Bayesian Semiparametric Local Clustering of Multiple Time Series Data

Jingjing Fan and Abhra Sarkar
jfan25@utexas.edu and abhra.sarkar@utexas.edu
Department of Statistics and Data Sciences
The University of Texas at Austin,
Welch 5.216, 105 East 24th Street D9800, Austin, TX 78705, USA

Abstract

In multiple time series data, clustering the component profiles can identify meaningful latent groups while also detecting interesting change points in their trajectories. Conventional time series clustering methods, however, suffer the drawback of requiring the co-clustered units to have the same cluster membership for all the time points in the data collection period. In contrast to these 'global' clustering methods, we develop a Bayesian 'local' clustering method that allows the functions to flexibly change their cluster memberships over time. We design a Markov chain Monte Carlo algorithm to implement our method. We illustrate the method in several real-world data sets, where time-varying cluster memberships provide meaningful inferences about the underlying processes. These include a public health data set to showcase the more detailed inference our method can provide over global clustering methods, and a temperature data set to demonstrate our method's utility as a flexible change point detection method. Supplemental materials for this article, including R codes implementing the method, are available online.

Key Words: Change point detection, Hidden Markov model, Local clustering, Time series

1 Introduction

As technology advances, multiple time series data are being routinely collected in diverse domains.[1] Clustering such data can provide valuable insights into the underlying data-generating processes. To give a few examples, clustering gene expression profiles can identify groups of genes co-regulating biological processes (Song *et al.*, 2007; Ma *et al.*, 2008); clustering bidding behavior in online auctions can reveal patterns associated with specific chances of obtaining items at reasonable prices (Peng *et al.*, 2008); clustering climate data can reveal the effects of climate change (Gorji Sefidmazgi *et al.*, 2015); etc.

Traditional unsupervised time series clustering methods use distance measures to assign individual time series to clusters. Partitional clustering methods, such as K-means, select random series as cluster centroids and minimize the distance between clustered series and the centroids (Watanabe, 2022). Hierarchical clustering creates a hierarchy of clusters by merging or splitting clusters based on within-cluster similarity (Maharaj *et al.*, 2019). More sophisticated nonparametric approaches, such as the functional Dirichlet process (FDP), treat each time series as a realization of a stochastic process and induce clustering by drawing the functions from a discrete probability measure (Scarpa and Dunson, 2014; Nguyen and Gelfand, 2014). There are also semiparametric methods where the time series are clustered around the corresponding cluster mean curves modeled as linear combinations of basis elements, and the individual series are treated as noisy realizations of the cluster means Song *et al.* (2007); Ma *et al.* (2008).

Though these methods differ substantially in their implementation, they share the assumption that each time series can only belong to one cluster. This restrictive assumption can prevent the recovery of true local complexities, such as when groups merge together or diverge from one another. As a hypothetical but easy-to-understand example, consider a longitudinal study tracking the body temperature of multiple people over time. A global clustering framework could generate separate mean profiles for healthy individuals, individuals with a fever, and individuals who transition between health and fever. This could lead one to infer a very complex underlying process with many

[1] This article concentrates specifically on the analysis of multiple time series data where each constituent series pertains to the same variable. Such data may be obtained, e.g., as (a) multiple univariate time series from a set of different but comparable sources over the same time period; or (b) multiple records collected from the same source over different recurrent time cycles of the same length. The two real data sets analyzed in Section 3 of the main paper here correspond to one each of these two scenarios.

different states affecting people's body temperatures when in fact there are only two underlying states. For another example, consider also the synthetic scenario shown in Figure 1 below which mimics a real data application considered later in Section 5.1. Here, although the four latent mean trajectories are all different from each other when viewed over the entire time domain, they are often locally identical over smaller sub-regions. The local clustering method we propose here is based on such local similarities and differences between the constituent time series as opposed to global clustering based on their global behavior over the entire time domain.

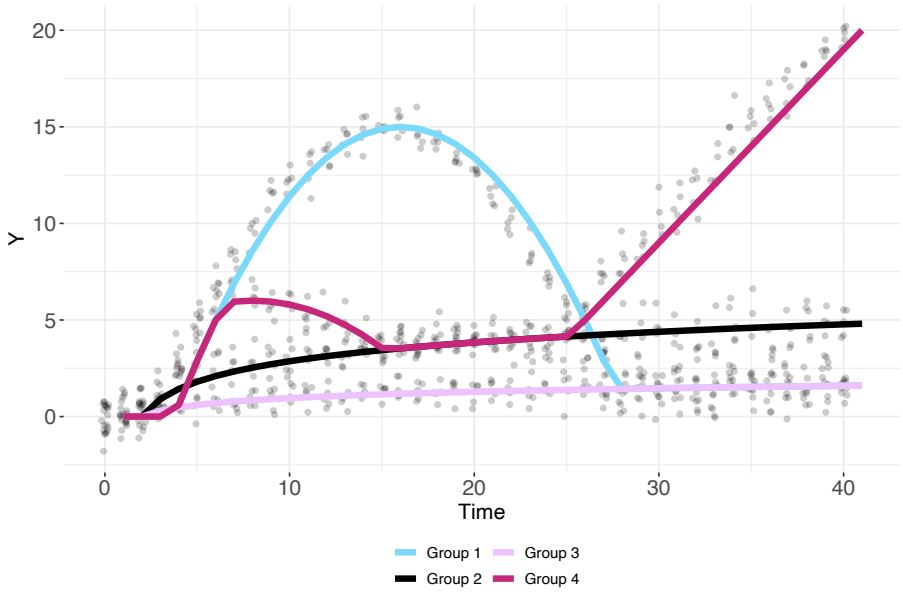


Figure 1: Simulated data (gray points) with true underlying mean functions overlaid. Group 1 is merged with Group 4 until $t = 5$, after which it splits off and maintains its own distinct profile until merging with Group 3 around $t = 27$. Group 4 begins merged with Group 1, splits off around $t = 5$, then merges with Group 2 for $t = [13, 24]$ after which it splits off and maintains its own identity for the remaining time.

A local clustering method that allows the latent mean curves to be differently clustered in different sub-domains has many advantages over a global clustering paradigm. For multiple time series data, where each constituent series corresponds to a different unit in the study, local clustering can eliminate redundant clusters by allocating two units to the same local cluster when they behave similarly and to two different local clusters when they behave differently.

This ability to locally cluster constituent time series also provides a natural framework for detecting interesting change points for time series with recurrent cycles. A time point within the cycles' domain is defined to be a change point when two or more series can be seen merging

or splitting at that point. Most existing methods for change point analysis for time series with cyclic patterns aim to identify change points such that the cycles that follow the change point are different from the cycles before in mean or variance. They often make restrictive assumptions, such as requiring that there be only one single change point in the entire measurement period, or requiring that sliding windows of a fixed size be specified to divide the time series into comparable segments, etc. (Aminikhanghahi and Cook, 2017; Reeves *et al.*, 2007; Li *et al.*, 2020). In contrast, the local clustering method proposed here is capable of detecting change points when the cycles deviate for only a portion of their domain. It requires minimal supervision and is able to detect multiple change points within the cycles' domain. To do this, we first cut the overall series at the points of recurrence to create multiple time series, then left-align the component series and apply our local clustering algorithm to identify dynamic changes in the cluster means. As we will see in applications, identifying change points in this manner often provides more nuanced and more useful information than what may be obtained using traditional methods.

The literature on local functional clustering is sparse. To our knowledge, existing methods are exclusively Bayesian and rely on variations and adaptations of Dirichlet process priors (Suarez and Ghosal, 2016; Petrone *et al.*, 2009; Nguyen and Gelfand, 2011). We propose a novel and computationally efficient method for local time series clustering that we call the hidden Markov functional local clustering model (HMFLCM). The HMFLCM works in two steps. In the first step, the data are liberally partitioned into global clusters. In the second step, the global cluster labels are locally re-clustered using a mixed-effects model. The data in each group are modeled by flexible mixtures of quadratic B-splines. To induce local clustering across different global cluster labels, we introduce latent cluster indicators for the basis coefficients corresponding to each global cluster label at each time point. These latent indicators are modeled using a hidden Markov model (HMM) where indicators for different pseudo-covariate levels at the same time point are allowed to take on the same value. At time points for which the latent indicators take on the same value, the basis coefficients for the corresponding fixed effects means are equal. This results in a mechanism where the cluster memberships as well as the number of clusters can change with time as the cluster means merge and split off as dictated by the data.

The methodology presented here is developed concurrently with the work of Paulon *et al.* (2023) and recycles some modeling techniques and ideas while also differing from them in a few key

ways in its motivation, detail, and application. The focus of Paulon *et al.* (2023) is on clustering the levels of categorical covariates to determine their local importance in longitudinal data with homoscedastic errors. In contrast, our work focuses on clustering a fixed number of time-series profiles with time-varying error variances in the complete absence of any covariates.

We use the HMFLCM to analyze a state-by-state Covid-19 daily case change data set to see whether state pandemic policies produced significant differences in the viral spread. We also demonstrate the HMFLCM's ability to detect change points using a long-range temperature data set from the National Climatic Data Center which measures the daily maximum temperature of Austin, Texas from 1938 to 2021. A third example, analyzing Federal Housing Finance Agency (FHFA) housing price index data for various metropolitan areas, is presented in the supplementary materials. We will see that, while global clustering can recover well-defined underlying groups, the global means can lack meaningful interpretations. By removing local redundancy and defining clear points of divergence between group means, much more meaningful cluster profiles can emerge.

The rest of this article is organized as follows. Section 2 develops the HMFLCM. Section 3 outlines posterior inference tools. Section 4 describes how the HMFLCM can be used for change-point analysis in a multiple time series setting. Section 5 demonstrates the analytical advantages of the HMFLCM in real-world applications. Section 6 demonstrates the accuracy of the HMFLCM on simulated data. Section 7 contains concluding remarks. Additional details, including the choice of hyper-parameters and Markov chain Monte Carlo (MCMC) algorithms for posterior inference, are deferred to the supplementary materials. R codes implementing our method are also included in the supplementary materials as separate files.

