothing parameter, the curve is estimated as a straight line, while the curve interpolates the observed data with a zero smoothing parameter. The innovation of our method is to measure the dissimilarity between subjects based on pair-by-pair varying curve estimates for a subject. The concepts are that (a) smoothing parameters of smoothing splines reflect the inverse of the signal-to-noise ratios and (b) the estimated curves for two similar subjects are expected to be close if an identical smoothing parameter is applied to both sets of observations. Specifically, if the unobserved true curves of subjects i and j are similar, their curve estimates should resemble with each other, no matter whether we use a smoothing parameter primarily for the i-th or the j-th subject. Our dissimilarity is then calculated through commuting between the smoothing parameters for a pair.

The rest of the article is organized as follows. Section 2 describes the proposed Smoothing Parameter Commutation dissimilarity and some of its properties. Its effectiveness is shown through simulations comparing to other dissimilarity measures in Section 3. An example of its application to methadone dosages observations is given in Section 4, where we also identify outliers with a simple but efficient method. Finally, Section 5 provides some concluding remarks and discussion concerning future directions.

2 Method

2.1 Smoothing splines

We utilize the smoothing spline to estimate curves of subjects. Assume that the curve of the i-th subject is observed as a set of measurements $\{y_{i,1}, \ldots, y_{i,K_i}\}$ contaminated by noises at distinct finite time points $\{t_{i,1}, \ldots, t_{i,K_i}\}$ in an interval $[T_L, T_U]$ according to the model

$$y_{i,k} = f_i(t_{i,k}) + \epsilon_{i,k}, \quad k = 1, \ldots, K_i, \quad i = 1, \ldots, n \tag{1}$$

where $f_i(\cdot)$ is the function of the true curve, and at time $t_{i,k}$ the noise $\epsilon_{i,k} \overset{iid}{\sim} N(0, \sigma^2)$ and the true value $f_i(t_{i,k})$ are both unobservable. A reasonable estimation of f_i is to minimize $\frac{1}{K_i}\sum_k (y_{i,k} - f_i(t_{i,k}))^2$ while controlling the smoothness of f_i by requiring $\int_{T_L}^{T_U} (f_i''(t))^2 \mathrm{d}t \le \rho$ for a positive ρ. This estimator is equivalent to a smoothing spline $\widehat{f_i}(\cdot; \lambda)$ that minimizes

$$\frac{1}{K_i}(\mathbf{y}_i - \mathbf{f}_i)'(\mathbf{y}_i - \mathbf{f}_i) + \lambda \int_{T_L}^{T_U} (f_i''(t))^2 \mathrm{d}t \tag{2}$$

where $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,K_i})'$ and $\mathbf{f}_i = \big(f_i(t_{i,1}), \ldots, f_i(t_{i,K_i})\big)'$ given a smoothing parameter $\lambda \ge 0$.[25,26] With an infinite λ the curve $\widehat{f_i}$ is a straight line, while with $\lambda = 0$, $\widehat{f_i}$ interpolates exactly all the data points. There are various methods to determine an appropriate λ in (2), and once it is chosen, $\hat{f}_i(t; \lambda)$ is completely determined over $t \in [T_L, T_U]$. We exploit a mixed-effects model representation[27] to choose λ in (2), which formulates $\mathbf{y}_i$ as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{u}_i + \boldsymbol{\epsilon}_i \tag{3}$$

where $\boldsymbol{\beta}_i$ is the fixed effect, $\mathbf{X}_i$ has two columns being 1's and $(t_{i,1}, \ldots, t_{i,K_i})'$, $\boldsymbol{\epsilon}_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,K_i})' \sim N(0, \sigma^2\mathbf{I})$, and $\mathbf{u}_i \sim N(\mathbf{0}, \sigma_u^2\boldsymbol{\Psi})$ with $\sigma_u^2 = \sigma^2/(K_i\lambda)$ and a specific correlation $\boldsymbol{\Psi}$. Speed[28] pointed out that minimizing (2) is equivalent to finding the minimum variance linear unbiased predictor of $\mathbf{y}_i$ in (3) with λ fixed. They dealt with coordinates based on transformed values. For $t, \tau \in [T_L, T_U]$, let $\tilde{t} = (t - T_L)/(T_U - T_L)$ and $\tilde{\tau} = (\tau - T_L)/(T_U - T_L)$, then $\tilde{t}, \tilde{\tau} \in [0, 1]$. They use a correlation matrix with its (k, k^*)-th element being $\int_0^1 (\tilde{t}_{i,k} - \tilde{\tau})_+(\tilde{t}_{i,k^*} - \tilde{\tau})_+ \mathrm{d}\tilde{\tau}$, where $a_+ = \max(0, a)$. That is, setting the (k, k^*)-th element of $\boldsymbol{\Psi}$ to be a function of $t_{i,k}$ and t_{i,k^*} as

$$\begin{aligned} &\int_{T_L}^{T_U} \frac{\{(t_{i,k} - T_L) - (\tau - T_L)\}_+}{T_U - T_L} \frac{\{(t_{i,k^*} - T_L) - (\tau - T_L)\}_+}{T_U - T_L} (T_U - T_L)^{-1} \mathrm{d}\tau \\ &= (T_U - T_L)^{-3} \int_{T_L}^{T_U} (t_{i,k} - \tau)_+(t_{i,k^*} - \tau)_+ \mathrm{d}\tau \end{aligned} \tag{4}$$

For any given $t_{i,k}$ and t_{i,k^*} in (4), $(t_{i,k} - \tau)_+$ and $(t_{i,k^*} - \tau)_+$ are two truncated linear functions over $\tau \in [T_L, T_U]$, and the integral is a convolution of these two functions, which does not depend on τ. Note that if one does not treat λ as fixed, λ can be expressed as a function of the variance of $\mathbf{u}_i$ in (3). Under the Gaussian assumption for $\boldsymbol{\epsilon}_i$ and $\mathbf{u}_i$, the two variance components σ_u^2 and σ^2 can be determined based on the restricted maximum likelihood method (REML), so that λ is also determined, and $K_i\lambda$ has a useful interpretation as the inverse of the signal-to-noise ratio σ_u^2/σ^2. Additionally, it has been shown that the smoothing results based on (3), (4) and REML are robust even when the correlation structure of $\boldsymbol{\epsilon}_i$ is mis-specified.[27,29]