2 Functional Clustering

In this section, we develop our Hidden Markov Functional Local Clustering (HMFLCM) method. The HMFLCM is a two-step methodology that first uses global functional clustering to generate pseudo-covariate groupings, which are subsequently used during the local clustering step.

In our functional framework, sequentially generated data $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,T})^{\mathrm{T}}$ for series $i \in \{1, \ldots, n\}$ are assumed to be noisy realizations of some true group mean curves plus their associated subject-specific random effects curves. As will also be the case with our real-world examples, we

assume the data points to be measured at the common set of time points $t \in \{1, \ldots, T\}$. To simplify notation, we let t denote both specific time stamps as well as generic time points in the domain of interest $[1, T]$. The recovery of the underlying mean curves and the associated locally varying clustering patterns is the primary objective of our analysis.

In this section, we first discuss our global clustering method. We then build our local clustering mixed-effects model, describing in detail how the fixed effects, random effects, and heteroscedastic error components are specified. In what follows, the subscript g serves to indicate model components, parameters, and variables associated with global clustering. Likewise, the subscript ℓ signifies model components, parameters, and variables associated with local clustering.

2.1 Step 1: Global Functional Clustering

In this section, we present the methodology for the creation of a pseudo-covariate that partitions the subjects into identifiable groups for local clustering. Specifically, we seek to partition N series $\mathbf{y}_{i=1,\ldots,N}$ into D underlying global clusters, each with a functional mean profile $f_{g,z}(t)$ that describes the average behavior of the response curve over the entire period of observation $[1, T]$. Conditional on the global cluster assignment $z_{g,i} = z$ for subject i, we characterize each $\mathbf{y}_i$ as an error-prone realization of an unknown functional mean $f_{g,z}$ observed at T time points,

$$(y_{i,t}|z_{g,i} = z) = f_{g,z}(t) + \varepsilon_{g,i,t}, \qquad \varepsilon_{g,i,t} \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2_{g,\varepsilon}). \tag{1}$$

We assume the errors, $\varepsilon_{g,i,t}$, to be independent and identically distributed to keep the model simple and the mean profiles easily identifiable. The main purpose of this first clustering step is to recover a liberal number of global clusters to be merged in the second local clustering stage. In all our experiments, the global model described in Equation (1) served that purpose sufficiently well.

Let the cluster membership variables be independently Dirichlet distributed as

$$z_{g,i} \sim \text{Dir}(\alpha_g, \ldots, \alpha_g).$$

We model the mean profiles $f_{g,z}(t)$ as mixtures of quadratic B-spline bases (Figure 2) as

$$f_{g,z}(t) = \textstyle\sum_{k=1}^{K_g} \beta_{g,z,k} b_{g,k}(t),$$

where $\mathbf{b}_g(t) = \{b_{g,1}(t), \ldots, b_{g,K_g}(t)\}^{\mathrm{T}}$ are a set of B-spline basis functions and $\boldsymbol{\beta}_{g,z} = \{\beta_{g,z,1}, \ldots, \beta_{g,z,K_g}\}^{\mathrm{T}}$ are a set of unknown coefficients to be estimated from the data. We choose $\mathbf{b}_g(t)$ to be a set of evenly spaced quadratic B-splines. Quadratic B-splines are easy to compute and are a popular choice for modeling flexibly varying smooth functions (de Boor, 1978).

Finally, we adopt hierarchical conditionally conjugate priors on the spline coefficients and the error variance terms

$$\boldsymbol{\beta}_{g,z} \sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_{g,\beta}), \quad \boldsymbol{\Sigma}_{g,\beta} \sim \mathrm{IW}(\nu_{g,0}, \mathbf{S}_{g,0}), \quad \sigma^2_{g,\varepsilon} \sim \text{Inv-Ga}(a_{g,\epsilon}, b_{g,\epsilon}).$$

Posterior inference is based on samples drawn using a Markov Chain Monte Carlo (MCMC) sampling scheme. The full conditionals for the parameters are listed in the supplementary materials.

To account for the possibility of label switching, we use the posterior co-clustering probabilities to determine the final global cluster assignments. The posterior co-clustering probability of the global groups is empirically computed by evaluating the proportion of posterior samples which cluster certain series together. The global cluster memberships are then treated as the levels of a pseudo-covariate for the purpose of local clustering, in which subjects belonging to different pseudo-covariate levels may merge into the same group at local neighborhoods in time. While it is desirable that the global and the local groups be learned simultaneously, it is not easy to achieve this in an aggregated posterior sampling scheme as the local groupings become difficult to identify when the subjects can potentially also change their global assignments in each draw.

The correctness of the overall local clustering results is dependent upon the global clustering method's ability to 1) correctly cluster together individuals that belong to the same global group but also more importantly 2) separate the individuals that do not belong to the same global group even at the expense of overestimating the number of global clusters. To illustrate this, we varied the number of spline knot points, K_g, and the maximum allowable number of global clusters, denoted henceforth by D, and evaluate the impact of these hyperparameters on the correctness of both the global and local clustering in Table 3 in Section 6. Additionally, while it is not necessary to use a B-spline mixture for the global clustering step, we find that the B-spline method produces global clusters better suited for local clustering than alternate methods such as K-means. We evaluate the impact of alternative global clustering methods on local clustering in the supplementary materials.

Consistent with Ruppert (2002), we found that when the smoothness is controlled by a data-

adaptive penalty parameter, the results are robust to the choice of K_g as long as a minimum number of knots is used. We use the Watanabe-Akaike information criterion (WAIC) (Watanabe and Opper, 2010) as well as leave-one-out cross-validation (LOO-CV) (Gelman *et al.*, 2014) to choose such lower bounds.

We also found that accurate local clustering results are achieved when D is over-specified. Similar to redundant mixture components in overfitted mixture models (Rousseau and Mengersen, 2011), superfluous latent classes become empty during the global clustering step if the Dirichlet prior on the class proportions is sufficiently vague. Additionally, by construction, the local clustering step can also merge redundant global clusters.

2.2 Step 2: Local Functional Clustering

A mixed-effects functional clustering method was recently introduced by Paulon *et al.* (2023) which performs local clustering of longitudinal functional profiles associated with different levels of an associated observed categorical covariate. Local clustering of the fixed effects was induced by allowing portions of the different fixed effects curves to merge. We adapt this method with some simplifications to locally cluster functional time series data in the complete absence of any observed covariate, where we instead treat the d non-empty global clusters recovered in stage one as the levels of a categorical pseudo-covariate $z_g \in \mathcal{Z}_g = \{1, \ldots, d\}$. Note that $d \leq D$ by design. The global cluster membership value of each subject is stored in the variable $z_{g,i}$, which doubles as the i-th series' associated pseudo-covariate assignment.

With the existence of a fixed pseudo-covariate assignment, a mixed-effects analysis becomes viable. The fixed effects describe the mean behavior of different pseudo-covariate levels, while the deviations of the individuals within each group are explained by the associated random effects. The random effects explain some of the variability at the individual level, allowing for smooth fixed effects recovery and meaningful fixed effecting merging. This is formalized as the following generic class of functional mixed models

$$(y_{i,t} \mid z_{g,i} = z_g) = f_{\ell,z_g}(t) + u_{\ell,i}(t) + \varepsilon_{\ell,i,t}, \quad u_{\ell,i}(t) \sim f_{\ell,u}, \quad \varepsilon_{\ell,i,t} \overset{ind}{\sim} f_{\ell,\varepsilon,t}, \tag{2}$$

where $f_{\ell,z_g}(t)$ is the time-varying fixed effects function corresponding to pseudo-covariate level z_g, $u_{\ell,i}(t)$ are subject-specific time-varying random effects with random effects distribution $f_{\ell,u}$, and

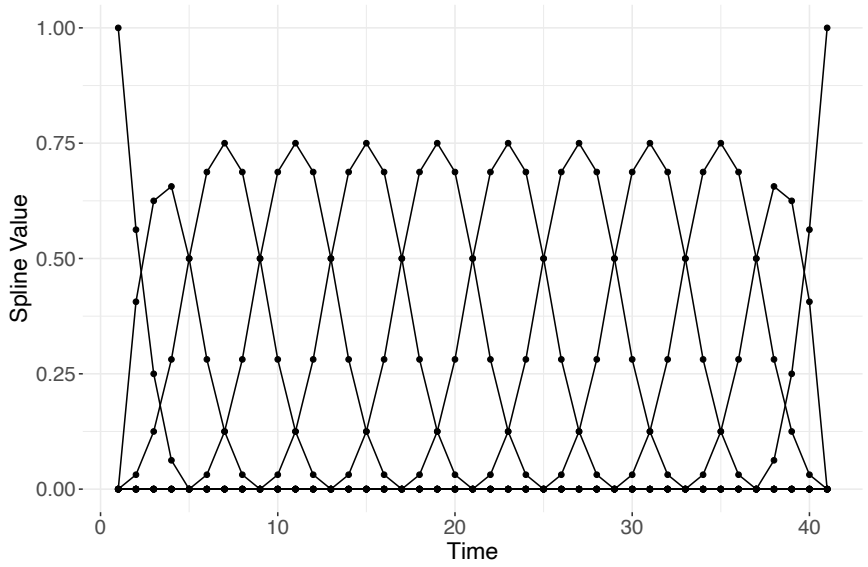


Figure 2: An illustration of quadratic B-splines used in our method. In this figure, they are created with internal knot points corresponding to the deciles of the observed time locations such that in total we have $K = 12$ bases.

$\epsilon_{\ell,i,t}$ are random errors with distribution $f_{\ell,\varepsilon,t}$.

2.2.1 Locally Varying Fixed Effects

Local clustering of the fixed effects is controlled by an HMM. Let the fixed effects for each z_g be constructed as

$$f_{\ell,z_g}(t) = \sum_{k=1}^{K_\ell} \beta_{\ell,z_g,k} b_{\ell,k}(t), \tag{3}$$

where $\mathbf{b}_\ell(t) = \{b_{\ell,1}(t), \ldots, b_{\ell,K_\ell}(t)\}^{\mathrm{T}}$ are a dense set of $K = T + 1$ quadratic B-splines with a knot point at every observed time location $t \in \{1, \ldots, T\}$, and $\boldsymbol{\beta}_{\ell,z_g} = \{\beta_{\ell,z_g,1}, \ldots, \beta_{\ell,z_g,K_\ell}\}$ are unknown coefficients to be estimated from the data. Choosing the knot points in this manner allows the cluster mean curves $f_{\ell,z_g}(t)$ for different values of z_g to merge or split at every observed time location. Thus, $K = T + 1$ is the total number of bases and the value of $f_{\ell,z_g}(t)$ at each $t = k$ is dependent upon the values of the $\{\beta_{\ell,z_g,k}, \beta_{\ell,z_g,(k+1)}\}$, each with a weight of 1/2 (Figure 3).

For two levels $z_g \neq z'_g$ of the pseudo-covariate, local clustering in the fixed effects is induced if at location t, $f_{\ell,z_g}(t) = f_{\ell,z'_g}(t)$. At other locations t' the fixed effects can be different, that is, $f_{\ell,z_g}(t') \neq f_{\ell,z'_g}(t')$. This is achieved by allowing elements of $\boldsymbol{\beta}_{\ell,z_g}$ and $\boldsymbol{\beta}_{\ell,z'_g}$ to be shared $\beta_{\ell,z_g,k} = \beta_{\ell,z'_g,k}$ at some locations k but be different $\beta_{\ell,z_g,k'} \neq \beta_{\ell,z'_g,k'}$ at the other locations k'.

The sharing of spline coefficients can be facilitated across any combination of pseudo-covariate levels and is not limited to the equivalence between only two levels of z_g. To facilitate such local clustering, a set of local latent variables $z_{\ell,k}^{(z_g)} \in \{1, \ldots, d\}$ denoting the local cluster assignment of a group z_g at knot point k is introduced such that

$$\begin{aligned} &\{f_{\ell,z_g}(t) \mid z_{\ell,k}^{(z_g)} = z_{\ell,k}\} = \textstyle\sum_{k=1}^{K_\ell} \beta^\star_{\ell,z_{\ell,k},k} b_{\ell,k}(t), \\ &\text{or equivalently,} \quad (\beta_{\ell,z_g,k} \mid z_{\ell,k}^{(z_g)} = z_{\ell,k}) = \beta^\star_{\ell,z_{\ell,k},k}. \end{aligned} \tag{4}$$

Let $\mathcal{Z}_{\ell,k} = \{z_{\ell,k} : z_{\ell,k}^{(z_g)} = z_{\ell,k}, z_g = 1, \ldots, d\}$ and $\mathcal{B}^\star_{\ell,k} = \{\beta^\star_{\ell,z_{\ell,k},k} : z_{\ell,k} \in \mathcal{Z}_{\ell,k}\}$. The set of B-spline coefficients to be estimated at location k is then $\mathcal{B}^\star_{\ell,k}$. When the $z_{\ell,k}^{(z_g)}$'s are assigned probability models supported on $\mathcal{Z}_g$, the number of local clusters, that is, the number of distinct values taken on by the $z_{\ell,k}^{(z_g)}$'s, $|\mathcal{Z}_{\ell,k}|$, must be less than the number of global clusters $|\mathcal{Z}_g| = d$. Specifically, when $z_{\ell,k}^{(z_1)} = z_{\ell,k}^{(z_2)}$ for two different levels z_1 and z_2 of z_g, the spline coefficients at location k are equal between the groups z_1 and z_2. Using quadratic B-splines, there are two splines contributing the function at each time point. At time point $t = k$, if both $z_{\ell,k}^{(z_1)} = z_{\ell,k}^{(z_2)}$ and $z_{\ell,(k+1)}^{(z_1)} = z_{\ell,(k+1)}^{(z_2)}$, the value of the underlying fixed effects curves are equal, i.e., $f_{\ell,z_1}(t) = f_{\ell,z_2}(t)$. The series belonging to the global clusters z_1 and z_2 are then co-clustered at time $t = k$. If, on the contrary, the $z_{\ell,k}^{(z_g)}$'s corresponding to two different levels of z_g are different, the underlying curves there will also be different.

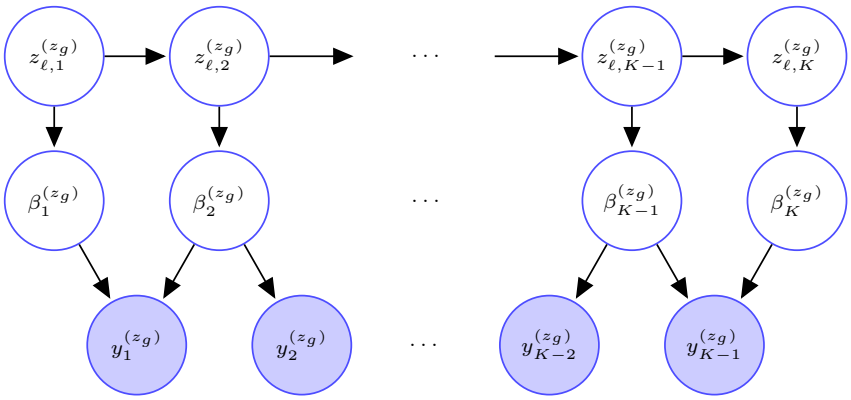


Figure 3: Graph of functional HMM model (4) using quadratic B-splines.

We now consider the problem of specifying probability models for the $z_{\ell,k}^{(z_g)}$'s and the $\beta^\star_{\ell,z_{\ell,k},k}$'s that appropriately accommodate temporal dependencies across k. The temporal evolution of the latent local cluster indicators $z_{\ell,k}^{(z_g)}, k = 1, \ldots, K$, are modeled using HMMs as

$$(z_{\ell,k}^{(z_g)} \mid z_{\ell,k-1}^{(z_g)} = z_{\ell,k-1}) \sim \text{Mult}(\pi_{z_{\ell,k-1},1}^{(z_g)}, \ldots, \pi_{z_{\ell,k-1},d}^{(z_g)}).$$

Hierarchical Dirichlet priors are assigned to the transition probabilities with gamma hyper-priors on the concentration parameters

$$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_d)^{\mathrm{T}} \sim \text{Dir}(\alpha/d, \ldots, \alpha/d), \quad \alpha \sim \text{Ga}(a, b).$$

The priors for the atoms $\beta^{\star}_{\ell,z_{\ell,k},k}$, conditional on the $z_{\ell,k}^{(z_g)}$'s and the coefficients at the previous locations, for $k = 2, \ldots, K$, are constructed sequentially as in Paulon *et al.* (2023) as

$$\beta^{\star}_{\ell,z_{\ell,k},k} \sim \begin{cases} \prod_{\{z_{\ell,k-1}^{(z_g)}:\ z \in \mathfrak{Z}_{\ell,k}^{(z_{\ell,k})}\}} \text{Normal}\left(\beta^{\star}_{\ell,k-1,z_{\ell,k-1}^{(z_g)}}, \sigma^2_{\beta,1}\right) & \text{if } |\mathfrak{Z}_{\ell,k}^{(z_{\ell,k})}| > 0, \\ \text{Normal}(\mu_{\beta,0}, \sigma^2_{\beta,0}) & \text{otherwise,} \end{cases}$$

where $\mathfrak{Z}_{\ell,k}^{(z_{\ell,k})} = \{z_g : z_{\ell,k}^{(z_g)} = z_{\ell,k}\}$ is the set of levels of z_g that, at the location k, are assigned the label $z_{\ell,k}$. The prior in (5) centers the B-spline coefficients around the ones 'expressed' at the previous location, penalizing their first-order differences. More specifically, the B-spline coefficient characterizing a local mean curve at knot location $k+1$ is encouraged to be close to the coefficient characterizing that same local mean curve at the previous knot point k. This helps ensure that the fixed effects curves are smooth and robust to outliers.

The coefficients that are not associated with any levels of z_g are assigned a normal prior with a large variance $\sigma^2_{\beta,0}$, which we set at 5 times the variance of all data points. The initial coefficients are assigned non-informative flat priors as $\beta^{\star}_{1,z_{\ell,k}} \sim 1$. The smoothness of the curves is controlled by the parameter $\sigma^2_{\beta,1}$, which is assigned a prior, allowing it to be informed by the data. Let

$$\sigma^2_{\beta,1} \sim \text{C}^{+}(0, 1),$$

where $\text{C}^{+}(a, b)$ denotes a half-Cauchy distribution (Gelman, 2006; Polson and Scott, 2012) with location parameter a and scale parameter b.

Borrowing information 'vertically' across subjects and 'horizontally' across time points, we are able to efficiently estimate the fixed effects as seen in Section 5.