2.2 Smoothing Parameter Commutation dissimilarity

The concept of our method is that if the 'true' f_i and f_j are similar, it is expected that $\hat{f}_i$ and $\hat{f}_j$ from $\mathbf{y}_i$ and $\mathbf{y}_j$ should be close, given an identical smoothing parameter. Our proposal starts with finding $\hat{\lambda}_i$ in (3) for a subject i based

on $\mathbf{y}_i$. The estimated curve is denoted by $\hat{f}_i(\cdot;\hat{\lambda}_i)$, which means that $\hat{f}_i(\cdot;\lambda)$ is estimated by setting $\lambda=\hat{\lambda}_i$ in (2) based on the observations $\mathbf{y}_i$. Given $\hat{\lambda}_i$, we can also obtain $\hat{f}_j(\cdot;\hat{\lambda}_i)$ based on the observations $\mathbf{y}_j$. Similarly, we exchange the roles of the two subjects to obtain $\hat{f}_j(\cdot;\hat{\lambda}_j)$ and $\hat{f}_i(\cdot;\hat{\lambda}_j)$. The dissimilarity between subjects i and j is defined as

$$d_{i,j}=\frac{1}{2}\left\{\left[\int_{T_L}^{T_U}\left(\hat{f}_i(t;\hat{\lambda}_i)-\hat{f}_j(t;\hat{\lambda}_i)\right)^2\mathrm{d}t\right]^{1/2}+\left[\int_{T_L}^{T_U}\left(\hat{f}_i(t;\hat{\lambda}_j)-\hat{f}_j(t;\hat{\lambda}_j)\right)^2\mathrm{d}t\right]^{1/2}\right\} \tag{5}$$

Due to the roles of $\hat{\lambda}_i$ and $\hat{\lambda}_j$ in (5), we call it a Smoothing Parameter Commutation dissimilarity. It takes the variation of smoothing into consideration with different λ's for different pairs of (i,j)'s, rather than focusing on the dissimilarity between (fixed) estimated curves. Note that $d_{i,j}\geq 0$ and $d_{i,j}=0$ if $i=j$, and it is clear that $d_{i,j}=d_{j,i}$, so conventional dissimilarity-based clustering methods can be applied. Note that the triangle inequality cannot be proved, in general, hence the term 'dissimilarity' rather than 'distance', but this is not required for dissimilarity-based clustering methods. The dissimilarity reduces to rooted integral squared difference of f_i and f_j when no missing values and measurement errors are present.

Our proposal has several advantages. First, data observed at irregular time points can be handled directly, because of the nature of smoothing splines. Second, the dissimilarity also serves as a useful tool for outlier detection (see Section 4). Third, the implementation is handy since subroutines for smoothing splines and numerical integration are widely available. Although the computational burden for (5) seems heavy at first glance, it can be done quite efficiently among n subjects. Given λ, a fast $O(K_i)$ algorithm to compute $\widehat{f_i}(t;\lambda)$ does exist.[30] Thus, one needs to solve $\hat{\lambda}_i$ in (3) only n times for the n subjects, and then one adopts the fast algorithm for $\left\{\hat{f}_j(t;\hat{\lambda}_i):\ i,j=1,\ldots,n\right\}$. Therefore, the computational complexity is proportional to that in treating $f=\widehat{f_i}(t;\lambda_i)$ as fixed and calculating the dissimilarity as square root of $\int_{T_L}^{T_U}\left(\hat{f}_i(t;\hat{\lambda}_i)-\hat{f}_j(t;\hat{\lambda}_j)\right)^2\mathrm{d}t$ (the latter procedure is referred to as d_{SS} in what follows).[31]

3 Simulation study

We conduct a simulation to investigate whether our proposed Smoothing Parameter Commutation measure is more capable than other dissimilarity measures when observations are contaminated with (independent or dependent) noises. If an analyst is interested in the relative shape patterns of curves, regardless of shift, shrinkage, expansion or magnitude, then several alignment, normalization and warping tools can be applied in preprocessing.[32–34] In order to not lose focus, we do not consider dissimilarity measures engaging with the preprocessing.

We consider the following four random curve models over $t\in[0,1]$:

$$\begin{aligned}
f^{(1)}(t;\eta)&=3\eta,\\
f^{(2)}(t;\eta)&=\sin(2\pi t)-t+2\eta\cos(4\pi t),\\
f^{(3)}(t;\eta)&=3t+2\eta t,\\
f^{(4)}(t;\eta)&=5\eta\left\{(t-0.5)^2-2t(1-t)\right\}
\end{aligned}$$

where $\eta\sim N(1,0.3^2)$. The four functional forms represent constant, periodic, linear and nonlinear (unobserved) true curves, respectively. The observed data are generated according to (1) at 200 time points, $t_k\in\{0,1/199,\ldots,198/199,1\}$, with noise coming from four mechanisms

$$\begin{aligned}
\text{WN}:&\quad \epsilon_k=\xi_k,\\
\text{AR}:&\quad \epsilon_k=0.8\epsilon_{k-1}+\xi_k,\\
\text{SARMA}:&\quad \epsilon_k=0.8\epsilon_{k-10}+0.8\xi_{k-10}+\xi_k,\\
\text{BILR}:&\quad \epsilon_k=0.8\epsilon_{k-1}+0.2\xi_{k-1}-0.2\epsilon_{k-1}\xi_{k-1}+\xi_k
\end{aligned} \tag{6}$$

where $\xi_k\overset{iid}{\sim}N(0,1)$, and ξ_k is independent of $\epsilon_{k'}$ for $k'\neq k$. That is, we set $K_i\equiv 200$, $t_{ik}\equiv(k-1)/199$. The four noise mechanisms are examples of usual assumption for noises: a purely independent process, a stationary process, a cyclostationary process and a non-stationary process. For each combination of $f\in\left\{f^{(1)},f^{(2)},f^{(3)},f^{(4)}\right\}$ and

Table 1. Dissimilarity measures to be compared.

Notation	Description	Literature
d_{EUCL}	Point-wise Euclidean distance	$\sqrt{\sum_k (y_{ik} - y_{jk})^2}$
d_{SPC}	Smoothing Parameter Commutation	The proposed method
d_{MAH}	Parametric testing of equality of processes	Maharaj[10]
d_{GLK}	Nonparametric equality testing of log-spectra	Fan and Zhang[14]
d_{SS}	Based on spline smoothing curves	Ramsay and Silverman[31]
d_{CORT}	Correlation-based modification of d_{EUCL}	Chouakriya and Nagabhushan[36]
d_{IP}	Based on integrated periodogram	de Lucas[16]
$d_{PRED,h}$	Based on predicted values at future	Vilar et al.[9]
d_{CID}	Complexity-based modification of d_{EUCL}	Batista et al.[37]
d_{PDC}	Permutation distributions of order patterns	Brandmaier[17]

mechanism ϵ_k, 10 series are generated according to 10 independent η's as well as 10 sets of ϵ_k's. In total, there are 160 series mimicking the longitudinal observations from 160 subjects.