2.2.2 Locally Varying Random Effects

We model the time-varying random effects components $u_{\ell,i}(t)$ as

$$\begin{aligned} &u_{\ell,i}(t) = \textstyle\sum_{k=1}^{K_\ell} \beta_{\ell,u,i,k} b_{\ell,k}(t), \\ &\boldsymbol{\beta}_{\ell,u,i} \sim \text{MVN}_{K_\ell}\{\mathbf{0}, (\sigma_{u,a}^{-2}\mathbf{I}_{K_\ell} + \sigma_{u,s}^{-2}\mathbf{P}_u)^{-1}\}, \\ &\sigma_{u,s}^2 \sim \text{C}^+(0,1), \qquad \sigma_{u,a}^2 \sim \text{C}^+(0,1), \end{aligned} \tag{5}$$

where $\boldsymbol{\beta}_{\ell,u,i} = (\beta_{\ell,u,i,1}, \ldots, \beta_{\ell,u,i,K_\ell})^{\text{T}}$ are series-specific spline coefficients, and $\text{MVN}_{K_\ell}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a K dimensional multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. We let, $K_\ell = T+1$ for the random effects curves as well. We choose $\mathbf{P}_u = \mathbf{D}_u^{\text{T}}\mathbf{D}_u$, where the $(K-1)\times K$ matrix $\mathbf{D}_u$ is a matrix such that $\mathbf{D}_u\boldsymbol{\beta}_{\ell,u,i}$ computes the first order differences in $\boldsymbol{\beta}_{\ell,u,i}$.

The model thus penalizes $\sum_{k=1}^{K}(\nabla\beta_{\ell,u,i})^2 = \boldsymbol{\beta}_{\ell,u,i}^{\text{T}}\mathbf{P}_u\boldsymbol{\beta}_{\ell,u,i}$, the sum of squares of first order differences in $\boldsymbol{\beta}_{\ell,u,i}$ (Eilers and Marx, 1996). The random effects variance parameter $\sigma_{u,s}^2$ models the smoothness of the random effects curves, smaller $\sigma_{u,s}^2$ inducing smoother $u_{\ell,i}(t)$'s. Additional variations from the constant zero curve are explained by $\sigma_{u,a}^2$. The absence of random effects is signified by the limiting case $\sigma_{u,s}^2 = \sigma_{u,a}^2 = 0$.

The prior hyper-parameters of our model are listed in Section S.1 of the supplementary materials. They are all chosen to keep the priors non-informative relative to the likelihood.

2.2.3 Heteroscedastic Random Errors

We assume the errors to be independently normally distributed with time-varying variance as

$$\varepsilon_{\ell,i,t} \overset{ind}{\sim} \text{Normal}\{0, \sigma_{\ell,\varepsilon}^2(t)\} \tag{6}$$

We model $\sigma_{\ell,\varepsilon}^2(t)$ using another smoothly varying mixture of quadratic B-splines as

$$\sigma_{\ell,\varepsilon}^2(t) = \textstyle\sum_{k=1}^{K_{\sigma^2}} \exp(\beta_{\sigma^2,k}) b_{\sigma^2,k}(t), \qquad \boldsymbol{\beta}_{\sigma^2} \sim \text{MVN}(0, \tau^2\mathbf{P}_{\sigma^2}). \tag{7}$$

Similar to the global clustering basis specification, we let $\mathbf{b}_{\sigma^2}(t) = \{b_{\sigma^2,1}(t), \ldots, b_{\sigma^2,K_{\sigma^2}}(t)\}$ be a set of quadratic B-splines with knot points corresponding the deciles of the time points. We set K_{σ^2} to the deciles of the time points to allow for flexible but smooth variance fluctuations. We choose $\mathbf{P}_{\sigma^2} = \mathbf{D}_{\sigma^2}^{\text{T}}\mathbf{D}_{\sigma^2}$, where the $(K_{\sigma^2}-1)\times K_{\sigma^2}$ matrix $\mathbf{D}_{\sigma^2}$ is such that $\mathbf{D}_{\sigma^2}\boldsymbol{\beta}_{\sigma^2}$ computes the

first order differences in $\boldsymbol{\beta}_{\sigma^2}$. The model thus penalizes $\sum_{k=1}^{K_{\sigma^2}}(\nabla\beta_{\sigma^2})^2 = \boldsymbol{\beta}_{\sigma^2}^{\mathrm{T}}\mathbf{P}_{\sigma^2}\boldsymbol{\beta}_{\sigma^2}$, the sum of squares of first order differences in $\boldsymbol{\beta}_{\sigma^2}$ (Eilers and Marx, 1996). The parameter τ^2 controls the smoothness of the dynamic variance. Smaller values of τ^2 result in a dynamic variance curve with a smaller range. We assign τ^2 a mildly informative Inv-Ga(a_{τ^2}, b_{τ^2}) prior but found the posterior of $\sigma^2_{\ell,\varepsilon}(t)$ to be fairly robust to changes in the values of τ^2. We show our model's ability to correctly recover the dynamic error variance structure of our simulation experiment in the supplementary materials. In additional numerical experiments reported in the supplementary materials, we show that allowing the error variance $\sigma^2_{\ell,\varepsilon}(t)$ to vary over time can significantly improve local clustering results by safeguarding against over-fitting the fixed and the random effects. An evaluation of the HMFLCM's robustness, when the true error variance stays constant over time, is also presented in the supplementary materials.

3 Posterior Inference

Posterior inference for the proposed functional mixed model framework is based on samples drawn from the posterior using a dynamic message passing MCMC algorithm which carefully exploits the conditional independence relationships encoded in the model. Since the levels of z_g create a partition of the data set according to the pseudo-covariate level, for different values of z_g, the latent sequences $\mathbf{z}_\ell^{(z_g)} = \{z_{\ell,k}^{(z_g)}, k = 1, \ldots, K\}$ also associate with disjoint data blocks and are conditionally independent in the posterior. The sequences can thus be updated using separate message-passing schemes for different values of z_g. That is, the HMM update for the latent sequence $\mathbf{z}_\ell^{(z_g)}$ corresponding to each z_g is dependent only on the data and parameters associated with that z_g, leading to d conditionally independent update blocks. Details have been deferred to Section S.2 in the supplementary materials.

Lastly, while our method is technically clustering curves, it does so by clustering the latent variables associated with each spline coefficient at each time point, and so the latent space scales with the number of global clusters and the number of knot points. The recursive message passing scheme we used to implement our functional hidden Markov model grows only quadratically with the number of global clusters and linearly with time. For the Covid-19 data set featuring 50 states and Washington DC, each with 156 time points, analyzed in Section 3.1 below, the run time for

1000 global clustering iterations, and 3000 local clustering iterations on a standard desktop with an Intel(R) Core(TM) i7-8700 3.2GHz with 16.0 GB of RAM was approximately 30 minutes.

4 Change Point Analysis via Local Clustering

The proposed local clustering method provides a principled albeit nontraditional framework for detecting meaningful change points for time series with recurring cycles. To apply the HMFLCM, each cycle within the larger series is treated to constitute a component in a multiple time series. These individual series can then be locally clustered to find local change points which affect only portions of each cycle. Here we define a time point to be a change point whenever a merging or a splitting behavior is observed for the local mean curves associated with different global cluster indices at that time point. More specifically, a time point t_0 is a change point if there are two global clusters, without loss of generality, say $z_g = 1$ and $z_g = 2$, for which either $f_{\ell,1}(t) = f_{\ell,2}(t)$ immediately before t_0 but $f_{\ell,1}(t) \neq f_{\ell,2}(t)$ immediately after t_0 (i.e., the mean curves split off), or conversely $f_{\ell,1}(t) \neq f_{\ell,2}(t)$ immediately before t_0 but $f_{\ell,1}(t) = f_{\ell,2}(t)$ immediately after t_0 (i.e., the mean curves merged together).

Alternatively, one could think of running a change point analysis on each constituent time series separately and then clustering the resulting fits. It would however be hard to cluster the sub-segments, especially when they do not line up. It is also not guaranteed that a change point will be identified for any individual series in the first place. More importantly, such an analysis, when possible, while potentially useful in many applications, would answer a different question from the one we focus on in the applications in Section 5. Traditional change point analysis identifies points for which the mean or variance of the entire series changes (Aminikhanghahi and Cook, 2017; Reeves *et al.*, 2007; Li *et al.*, 2020). However, we are primarily interested in capturing the local changes between different cycles within the same series. In particular, for recurrent events data, where each constituent time series is defined as one iteration of a recurring cycle, we want to find change points that occur when the trajectory changes significantly between different cycles. In other words, we are interested in how the cycles differ from each other and where in the cycles these differences occur, but not in the individual change points within each cycle.

5 Applications

In this section, we discuss the results obtained by our method when applied to two real-world data sets from two different application domains, public health and climate science. In the public health application, we focus on demonstrating how the ability to partition data into flexible time groups reveals interesting patterns that align with reasonable explanations arising from domain knowledge. In the climate science application, we focus on demonstrating how local clustering can also detect meaningful change points in the constituent trajectories, which can potentially provide very useful information for energy engineers.

5.1 Covid-19 Data Set

This data set tracks the daily change in confirmed cases of Covid-19 in every state in the United States plus Washington DC. The data comes from the Johns Hopkins University Center for System Science and Engineering (JHU CSSE).

JHU CSSE collects their confirmed cases from a number of sources, including the World Health Organization, and the U.S. Centers for Disease Control and Prevention. The confirmed cases are tracked at the county level, so we aggregated the data to compare caseloads by state. For each state, we use the number of new daily cases per hundred thousand people to create a comparable rate of contagion across states.

At the time of writing of the first version of this article, there were few scientifically proven policies for controlling the spread of the SARS-COV2 virus responsible for the Covid-19 pandemic. As such, the daily case-load data provide a rough benchmark for the efficacy of epidemiological policies instigated in response to the pandemic. The United States presents a uniquely interesting case study in disease control because the patchwork, state-by-state nature of the U.S. Covid-19 response allows many different response strategies to be easily juxtaposed and compared.

Below we cluster the 51 (50 states plus Washington D.C.) temporal profiles, each collected from January, 22nd 2020 to June, 29th 2020. We use seven latent classes for the global clustering stage as determined by the LOO-CV metric. Table 1 lists the states belonging to each of these global clusters. Figure 4 shows the functional means recovered by the global and local clustering steps. The estimated posterior co-clustering probabilities in Figure 5 quantify the evolution of local

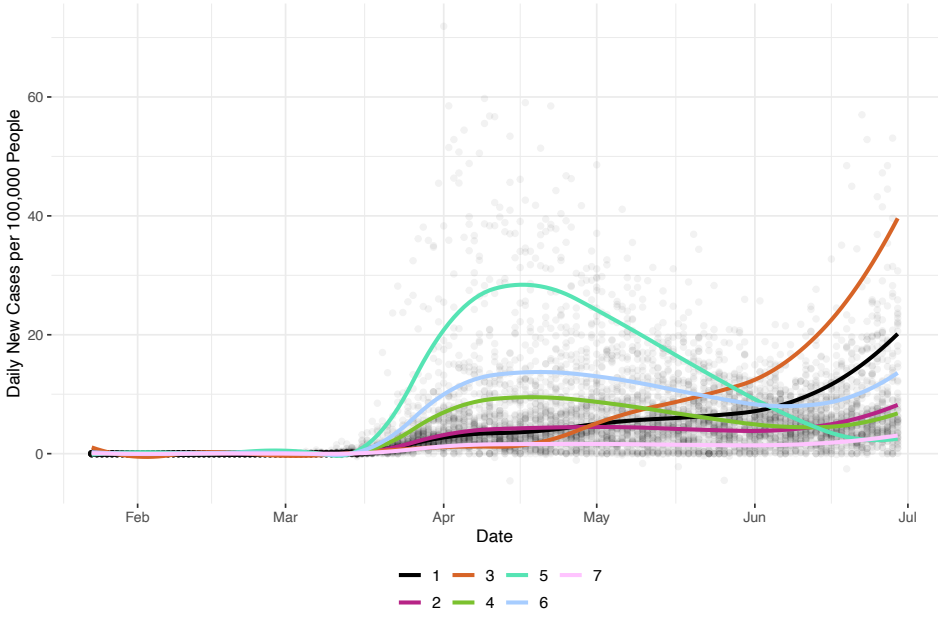
co-clustering assignments over time.

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
Alabama	Alaska	Arizona	Colorado	Connecticut	Delaware	Idaho
Arkansas	Hawaii		Indiana	Massachusetts	Illinois	Kansas
California	Maine		Michigan	New Jersey	Iowa	Kentucky
Florida	Montana		Pennsylvania	New York	Louisiana	Minnesota
Georgia	Oregon		South Dakota	Rhode Island	Maryland	Missouri
Mississippi	Vermont		Virginia	District of Columbia	Nebraska	New Hampshire
Nevada	West Virginia					New Mexico
North Carolina	Wyoming					North Dakota
South Carolina						Ohio
Tennessee						Oklahoma
Texas						Washington
						Wisconsin

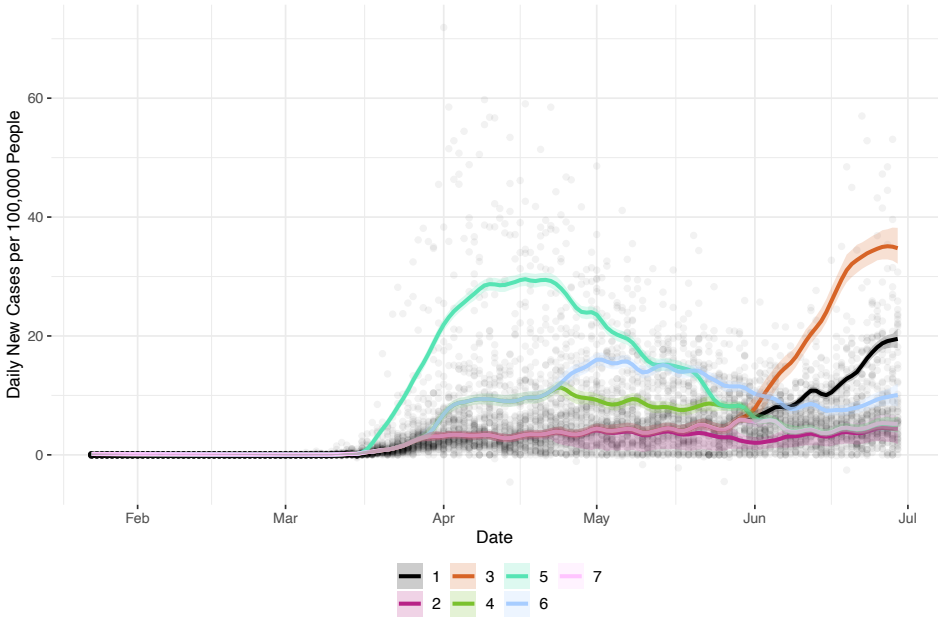
Table 1: The global clusters as found by step one of the algorithm. Each group functions as a level of the pseudo-covariate used for local clustering. Refer to Figure 4b for local clustering patterns.

The mean profile of each group can be seen in Figure 4a and the group memberships are listed in Table 1. The states in Groups 2 and 7 maintain the lowest and flattest mean profiles. This may be explained by the fact that the states in Group 2 have relatively low population densities and states in Group 7 mostly sit in the center of the contiguous United States. The states in Group 1 have a higher average caseload but do not exhibit a large increase in cases in the month of June. Groups 1 and 3 feature states whose caseloads increase sharply throughout June. Groups 4 and 5 contain states that had higher new case rates in April but had reduced their contagion rate by June.

Although the information on geographical closeness was not incorporated in our model, Group 5 contains many geographically proximate states with New York at its center. This is not surprising, given the high volume of traffic between these states, which likely contributed to their similar case trajectories. Although the global clustering method effectively partitions states by their distinctive overall trajectories, the local clustering results provide a clearer picture of how changes in the pandemic policy of certain states produced different caseload trends in those states. The results of the local clustering method lend support to the argument that many states opened nonessential businesses prematurely, leading to a significant uptick in the spread of the virus. This conclusion is largely consistent with medical news reports from that time (Berger, 2020). Specifically, the method was able to identify many of the first reopened states that opened non-essential businesses on or soon after May 1st (Lee *et al.*, 2020). Texas, California and Florida (Group 1) saw marked increases



(a) Global clusters



(b) Local clusters with 95% credible intervals

Figure 4: Global versus local clustering means for the Covid-19 data set. The local clustering mean curves support a different interpretation than the global clustering means. While both figures display an increasing trend for Groups 1 and 3, the local clustering means show the divergence of the Groups 1 and 3 means from their previously merged clusters in June, indicating that the states belonging to these groups started behaving significantly differently around that time. This information is not clear from the global clustering results alone.

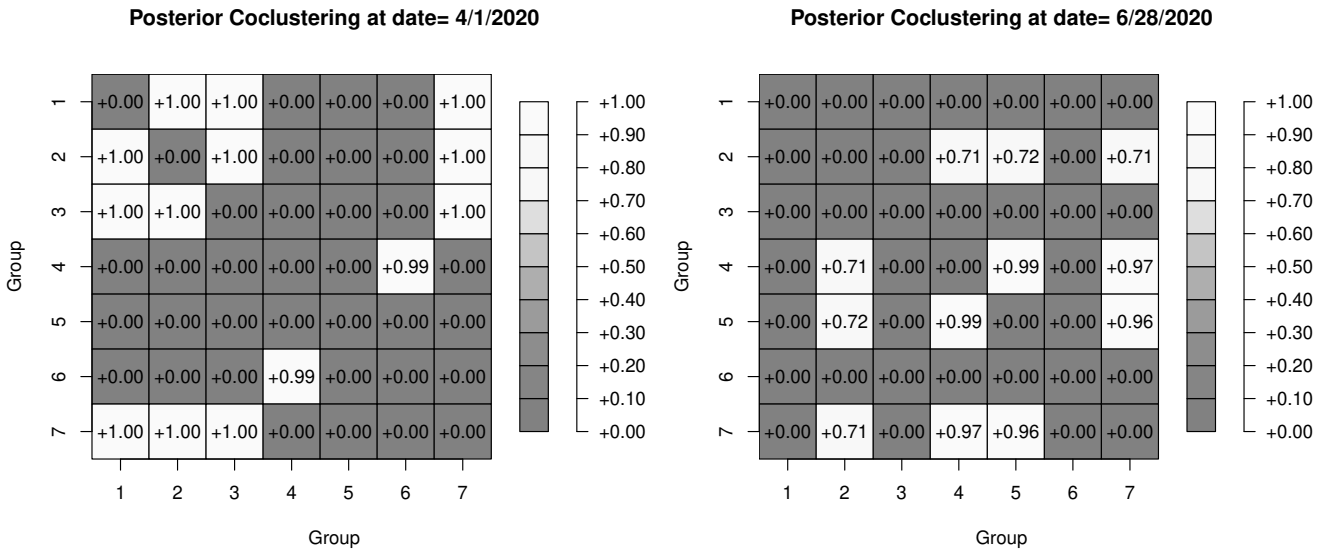


Figure 5: Results for Covid-19 Data: Co-clustering patterns changed remarkably between April and June, 2020. While it is interesting to observe the merging changes in the fixed effects means, posterior co-clustering probabilities such as the ones in this figure give a more quantitative picture of how the states' co-clustering behavior changes over time.

in confirmed cases a month after relaxing shutdown measures. Arizona (Group 3) experienced a particularly rapid rate of increase in new cases following reopening. Considering the 2-14 day incubation period for the SARS-COV2 virus (Center for Disease Control, 2020), the typical delay between the implementation of a new policy and a visible, large-scale effect on the populace is approximately one month. The dates corresponding to the divergence in group means for Groups 1 and 3 are in line with the expected lag following the implementation of their divergent reopening policies. Our clustering results thus provide empirical evidence that premature state reopening was an important factor contributing to the fast spread of the virus.

5.2 Austin Temperature Data Set

This data set comes from the National Oceanic and Atmospheric Administration (NOAA) and can be accessed using the Climate Date Online tool (https://www.ncei.noaa.gov/cdo-web/search). The data set includes the daily maximum temperature recorded at the Camp Mabry weather station located in Austin, Texas, from January 1, 1938, to December 31, 2021. We segment the daily

temperature series by year to form 83 yearly cycles. For simplicity, we truncate some of the series so that there are 365 days in each year. We let $D = 6$ and $K_g = 12$ for the global clustering step. The goal is to determine if there are significant differences between the temperature series of different years, and if so, during what parts of the year are such differences the most pronounced.