Ten dissimilarity measures as listed in Table 1 are applied to the simulated data. They include seven measures in Montero and Vilar,[35] the proposed Smoothing Parameter Commutation method (d_{SPC}), point-wise Euclidean distance (d_{EUCL}), and d_{ss} as mentioned in the last section. Two comparison criteria are defined as follows:

$$Q = n^{-2} \min_{a,b} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{\left(a + b\hat{d}_{i,j} - d_{i,j}\right)^2}{d_{i,j}},$$

$$R = n^{-2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} (\hat{r}_{i,j} - r_{i,j})^2$$

where $\hat{d}_{i,j}$ is one of the considered dissimilarity measures between the i-th and j-th subjects, $d_{i,j} = \sqrt{\sum_k (f_i(t_{ik}) - f_j(t_{jk}))^2}$ is the true dissimilarity without noise, and $\hat{r}_{i,j}$ and $r_{i,j}$ are the corresponding ranks of $\hat{d}_{i,j}$ and $d_{i,j}$ among all pairs of (i, j)'s, respectively. The quantity Q reflects the loss, normalized by the true dissimilarity scales, for (linear) approximation to all the pairs of true dissimilarities, while R measures the deviation from monotonicity between $\hat{d}_{i,j}$ and $d_{i,j}$. A good measure should have a small value of Q and R. The averaged Q- and R-values for the 10 measures over 200 simulation replicates are given in Tables 2 and 3, respectively.

The Q value is a loss function based on the best affine approximation, which measures how close optimally linearly transformed dissimilarity measures are to the true d_{ij}. Obviously, this would be a good feature for a dissimilarity measure in such cases, but one may argue that achieving this is not the aim of some of the measures listed in Table 1, so we propose alternative measures, see below. Q favours Euclidean distance based dissimilarities. R measures by and large the same feature but does not rely on metric approximation. Note that in areas of high density of curves small changes in the metric approximation quality can lead to much larger changes in ranks, so that differences in R are often much clearer than differences in Q. In any case, the two comparison criteria are highly coherent in that they almost always detect the same best and worst measures. As expected, without missing data, d_{EUCL} is often among the best measures and it looks unbiased for estimating the true distance between curves in many situations. But it does not do well enough if the signal or noise is periodic ($f^{(2)}$, SARMA, respectively). Our d_{SPC} method and d_{SS} always are among the best three measures, both for the 10 curves within an individual group and for 160 curves as a whole. The non-stationarity of BILR can occasionally lead to larger-variance noises. As a result, noises mask the signal at some time points. When the signal-to-noise ratio becomes lower, most dissimilarity measures cannot perform well. As expected, if noises come from BILR, d_{SPC} and d_{SS} have a smaller advantage, that is, their performances are not very different from other methods. Note that d_{SS} and d_{SPC} yield an almost identical result within groups, due to them both utilizing the mixed-effects model representation of smoothing splines. The difference lies in that d_{SS} regards $\hat{f}_i(t; \hat{\lambda}_i)$ as a fixed estimate of f_i. Our d_{SPC} method

Table 2. Averaged Q-values over 200 simulated replicates among 10 dissimilarity measures for each combination of f and ϵ_k (with 10 random curves), and all the 160 curves.

	d_{EUCL}	d_{SPC}	d_{MAH}	d_{GLK}	d_{SS}	d_{CORT}	d_{IP}	$d_{PRED,h}$	d_{CID}	d_{PDC}
$f^{(1)}+$W	**1.59**	**0.36**	8.45	8.55	**0.37**	3.82	9.32	1.85	2.19	8.63
$f^{(1)}+$A	**5.61**	**5.66**	8.07	8.02	**5.66**	6.4	9.54	**4.61**	5.82	7.96
$f^{(1)}+$S	8.07	**6.15**	8.67	8.55	**6.16**	8.68	10.53	**7.32**	8.19	8.78
$f^{(1)}+$B	**7.65**	**7.66**	8.48	8.48	**7.65**	7.89	11.88	**5.62**	7.87	8.48
$f^{(2)}+$W	2.21	**0.87**	3.79	3.56	**0.83**	3.54	**1.35**	3.96	2.76	4.06
$f^{(2)}+$A	**3.94**	**3.94**	**3.91**	3.97	**3.94**	**3.94**	5.69	4.01	**3.95**	3.97
$f^{(2)}+$S	3.96	**3.81**	3.83	**3.83**	**3.63**	3.94	5.71	4.04	3.93	3.95
$f^{(2)}+$B	**3.99**	**3.99**	4.05	4.06	**3.99**	**3.99**	12.32	4.04	4.04	4.09
$f^{(3)}+$W	1.49	**1.05**	1.40	1.40	**0.99**	1.52	**1.38**	1.58	1.50	1.53
$f^{(3)}+$A	**1.49**	**1.49**	**1.47**	**1.49**	**1.49**	1.50	2.38	1.51	**1.49**	**1.49**
$f^{(3)}+$S	1.53	**1.49**	**1.49**	**1.49**	**1.49**	1.52	4.07	1.54	1.52	1.51
$f^{(3)}+$B	**1.56**	**1.56**	**1.55**	**1.56**	**1.56**	1.57	12.44	1.58	**1.56**	**1.56**
$f^{(4)}+$W	**2.31**	**0.79**	3.17	3.19	**0.81**	2.94	3.53	2.69	2.54	3.18
$f^{(4)}+$A	**3.20**	**3.20**	3.21	3.27	**3.20**	3.23	4.01	**3.04**	3.23	3.22
$f^{(4)}+$S	3.29	**3.19**	3.23	**3.22**	**3.18**	3.29	4.74	3.32	3.28	3.25
$f^{(4)}+$B	**3.35**	**3.35**	3.38	3.38	**3.35**	3.36	9.01	3.31	3.38	3.37
ALL	**24.83**	**23.92**	29.29	29.30	**24.45**	26.13	32.38	25.63	28.86	29.28

W, A, S and B in the first column stand for WN, AR, SARMA and BILR in (6), respectively. Bold digits are the best 3 within each row.