Accurate determination of change points in the temperature record is important for climate scientists to distinguish between detectable changes in climatic trends and short-term stochastic variations inherent in natural systems (Yu and Ruggieri, 2019). Pinpointing the changing trends in temperature is also an important problem for the energy sector. In hotter climates such as Texas, air conditioning energy loads make up a significant portion of overall electric grid demand. Perez *et al.* (2017) showed that AC energy use increases linearly with outdoor dry-bulb temperatures once temperatures are higher than a critical degree. Therefore, accurate forecasting of when the outdoor temperatures will be high is critical for electricity providers to plan their generation capacity appropriately and avoid costly mistakes.

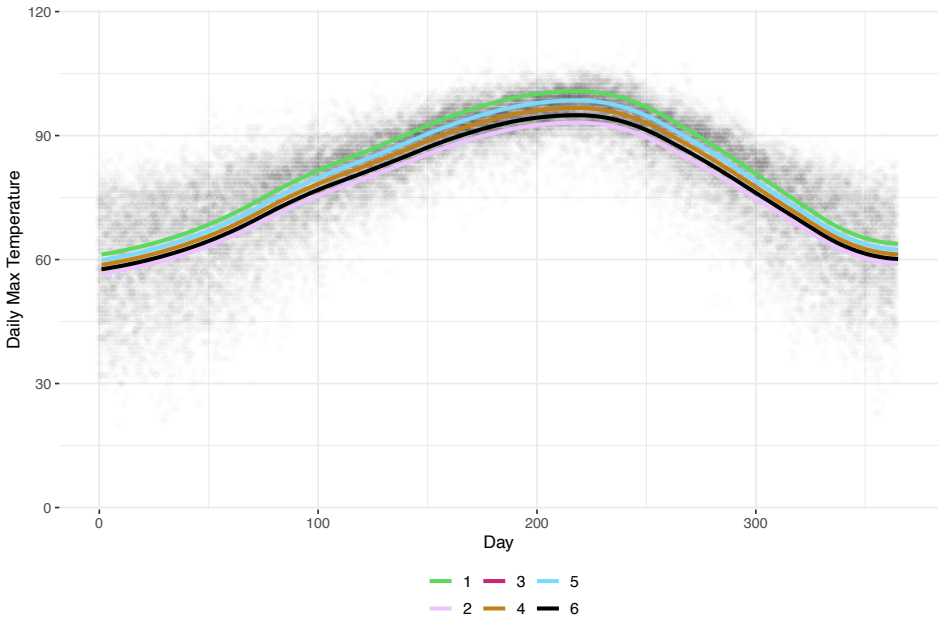
The results of our clustering analysis are in accordance with the globally observed warming trend. Table 2 includes the cluster memberships for each global cluster. Though the member years of each cluster are not all contiguous, the average of the member years in each group provides a good heuristic for comparing the temporal order of each cluster. We see from the local clustering patterns in Figure 6b that temperature increases are not uniform throughout the year. During the spring and the fall, all of the years collapse into a single cluster. Marked differences in temperature patterns include distinctively warmer weather in January, earlier onset of warming weather in spring, and extended hotter summers. Warmer winters are indicated by the co-clustering pattern starting January 1st ($t = 1$, Figure 6b), which shows Group 1 and Group 5 merged with 99% probability. Groups 1 and 5 have the highest average membership year and correspond to the blue line peaking above the rest of the clusters at $t = 1$ in Figure 6b. By the 100^{th} day of the year, Groups 1, 3, and 5 split off from Groups 2, 4, and 6, with higher temperatures (blue line). By the summer solstice (the 172^{nd} day of the year), Group 5 (blue) splits off from Groups 1 and 3 (magenta) to form an extra hot summer cluster. Group 1 leaves Group 3 (magenta) to rejoin Group 5 (blue) around day 225 resulting in extended high temperatures that last through autumn. These results give a more nuanced picture of how temperatures have increased in the past few decades. Spring temperatures rose first, followed by hotter autumns, followed by more intensely hot summers. The

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
1939	1940	1948	1943	1954	1968
1951	1941	1950	1945	1956	1976
1955	1942	1952	1946	2006	1979
1963	1944	1953	1947	2008	1983
1990	1957	1967	1949	2009	1997
1998	1958	1971	1959	2011	
1999	1960	1980	1962	2012	
2000	1961	1984	1964	2017	
2005	1966	1988	1965	2020	
2013	1970	1989	1969		
2016	1973	1994	1972		
2018	1974	1996	1977		
2019	1975	2003	1981		
2021	1978	2010	1982		
		2014	1985		
		2015	1986		
			1987		
			1991		
			1992		
			1993		
			1995		
			2001		
			2002		
			2004		
			2007		
Mean = 1991.93	Mean = 1961.67	Mean = 1982.13	Mean =1974.19	Mean =1999.22	Mean = 1980.60

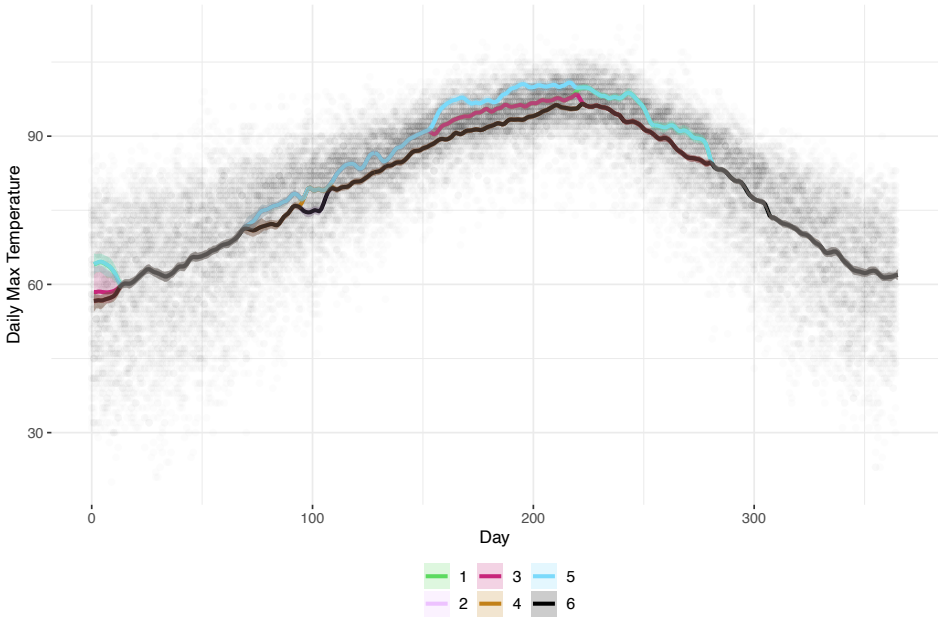
Table 2: Global clusters in the Austin temperature data found by our method. Local clusters subsequently found by our method are shown in Figure 4b.

results of this analysis suggest that energy companies should plan for hotter temperatures to begin earlier in the year, and for summer temperatures to persist for longer. Temperatures in early spring and late autumn/early winter appear to have a wider variance than temperatures in the middle of the year. However, it appears that the dynamic error variance of the HMFLCM was able to absorb the heterogeneity, allowing the fixed effects means to merge into one cluster during portions of the year where temperatures have a wide spread. This could be indicative of the fact that while these portions of the year experience more variable temperature changes within a year, there is not a systematic change between earlier and later years.

In both real-world examples, global clustering produced well-fitted but somewhat obfuscated results while the local clustering results allowed us to draw more precise conclusions. In particular, the Covid-19 local clustering results are able to provide time-specific cluster-changing behavior, lending some credibility to the argument that the hasty reopening of the economy led to the



(a) Global clusters



(b) Local clusters with 95% credible intervals

Figure 6: Global versus local clustering means for the daily maximum temperature of Austin data set. The data are segmented by year to form comparative units. The global clustering results show that there is a general warming trend, but the local clustering results are able to pinpoint what portions of the year are experiencing the largest degree of change.

proliferation of cases. Likewise, the Austin temperature data results demonstrate the ability of our approach to find interesting change points in the data. However, both applications discussed here

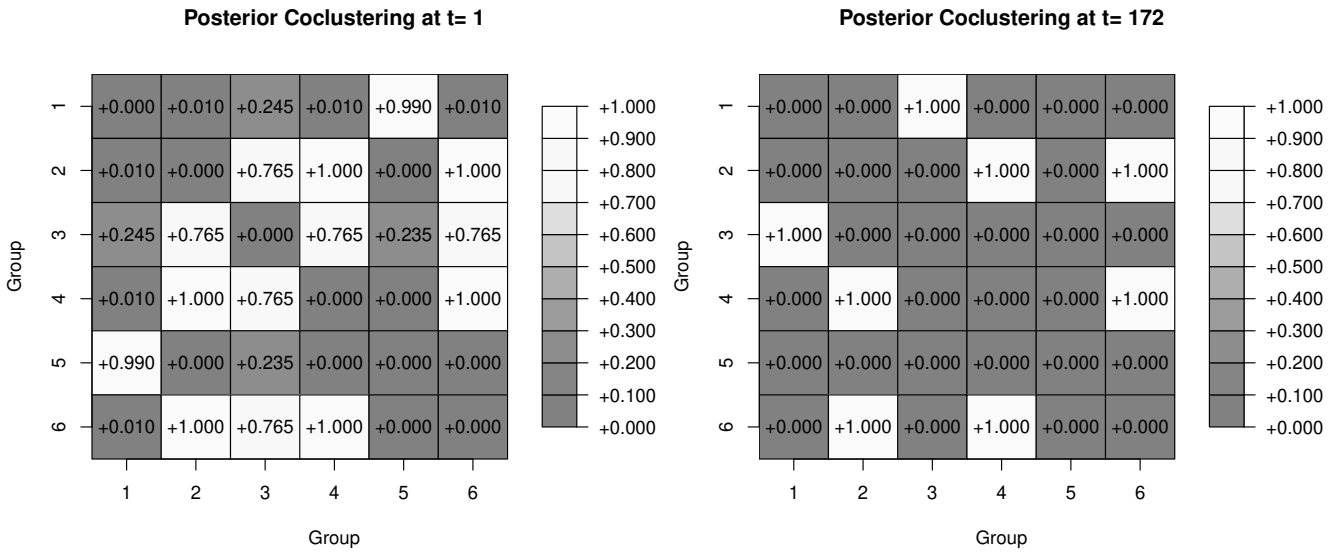


Figure 7: Results for Austin temperature data: Co-clustering patterns show distinct co-clustering profiles in winter (t=1) versus summer (t=172).

involve highly complex dynamics influenced by many unobserved factors and our results must not be over-interpreted as definitive evidence in favor of our drawn conclusions over other plausible alternatives. For instance, it would have been interesting to also be able to incorporate a-priori known cross-constituent dependency patterns in the model (e.g., the geographical proximity of the states in the Covid-19 data set, or the proximity in time of different years in the Austin temperature data set) which could potentially provide even more meaningful clustering patterns and inferences. Particularly for the Covid-19 data set, it would have also been more rigorous to work with a likelihood function that accommodates zero inflation. Such adjustments are however not straightforward, both from a modeling and a computational perspective, and are left as directions for future work.

6 Simulation Studies

In this section, we evaluate the performance of our local clustering method on a simulated data set that mimics the latent mean trajectories seen in the Covid-19 data set. As described in Equation (2), each y_{it} is the summation of one of the four mean functions shown in Figure 1, an individual-specific random effects curve and a Gaussian error with time-varying variance (see Figure S.3 in

the supplementary materials). We simulated five constituent time series for each mean function.

We compare the performance of our local clustering method to three other methods - partitional K-means clustering, divisive hierarchical clustering (DHC), and spectral clustering (Maharaj *et al.*, 2019, Chapter 3). The data are generated with random effects and noise, which can make it difficult for traditional clustering methods to properly align the different series during clustering. To ensure that these comparison methods are not at a disadvantage, we use dynamic time warping to improve the shape-matching ability of partitional clustering, and a shape-based distance metric (Paparrizos and Gravano, 2015) to measure the similarity between the constituent time series to be used in hierarchical clustering. We implemented the dynamic time-warped K-means clustering using the `tsclust` function from the `dtwclust` package in `R`. DHC is implemented using the `hclust` package in `R`. We also included a spectral clustering method in our comparisons, implemented using the `Spectrum` package in `R`.

We evaluated the performance of these methods in a variety of simulation scenarios. In Table 3, we report the average point-wise root-mean-squared error (RMSE) between the true data-generating cluster means and the cluster centroids found by the other methods, and between the true data-generating cluster means and the local cluster means estimated by our method. Additionally, we report the posterior co-clustering error, defined as L_2-norm of the difference between the true co-clustering matrix and the estimated co-clustering matrix. For each time point t, we define the estimated subject-level co-clustering matrix $\mathbf{M}^{N\times N} = ((m_{i,j}))$, where, for $i \leq N, j \leq N, i \neq j$, $m_{i,j}$ is the probability that y_i and y_j belong to the same group. The proposed HMFLCM consistently substantially outperformed the other methods.

The results of these experiments exhibit the HMFLCM's ability to recover the true underlying mean functions and associated clustering patterns with great accuracy. For the global clustering first stage of our approach, we initiated the MCMC algorithm with $D = 6$ allowable latent classes. In simulation experiments, our proposed approach substantially outperforms the competitors which generates extra clusters, leading to many mistakes in co-clustering the component series. See, for example, Figure 8 which shows the estimates obtained by the K-means method, our most robust competitor. By allowing for local clustering while also accounting for unit-specific random effects, the proposed HMFLCM method is able to produce much better results. We see that while the K-means method is able to capture the overall shapes of the time series trajectories, it

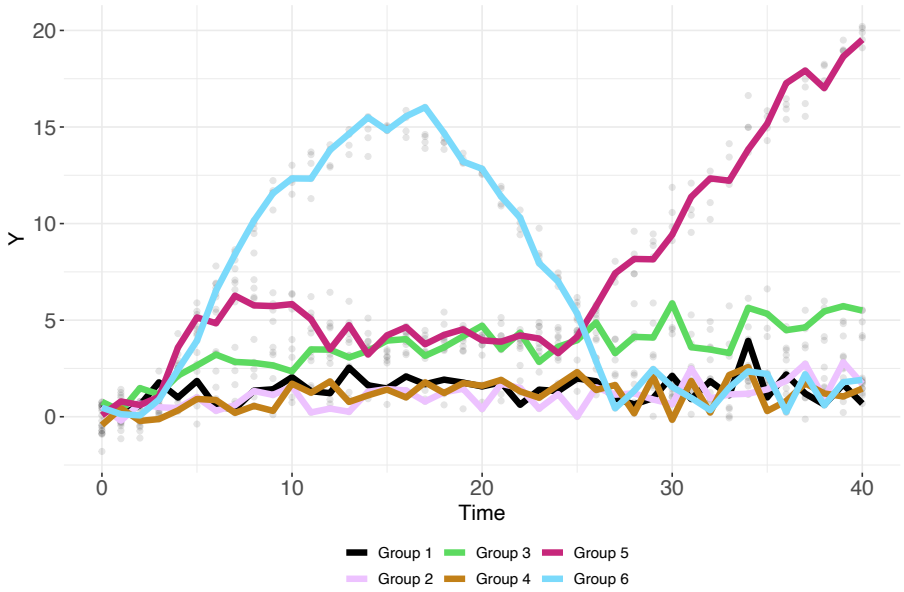


Figure 8: Results obtained by the K-means method for simulated data generated with time-varying error variance. Global means recovered with $D = 6$ clusters. While the cluster centroids picked out by the K-means model match the general shapes of the data-generating functions, the overall profiles are jagged and include extraneous groups in the bottom clusters.

produced a few redundant global clusters, and the estimated cluster mean curves are quite 'pointy' as well. In contrast, Figure 9 shows that the HMFLCM is able to identify and eliminate the extra groups obtained by the global clustering method. The estimated local mean curves are also very smooth. Additionally, the resultant co-clustering pattern between groups is in agreement with the underlying truth with a high degree of accuracy across all time points. Figures S.4, S.5, and S.8 in the supplementary materials illustrate that our method is also able to recover the true underlying individual-specific curves very well.

We also test the local clustering method's sensitivity to hyperparameters by varying D from 3 to 7 ($D_{true} = 4$), and varying the global clustering knot locations to correspond to the quantiles, septiles, and deciles of the time points $K_g = 6$, $K_g = 9$, and $K_g = 12$, respectively. Table 3 shows that our proposed local clustering method produces vastly superior results compared to other global clustering methods. Our local clustering method is also fairly robust to the specification of K_g. The local clustering method naturally achieves the best results when the allowable number of global groups is equal to the true number of groups, $D = D_{true}$, but allowing $D > D_{true}$ achieves better results than allowing $D < D_{true}$. In contrast to the K-means clustering, the performance of the

local clustering method does not deteriorate as D increases. Thus, we recommend choosing a large initial D when using our method.

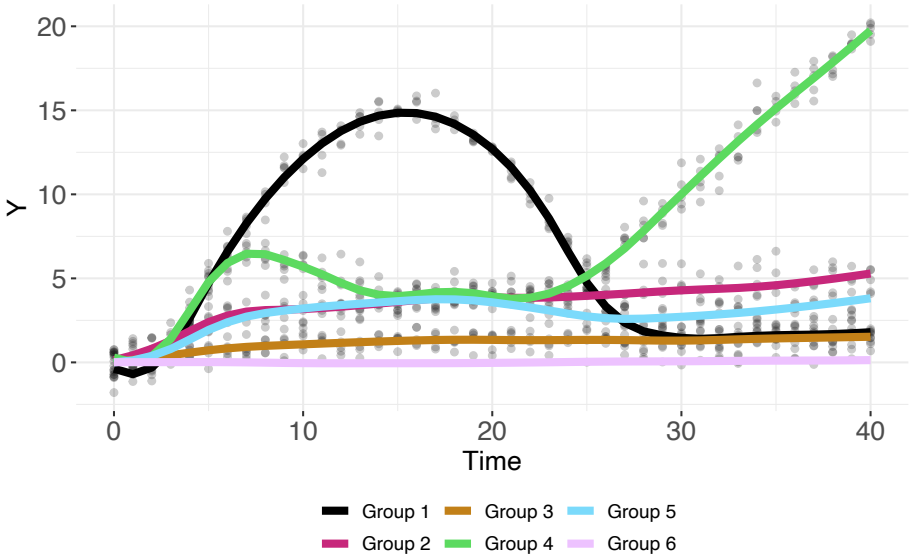
	Mean Point-wise RMSE					
	Local Clustering			K-means	DHC	Spectral
	$K_g = 6$	$K_g = 9$	$K_g = 12$			
$D = 3$	0.21	0.21	0.21	0.67	1.58	0.73
$D = 4$	0.14	0.14	0.14	0.65	3.19	0.30
$D = 5$	0.14	0.14	0.14	0.61	2.97	0.62
$D = 6$	0.14	0.14	0.14	0.63	2.80	0.90
$D = 7$	0.14	0.14	0.14	0.65	2.68	0.97
	Mean Co-clustering Error					
	Local Clustering			K-means	DHC	Spectral
	$K_g = 6$	$K_g = 9$	$K_g = 12$			
$D = 3$	6.19	6.19	6.19	4.41	11.87	6.59
$D = 4$	1.40	1.40	1.40	5.52	8.45	4.27
$D = 5$	1.91	1.91	1.91	5.36	7.56	5.13
$D = 6$	1.91	1.91	1.91	5.60	7.01	5.59
$D = 7$	1.91	1.91	1.91	6.14	6.29	6.21

Table 3: Average point-wise RMSE and co-clustering error for K-means clustering, divisive hierarchical clustering (DHC), spectral clustering, and our proposed local clustering method. For our approach, we varied the number of total global groups D as well as the number of knots K_g used in the global clustering stage. The local clustering naturally performs best when $D = D_{true} = 4$, and better when $D \geq D_{true}$ than when $D < D_{true}$. Allowing D to be over-specified however does not negatively impact the results of the local clustering significantly.