Table 3. Averaged R-values over 200 simulated replicates among 10 dissimilarity measures for each combination of f and ϵ_k (with 10 random curves), and all the 160 curves.

	d_{EUCL}	d_{SPC}	d_{MAH}	d_{GLK}	d_{SS}	d_{CORT}	d_{IP}	$d_{PRED,h}$	d_{CID}	d_{PDC}
$f^{(1)}+$W	**0.73**	**0.24**	12.26	12.15	**0.24**	2.22	12.01	1.11	1.16	12.39
$f^{(1)}+$A	**4.89**	**4.89**	12.11	12.02	**4.89**	5.85	12.29	**3.98**	5.15	12.29
$f^{(1)}+$S	7.79	**5.21**	11.82	11.94	**5.24**	10.20	12.29	**7.23**	8.87	12.18
$f^{(1)}+$B	**7.73**	**7.73**	12.27	12.15	**7.73**	8.30	12.25	**4.70**	8.20	12.43
$f^{(2)}+$W	3.01	**1.04**	8.88	6.69	**1.01**	6.27	**1.29**	11.69	4.21	12.27
$f^{(2)}+$A	**9.20**	**9.20**	10.24	10.59	**9.19**	9.80	8.15	12.66	9.45	12.05
$f^{(2)}+$S	10.99	**8.19**	10.14	10.18	**7.88**	11.71	**7.85**	13.45	11.33	12.35
$f^{(2)}+$B	**10.59**	**10.6**	11.62	11.63	**10.6**	10.89	10.77	12.66	10.75	12.10
$f^{(3)}+$W	9.18	**4.49**	8.16	8.10	**4.27**	10.79	**6.78**	14.54	10.09	12.35
$f^{(3)}+$A	**11.5**	**11.53**	11.89	11.87	**11.53**	11.69	12.06	13.88	**11.53**	12.22
$f^{(3)}+$S	11.90	**11.53**	**11.72**	12.17	**11.34**	11.95	12.10	14.09	11.87	11.96
$f^{(3)}+$B	**11.99**	**12.02**	12.12	12.05	**12.02**	**11.94**	12.26	13.51	12.06	12.31
$f^{(4)}+$W	**4.63**	**1.31**	11.56	12.23	**1.32**	7.87	12.29	7.49	5.79	12.21
$f^{(4)}+$A	**9.89**	**9.89**	11.59	12.28	**9.88**	10.45	12.47	10.00	10.02	12.18
$f^{(4)}+$S	11.71	**10.23**	**11.34**	12.11	**10.23**	11.87	12.19	13.08	11.72	12.20
$f^{(4)}+$B	**11.24**	**11.26**	12.24	12.20	**11.26**	11.52	12.24	**10.83**	11.41	11.68
ALL	**1155.6**	**874.7**	4160.5	4063.7	**901.0**	1315.4	3768.6	1239.8	2693.1	4191.6

W, A, S and B in the first column stand for WN, AR, SARMA and BILR in (6), respectively. Bold digits are the best 3 within each row.

outperforms the others for between-group dissimilarity, which indicates the advantage of accounting for smoothing variation via smoothing parameter commutation. In certain cases $d_{PRED,h}$ and d_{MAH} are good measures, which also take estimation uncertainty into consideration.

For cases where the heterogeneities in magnitude and in shape are of interest, criteria targeting at the integrated euclidean distances are reasonable to use as a comparison index, whereas the criteria might not be sensible in applications such as segregating relative shape changes regardless of magnitudes in microarray experiments.[38] Therefore, in addition to Q and R, we examined the cluster recovery ability among measures list in Table 1 with the Nowak index, Rand index and adjusted Rand index. The Nowak index focuses on the largest cluster, the Rand

Table 4. Averaged measures over 200 simulated replicates among 10 dissimilarity measures for four groups of f among all the 160 curves.

	d_{EUCL}	d_{SPC}	d_{MAH}	d_{GLK}	d_{SS}	d_{CORT}	d_{IP}	$d_{PRED,h}$	d_{CID}	d_{PDC}
RAND	0.7705	0.8034	0.6179	0.6174	0.8016	0.772	0.6156	0.7295	0.7015	0.6139
	(0.0038)	(0.0015)	(0.0025)	(0.0026)	(0.0013)	(0.0039)	(0.0027)	(0.0034)	(0.00329)	(0.00256)
adjRAND	0.4154	0.4998	−0.0156	−0.014	0.4962	0.4205	0.0269	0.3129	0.2332	−0.0062
	(0.0043)	(0.0034)	(0.0002)	(0.0003)	(0.0027)	(0.0046)	(0.0009)	(0.0029)	(0.00415)	(0.00044)
Nowak	0.5679	0.6181	0.2607	0.2654	0.6165	0.573	0.3079	0.467	0.428	0.2786
	(0.0042)	(0.0040)	(0.0011)	(0.0012)	(0.0035)	(0.0046)	(0.0019)	(0.0027)	(0.0022)	(0.0016)

Digits in parentheses are the standard errors.

index measures the average performance, and the adjusted Rand index is a chance-corrected version of the Rand index. All are implemented in Dudek[39] and the reference therein gives the detailed definitions. There are four signal patterns in our simulation setup, despite various noise mechanisms. We applied 'partitioning around medoids' clustering (PAM[40]) with four clusters to the pairwise dissimilarity matrix of each measure. The true clusters were defined by the four random curve models. The average values and standard errors of indices for the 10 measures over 200 simulation replicates are given in Table 4. d_{SPC}, d_{SS}, d_{CORT} and d_{EUCL} outperform the other measures, and d_{SPC} is the best among the four by a small margin.

4 Real data application with outlier detection

We apply d_{SPC} defined in (5) to a methadone maintenance therapy dataset analysed in Lin et al.[41] Daily methadone dosages in milligrams (mg) for 314 participants between 01 January 2007 and 31 December 2008 were collected. Dosage records for each patient from day 1 to 180 were used for clustering. Lin et al.[41] categorized the dosages into seven levels, one of which is for missing values, and proposed a dissimilarity measure for clustering such ordinal data with extra missingness category. The ordering of time coordinates, however, was ignored in their approach. In this example, we use the daily dosage taken by patients, and do not recode missing values separately. Smoothing splines take care of the irregular follow-up time points of patients automatically, which may not be an easy task for some other measures listed in Table 1.

Real data, inevitably, are prone to have outliers. Garcia-Escudero et al.[42] give an overview of the impact of outliers on clustering and some approaches to deal with them. A clustering procedure with outlier removal consists of three steps: (a) calculating the dissimilarity matrix, (b) detecting and removing outliers and (c) grouping the remaining subjects into a desired number of clusters. An outlying curve in functional data is not only one with a few unusual high or low measurements, but also one that has an overall atypical magnitude or shape. Hubert et al.[43] distinguish the former as 'isolated outlier' and the latter as 'persistent outlier'. Among 'persistent outliers', Hyndman and Shang[44] call curves lying outside the range of the vast majority of the data 'magnitude outliers' and call those having a very different shape from other curves 'shape outliers'. Much research has been done on detection of persistent outliers.