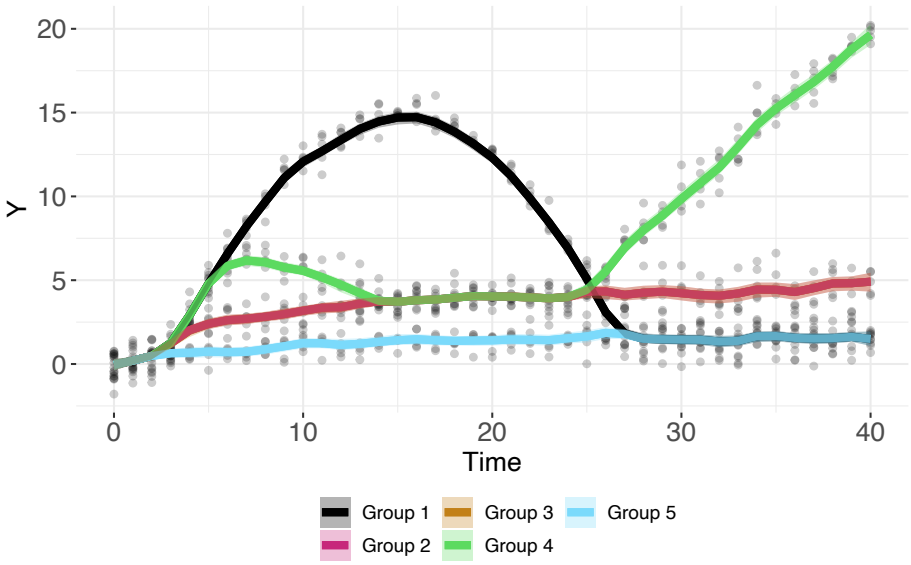
The results of some additional simulation experiments and some additional graphical summaries are presented in Sections S.3 and S.4 in the supplementary materials.

7 Discussion

In this article, we proposed a two-stage method, HMFLCM, for locally clustering functional time series data. Previously existing methods for clustering multiple time series data require that the cluster membership of each component remains the same for the entirety of the data collection period. HMFLCM combines B-spline mixtures with HMMs through the use of an intermediary pseudo-covariate to perform local clustering of multiple time series data, where the cluster membership of each constituent time series, and hence also the total number of clusters, can vary with time. Using the HMFLCM, a constituent time series is able to separate from its original cluster



(a) Global clusters obtained by our method



(b) Local clusters and 95% credible intervals obtained by our method

Figure 9: Results obtained by the proposed HMFLCM for simulated data generated with time-varying error variance. Panel (a): Results obtained by the first-stage global clustering method with $D = 6$ initial clusters. Note that the trajectories recovered for the later time points were particularly inaccurate. Group 6 was empty and is therefore not included in the subsequent local clustering step. Panel (b): Results obtained by the final-stage local clustering method. The issues with global clustering were mostly rectified by co-clustering the wayward units with correct groupings at every time point, which resulted in the merger of Groups 2 and 3, effectively eliminating the additional extraneous group obtained in the global clustering step. In addition, the true underlying merging and splitting behavior of the group means were also accurately recovered. Compared to K-means clustering, the cluster means recovered here are also much smoother.

and form its own unique mean if it begins to behave very differently from its original cluster. Conversely, a time series may merge with a pre-existing cluster if its behavior changes to match the trajectory represented by that cluster. While this ability is in and of itself a novel contribution, our method also offers the additional advantage of increasing the interpretability of the clusters. In all three real-world data examples considered in the main article and the supplementary material, the local clustering results offered evidence for conclusions that would be difficult to draw looking only at global clustering results. Being able to locally merge redundant clusters allows for a quantitative assertion of equivalence between clusters. Similarly, the divergence of a new cluster from a pre-existing cluster indicates that a significant change has occurred. The posterior co-cluster probabilities provide a metric of uncertainty for both possibilities. Overall, we believe our proposed method is a nice addition to the arsenal of statistical methods available for scientific research with multiple time series data. The method, however, makes no assumption on the long-term behavior of the data dynamics such as stationarity, etc. It is thus not suited to make long-term predictions which are nevertheless not very meaningful in settings where local complexities are present.

Avenues of future research include local clustering strategies that preclude the need for a separate first stage for global clustering. In addition, we are considering extensions to allow autocorrelated errors, incorporate a-priori known cross-constituent dependency patterns, etc. We are also actively working on extensions to local clustering models with functional domains in higher dimensional spaces, e.g., for spatial and spatiotemporal data, etc.

Supplementary Materials

The supplementary materials detail the choice of hyper-parameters and the MCMC algorithm used to sample from the posterior. We also include additional figures demonstrating the local clustering method's ability to recover individual-specific curves. The data for our simulation experiment can be accessed as a separate csv file from the online supplementary materials accompanying this paper. `R` codes implementing and demonstrating the methods developed in this article are also included in the online supplementary materials. Manuals for the codes and a ReadMe file providing additional details on how data should be formatted for compatibility with our codes are also included.

Funding

This work was supported in part by grant DMS-1953712 from the National Science Foundation.

Acknowledgement

We thank the Editor, Dr. Robert Gramacy, an anonymous Associate Editor, and three anonymous referees for their thorough review of the originally submitted manuscript and their many constructive comments and suggestions which led to a significantly improved final paper.

References

Aminikhanghahi, S. and Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and Information Systems*, **51**, 339–367.

Berger, M. (2020). States are learning what happens to COVID-19 cases if you reopen too early. Available at https://www.healthline.com/health-news/covid19-cases-rising-states-reopened (2020/07/22).

Center for Disease Control (2020). Clinical questions about covid-19: Questions and answers. Available at https://www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html#:~:text=Based%20on%20existing%20literature%2C,2%E2%80%9314%20days. (2020/07/22).

de Boor, C. (1978). *A Practical Guide to Splines*. Springer-Verlag.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, **11**, 89–102.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**, 515–534.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, **24**, 997–1016.

Gorji Sefidmazgi, M., Moradi Kordmahalleh, M., Homaifar, A., and Liess, S. (2015). Change detection in climate time series based on bounded-variation clustering. In *Machine Learning and Data Mining Approaches to Climate Science*, pages 185–194. Springer.

Lee, J. C., Mervosh, S., Avila, Y., Harvey, B., and Matthews, A. L. (2020). See how all 50 states are reopening (and closing again). Available at https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html (2020/07/22).

Li, Q., Yao, K., and Zhang, X. (2020). A change-point detection and clustering method in the recurrent-event context. *Journal of Statistical Computation and Simulation*, **90**, 1131–1149.

Ma, P., Zhong, W., Feng, Y., and Liu, J. S. (2008). Bayesian functional data clustering for temporal microarray data. *International Journal of Plant Genomics*, **2008**.

Maharaj, E. A., D'Urso, P., and Caiado, J. (2019). *Time Series Clustering and Classification*. Chapman and Hall/CRC Boca Raton.

Nguyen, X. and Gelfand, A. E. (2011). The Dirichlet labeling process for clustering functional data. *Statistica Sinica*, **21**, 1249–1289.

Nguyen, X. and Gelfand, A. E. (2014). Bayesian nonparametric modeling for functional analysis of variance. *Annals of the Institute of Statistical Mathematics*, **66**, 495–526.

Paparrizos, J. and Gravano, L. (2015). k-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pages 1855–1870.

Paulon, G., Mueller, P., and Sarkar, A. (2023). Bayesian hidden Markov tensor partition models for longitudinal data with local variable selection. *Bayesian Analysis*. To appear.

Peng, J., Müller, H.-G., *et al.* (2008). Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *The Annals of Applied Statistics*, **2**, 1056–1077.

Perez, K. X., Cetin, K., Baldea, M., and Edgar, T. F. (2017). Development and analysis of residential change-point models from smart meter data. *Energy and Buildings*, **139**, 351–359.

Petrone, S., Guindani, M., and Gelfand, A. E. (2009). Hybrid Dirichlet mixture models for functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 755–782.

Polson, N. G. and Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, **7**, 887–902.

Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. Q. (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of Applied meteorology and Climatology*, **46**, 900–915.

Rousseau, J. and Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(5), 689–710.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, **11**, 735–757.

Scarpa, B. and Dunson, D. B. (2014). Enriched stick-breaking processes for functional data. *Journal of the American Statistical Association*, **109**, 647–660.

Song, J. J., Lee, H.-J., Morris, J. S., and Kang, S. (2007). Clustering of time-course gene expression data using functional data analysis. *Computational Biology and Chemistry*, **31**, 265–274.

Suarez, A. J. and Ghosal, S. (2016). Bayesian clustering of functional data using local features. *Bayesian Analysis*, **11**, 71–98.

Watanabe, N. (2022). A k-means method for trends of time series: An application to time series of COVID-19 cases in Japan. *Japanese Journal of Statistics and Data Science*, **5**, 303–319.

Watanabe, S. and Opper, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, **11**.

Yu, M. and Ruggieri, E. (2019). Change point analysis of global temperature records. *International Journal of Climatology*, **39**, 3679–3688.