Some work exploits the notion of data depth for sorting subjects into layers with a more outward layer more likely to be atypical (first proposed by Tukey[45]; see Gervini[46] and Hubert et al.[43] for an overview in the functional setting), often equipped with a functional boxplot as a visualization tool. Some methods rely on robust functional principal components,[47,48] which brings about various visualization tools such as bagplots and highest-density-region boxplots.[44] While most of the above focus on magnitude outliers, Arribas-Gil and Romo[49] propose the 'outliergram' to detect shape outliers as well as magnitude outliers. Other methods deal with phase heterogeneity relating to warping and alignment preprocessing,[50,51] which is beyond the scope of this work.

The aforementioned methods regard outliers as subjects that are very different to the majority. Alternatively, Ramaswamy et al.[52] propose a dissimilarity-based outlier detection method considering the dissimilarity to nearest neighbours. Their outlier definition is that outliers are those with no or only so few subjects nearby that these could not be interpreted as forming a relevant cluster. This method has been shown to be effective in the pattern recognition literature.[53,54] The difficulty in applying it to functional data is the construction of an appropriate dissimilarity, which is the main theme of the present study. By virtue of the good performance of d_{SPC} in the previous section, we consider a dissimilarity-based outlier detection method similar to Ramaswamy et al.[52] We obtain a pairwise dissimilarity matrix based on (5), and calculate the average dissimilarity of each participant to

their k nearest neighbours. The minimum size of clusters to be meaningful (k should be this size minus one) in principle depends on the application and the size of the dataset. Averaging the dissimilarities to the nearest neighbours implies that a collection of up to k subjects needs to be further away from the remaining subjects in order to be considered outliers than a single isolated subjects, because a single isolated outlier forms an even less relevant pattern. Note that in cluster analysis with outliers, there is an essential ambiguity about whether a small group of atypical subjects is a cluster on its own, or rather a group of outliers, see Garcia-Escudero et al.[42] Therefore, outlier detection methods in clustering will always require some tuning of this kind.

Two participants had average d_{SPC} values to the three nearest neighbours as 498 and 989, while the others had values between 34 and 282. The two participants were, therefore, considered as potential outliers. We assess the outlyingness of observations visually, using boxplots, without the need of involving formal model assumptions. Figure 1 shows the average dissimilarities to k nearest neighbours, where $k = 2, \ldots, 10$. As is clear from the graph, the two participants with the largest nearest neighbour dissimilarities are very far away from the other patients, regardless of k.

Hennig and Lin[55] used a flexible parametric bootstrap method to assess the number of clusters for the data in Lin et al.[41] The PAM solutions between 2 and 20 clusters were compared regarding the average silhouette width[40] and the prediction strength.[56] This suggested either five or six clusters. We use PAM with six clusters and use the adjusted Rand index to measure the similarity between clusters found in our study. We excluded the two outlying participants and applied PAM with six clusters to the pairwise d_{SPC} matrix of the 312 participants. Figure 2 shows the clustering result. Each horizontal line represents a dosage curves from day 1 to 180. Figure 3(a) shows the average dosage of each cluster. As seen, Groups 1 and 2 are more stable, remaining at dosages in [10,40] and [40,80], respectively. Group 3 has an upward trend while Group 4 has a downward trend, and their average dosages represented by the two curves cross around day 85. Group 5 goes up quickly and stays at dosage of around 80 mg. Although Group 6 has a similar trend to Group 5, it fluctuates heavily over a larger range and looks less stable. Overall, these figures indicate that a patient with early higher dosage taken (roughly above 60 mg before day 45) does not tend to reduce the level afterwards, and a monitoring between the second and third month can be critical.

Results based on \a model-based functional data clustering are also given for reference. We used the 'funcit' function in the 'funcy' package[57] on the Comprehensive R Archive Network (CRAN; R Core Team[58]). The model option of 'funcit' is set to be 'iterSubspace', i.e. an implementation of the algorithm in Chiou and Li.[22]

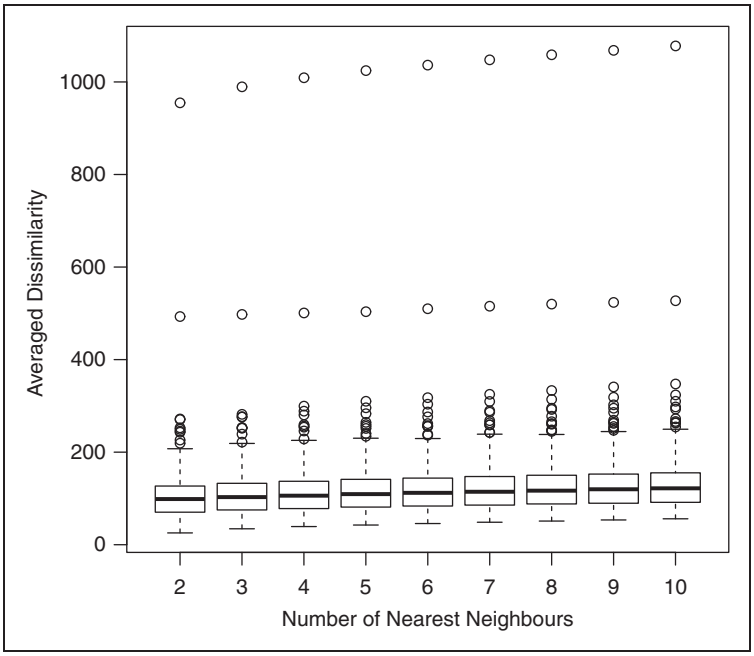

Figure 1. Distribution of the averaged dissimilarity d_{SPC} to the nearest k neighbours, where $k = 2, \ldots, 10$.

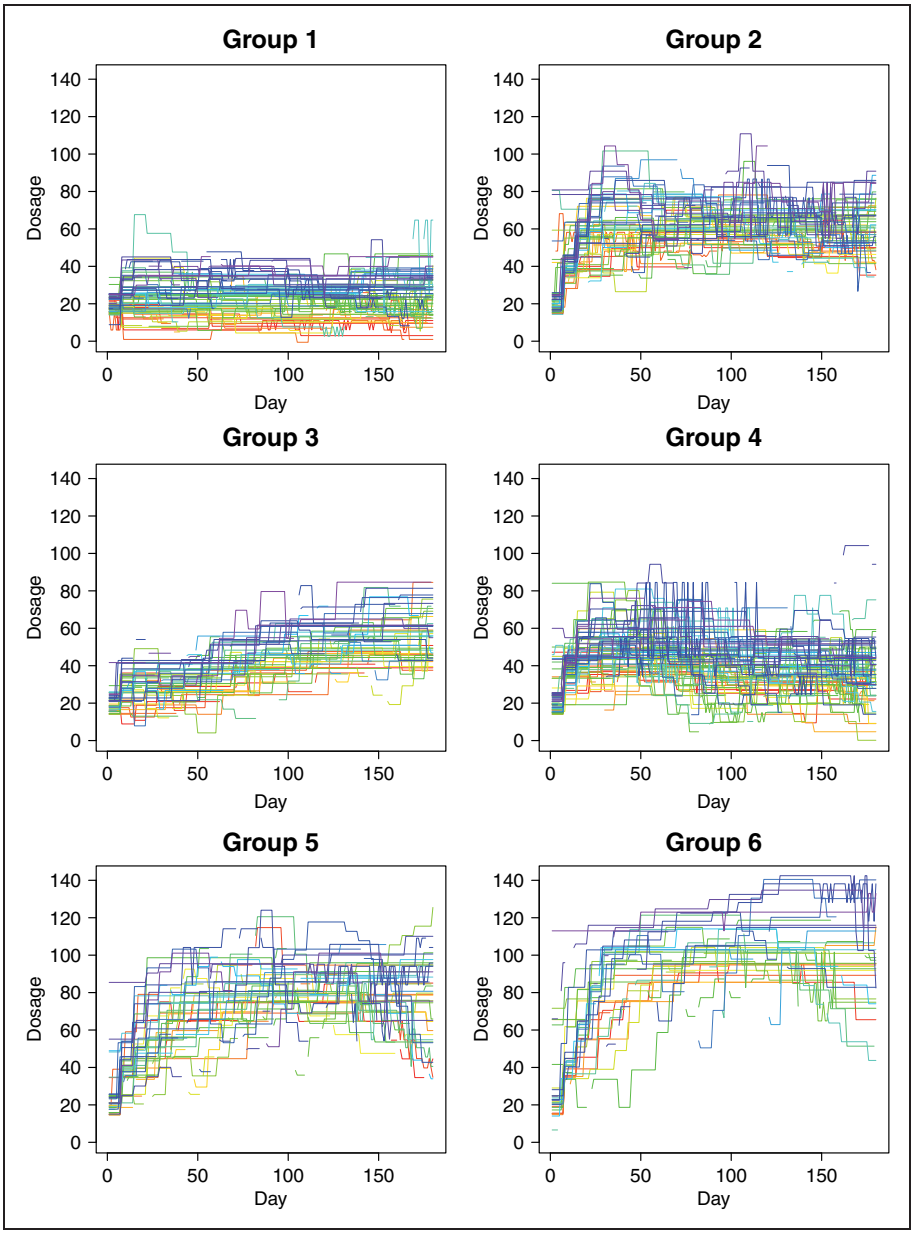

Figure 2. Subgroups from PAM clustering of the 312 patients in methadone maintenance therapy.

The 'funcy' package integrates several model-based clustering methods for functional data, but most of them require regular measurements and do not fit the methadone dosage example with many missing values. The only two methods of the package allowing irregular measurements are 'fitfclust' and 'iterSubspace', and we used the latter because of the length of time and amount of memory needed to run one iteration.

Profiles of the two aforementioned potential outliers are shown in Figure 3(b). The theoretical mean profiles of clusters obtained from the model-based clustering with and without outliers are shown in Figure 3(c) and (d), respectively. Excluding the two outliers improved the model-based method in the sense that the average dissimilarity to the mean profile was reduced by 7.6% from 166.7 to 154.9, which yielded more compact clusters. It seems inappropriate to assign the two outlying curves into the found groups. Enforcing their inclusion will exaggerate the within-group variation, no matter which groups they are assigned to. Then the boundaries of groups are more blurred, and the mean profiles are less representative of their groups.

We also applied the outliergram[49] for outlier detection. Five outliers were detected. Two of them were the same participants as identified by our method shown in Figure 3(b), confirming their outlyingness. The two grey curves

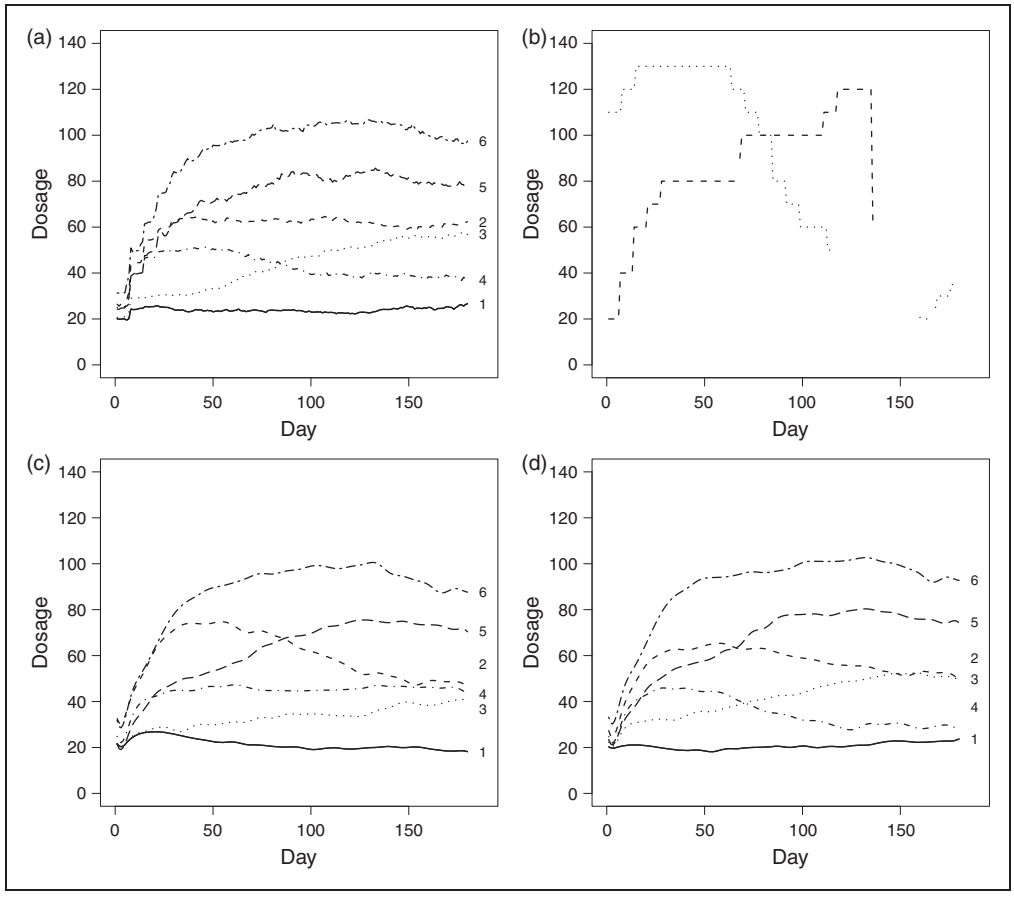

Figure 3. (a): Curves of average dosage of the six groups obtained from PAM and the pairwise dissimilarity matrix in Figure 2; (b) dosage profiles of the two potential outliers; (c) mean profiles of a model-based clustering method including outliers; (d) mean profiles of the model-based clustering method excluding outliers.

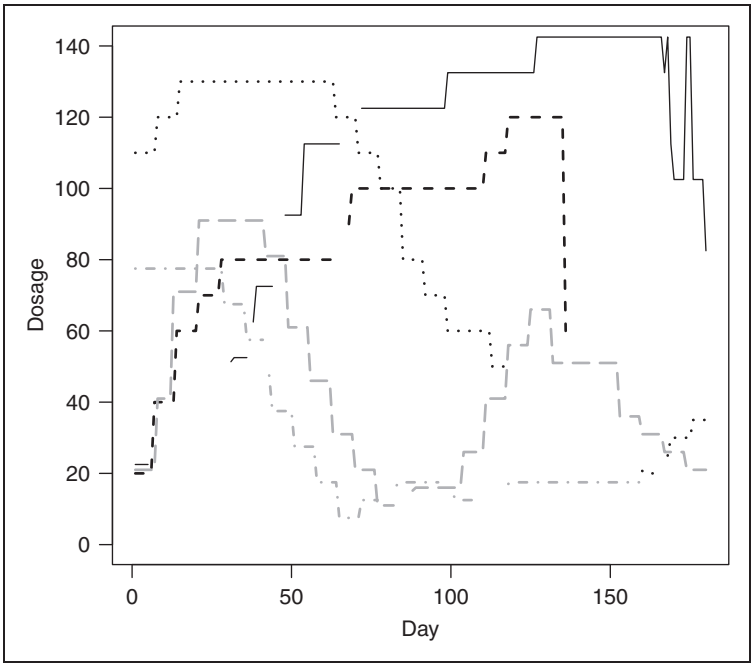

Figure 4. Dosage profiles of the five potential outliers identified by using outligram.

in Figure 4 were clustered as Group 4 in Figure 2 but are detected as outlier by the outliergram because of an atypical downward trend, and the black solid curve was in Group 6. For applications where both magnitude and shape heterogeneities are important, our method taking into account both is preferred. In situations where separation is important, the outliergram may be more suitable.

Identifying outliers in model-based clustering is hard, because the assessment of outlyingness depends on the model fit, which in turn is based on the outlying subjects (see Garcia-Escudero et al.[42] for various iterative approaches for doing this, which depends on some initialization). In contrast, in dissimilarity-based clustering, outliers can be identified based on the dissimilarities alone, without having to rely on knowing the clusters. The simulations above show a strong and stable performance of the proposed dissimilarity. It can also be used for a beneficial pre-cleaning step for model-based clustering.

5 Conclusion and discussion

We have shown that the proposed Smoothing Parameter Commutation dissimilarity is good at reproducing the true distances between noiseless curves that change gradually. On the real dataset, dissimilarity-based clustering on smoothed data is superior to model-based clustering under specific time series assumptions. The concept of the proposed dissimilarity measure is simple and easy to implement. We also demonstrated a simple method for outlier detection that helps model-based functional data clustering to form more compact subgroups.

There are many nonparametric regression methods other than smoothing splines, e.g. local polynomial regressions and wavelet analysis. Different techniques stand out in different situations. It is of interest to study whether there exist analogous parameter commutation operations and with similar advantages when applying other nonparametric regressions. This direction is left as a future work.

Acknowledgements

The authors thank the two anonymous reviewers and the editor for their helpful comments on improving the paper. The authors also thank Dr. Chieh-Liang Huang in China Medical University Hospital for permission to use the methadone dosage data.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: ST was supported in part by Taiwan Ministry of Science and Technology grant MOST103-2118-M-001-007-MY3. CH was supported by Engineering and Physical Sciences Research Council Grant EP/K033972/1. Y-FL was supported in part by China Medical University (grant CMU105-S-49) and Taiwan Ministry of Science and Technology (grant MOST 104-2314-B-039-037). CL was supported by Medical Research Council (grant number MC_UP_1302/4) and Cancer Research UK (grant number C48553/A18113).

References

1. Berkhin P. A survey of clustering data mining techniques. In: Kogan J, Nicholas C and Teboulle M (eds) *Grouping multidimensional data*. Berlin Heidelberg: Springer, 2006, pp.25–71.
2. Bouveyron C and Brunet-Saumard C. Model-based clustering of high-dimensional data: A review. *Comput Stat Data Anal* 2014; **71**: 52–78.
3. Murtagh F and Contreras P. Algorithms for hierarchical clustering: an overview. *Wiley Interdiscip Rev Data Min Knowl Discov* 2012; **2**: 86–97.
4. Hennig C, Meila M, Murtagh F, et al. (eds) *Handbook of cluster analysis*. Boca Raton, FL: Taylor & Francis, 2015.
5. Ferraty F and Vieu P. *Nonparametric functional data analysis: theory and practice*. New York: Springer Science & Business Media, 2006.
6. Hitchcock DB, Booth JG and Casella G. The effect of pre-smoothing functional data on cluster analysis. *J Stat Comput Simul* 2007; **77**: 1043–1055.

7. Ferreira L and Hitchcock DB. A comparison of hierarchical methods for clustering functional data. *Commun Stat Simul Comput* 2009; **38**: 1925–1949.
8. Alonso AM, Berrendero JR, Hernández A, et al. Time series clustering based on forecast densities. *Comput Stat Data Anal* 2006; **51**: 762–776.
9. Vilar JA, Alonso AM and Vilar JM. Non-linear time series clustering based on non-parametric forecast densities. *Comput Stat Data Anal* 2010; **54**: 2850–2865.
10. Maharaj EA. A significance test for classifying ARMA models. *J Stat Comput Simul* 1996; **54**: 305–331.
11. Abraham C, Cornillon PA, Matzner-Løber E, et al. Unsupervised curve clustering using B-splines. *Scand J Stat* 2003; **30**: 581–595.
12. James GM, Hastie TJ and Sugar CA. Principal component models for sparse functional data. *Biometrika* 2000; **87**: 587–602.
13. Warren Liao T. Clustering of time series data—a survey. *Pattern Recognit* 2005; **38**: 1857–1874.
14. Fan J and Zhang W. Generalised likelihood ratio tests for spectral density. *Biometrika* 2004; **91**: 195–209.
15. Caiado J, Crato N and Peña D. A periodogram-based metric for time series classification. *Comput Stat Data Anal* 2006; **50**: 2668–2684.
16. de Lucas DC. *Classification techniques for time series and functional data*. PhD Thesis, Universidad Carlos III de Madrid, 2010.
17. Brandmaier AM. *Permutation distribution clustering and structural equation model trees*. PhD Thesis, Saarland University, Saarbruecken, Germany, 2012.
18. Jones BL and Nagin DS. Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Sociol Methods Res* 2007; **35**: 542–571.
19. James GM and Sugar CA. Clustering for sparsely sampled functional data. *J Am Stat Assoc* 2003; **98**: 397–408.
20. Bouveyron C and Jacques J. Model-based clustering of time series in group-specific functional subspaces. *Adv Data Anal Classif* 2011; **5**: 281–300.
21. Jacques J and Preda C. Funclust: A curves clustering method using functional random variables density approximation. *Neurocomputing* 2013; **112**: 164–171.
22. Chiou JM and Li PL. Functional clustering and identifying substructures of longitudinal data. *J R Stat Soc Series B Stat Methodol* 2007; **69**: 679–699.
23. McNicholas PD and Murphy TB. Model-based clustering of longitudinal data. *Can J Stat* 2010; **38**: 153–168.
24. Delaigle A and Hall P. Defining probability density for a distribution of random functions. *Ann Stat* 2010; **38**: 1171–1193.
25. Wahba G and Wendelberger J. Some new mathematical methods for variational objective analysis using splines and cross validation. *Mon Weather Rev* 1980; **108**: 1122–1143.
26. Green PJ and Silverman BW. *Nonparametric regression and generalized linear models: a roughness penalty approach*. New York: CRC Press, 1993.
27. Wang Y. Smoothing spline models with correlated random errors. *J Am Stat Assoc* 1998; **93**: 341–348.
28. Speed T. Comment on ''That blup is a good thing: The estimation of random effects''. *Stat Sci* 1991; **6**: 42–44.
29. Krivobokova T and Kauermann G. A note on penalized spline smoothing with correlated errors. *J Am Stat Assoc* 2007; **102**: 1328–1337.
30. Hutchinson MF and De Hoog F. Smoothing noisy data with spline functions. *Numer Math* 1985; **47**: 99–106.
31. Ramsay J and Silverman B. Smoothing functional data with a roughness penalty. In: Ramsay J and Silverman B (eds) *Functional data analysis*. New York: Springer, 2005, pp.81–109.
32. Berndt DJ and Clifford J. Using dynamic time warping to find patterns in time series. *KDD-94: AAAI Workshop Knowledge Discov Databases* 1994; **10**: 359–370.
33. Gaffney SJ and Smyth P. Joint probabilistic curve clustering and alignment. In: Saul L, Weiss Y and Bottou L (eds) *Advances in neural information processing systems 17*. Cambridge, MA: MIT Press, 2004, pp.473–480.
34. Liu X and Yang MC. Simultaneous curve registration and clustering for functional data. *Comput Stat Data Anal* 2009; **53**: 1361–1376.
35. Montero P and Vilar JA. TSclust: An R package for time series clustering. *J Stat Softw* 2014; **62**: 1–43.
36. Chouakria AD and Nagabhushan PN. Adaptive dissimilarity index for measuring time series proximity. *Adv Data Anal Classif* 2007; **1**: 5–21.
37. Batista GE, Wang X and Keogh EJ. A complexity-invariant distance measure for time series. In: *Proceedings of the 11th SIAM international conference on data mining*, Mesa, Arizona, USA, 2011, pp.699–710.
38. Hardin J, Mitani A, Hicks L, et al. A robust measure of correlation between two genes on a microarray. *BMC Bioinform* 2007; **8**: 220.
39. Dudek MWA. clusterSim: Searching for optimal clustering procedure for a data set, https://CRAN.R-project.org/package=clusterSim. R package version 0.45-1 (2016, accessed 4 May 2017).
40. Kaufman L and Rousseeuw PJ. Partitioning around medoids (program pam). *Finding groups in data: an introduction to cluster analysis*. Hoboken,NJ,USA: John Willey & Sons, Inc., Wiley Online Library, 1990, pp.68–125.
41. Lin CJ, Hennig C and Huang CL. Clustering and a dissimilarity measure for methadone dosage time series. In: *Proceedings of ECDA-2014*. Bremen, Germany. Berlin: Springer, 2015, pp.31–41.

42. Garcia-Escudero LA, Gordaliza A, Matran C, et al. Robustness and outliers. In: Hennig C, Meila M, Murtagh F, et al. (eds) *Handbook of cluster analysis*. London: Taylor & Francis, 2015, pp.653–678.
43. Hubert M, Rousseeuw PJ and Segaert P. Multivariate functional outlier detection. *Stat Methods Appt* 2015; **24**: 177–202.
44. Hyndman RJ and Shang HL. Rainbow plots, bagplots, and boxplots for functional data. *J Comput Graph Stat* 2010; **19**: 29–45.
45. Tukey JW. Mathematics and the picturing of data. *Proc Int Congr Mathemat* 1975; **2**: 523–531.
46. Gervini D. Outlier detection and trimmed estimation for general functional data. *Stat Sin* 2012; **22**: 1639–1660.
47. Gervini D. Detecting and handling outlying trajectories in irregularly sampled functional datasets. *Ann Appl Stat* 2009; **3**: 1758–1775.
48. Sawant P, Billor N and Shin H. Functional outlier detection with robust functional principal component analysis. *Comput Stat* 2012; **27**: 83–102.
49. Arribas-Gil A and Romo J. Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics* 2014; **15**: 603–619.
50. Slaets L, Claeskens G and Hubert M. Phase and amplitude-based clustering for functional data. *Comput Stat Data Anal* 2012; **56**: 2360–2374.
51. Tucker JD, Wu W and Srivastava A. Generative models for functional data using phase and amplitude separation. *Comput Stat Data Anal* 2013; **61**: 50–66.
52. Ramaswamy S, Rastogi R and Shim K. Efficient algorithms for mining outliers from large data sets. *ACM SIGMOD Record* 2000; **29**: 427–438.
53. Angiulli F and Fassetti F. Distance-based outlier queries in data streams: the novel task and algorithms. *Data Min Knowl Discov* 2010; **20**: 290–324.
54. Goldstein M and Uchida S. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PloS One* 2016; **11**: e0152173.
55. Hennig C and Lin CJ. Flexible parametric bootstrap for testing homogeneity against clustering and assessing the number of clusters. *Stat Comput* 2015; **25**: 821–833.
56. Tibshirani R and Walther G. Cluster validation by prediction strength. *J Comput Graph Stat* 2005; **14**: 511–528.
57. Yassouridis C. *funcy: Functional Clustering Algorithms*, https://CRAN.R-project.org/package=funcy. R package version 0.8.6 (2017, accessed 4 May 2017).
58. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, https://www.R-project.org/ (2016, accessed 4 May 2017).