Clustering Random Curves Under Spatial Interdependence
with Application to Service Accessibility

Huijing Jiang
Business Analytics & Mathematical Sciences
IBM Thomas J. Watson Research Center
huijiang@us.ibm.com

Nicoleta Serban
H. Milton Stewart School of Industrial Systems and Engineering
Georgia Institute of Technology
nserban@isye.gatech.edu

Abstract: Service accessibility is defined as the access of a community to the nearby site locations in a service network consisting of multiple geographically distributed service sites. Leveraging new statistical methods, this paper estimates and classifies service accessibility patterns varying over a large geographic area (Georgia) and over a period of 16 years. The focus of this study is on financial services but it generally applies to any other service operation. To this end, we introduce a model-based method for clustering random time-varying functions which are spatially interdependent. The underlying clustering model is nonparametric with spatially correlated errors. We also assume that the clustering membership is a realization from a Markov random field. Under these model assumptions, we borrow information across functions corresponding to nearby spatial locations resulting in enhanced estimation accuracy of the cluster effects and of the cluster membership as shown in a simulation study.

Keywords: functional data analysis, Markov random field, model-based clustering, service accessibility, spatial dependence, semi-parametric modeling.

1We would like to thank Alexander Gray and Nikolaos Vasiloglou II for their helpful suggestions and for their support in developing the C++ code for this paper using Fastlib library. The authors are thankful to the two referees who provided very useful feedback, and to the associate editor whose input greatly helped improving the presentation of this paper. Huijing Jiang has been supported by Temenbaum Institute and Statistical and Applied Mathematical Sciences Institute fellowships while performing this research and she is now a postdoctoral fellow at IBM TJ Watson Research Center. Nicoleta Serban has been supported by the National Science Foundation under Grant CMMI-0954283.
1 Introduction

Research in service accessibility has emerged as economic and social equity advocates recognized that where people live influences their opportunities for economic development, access to quality healthcare, and political participation (Blackwell and Treuhaft, 2008; Frumkin et al., 2004; Lee and Rubin, 2007). In this paper, we leverage new statistical methods for estimating and describing service accessibility trends that can be used to inform about potential business opportunities as well as about the extent of service accessibility disparities.

Many existing studies have analyzed service accessibility but they are limited to small geographic areas such as towns and only one year of data (Graves, 2003; Larson, 2003; Moore et al., 2006; Small and McDermott, 2006; Talen, 2001; Talen and Anselin, 1998). Various measure of service accessibility have been introduced in the existing literature in urban planning, economic geography and healthcare. Generally, the accessibility measures group into two categories: congestion (Gualiardo, 2004; Luo and Qi, 2009; Yang et al., 2006) and travel cost measures (Hyndman, Holman, 2001; Lovett, 2002; Talen and Anselin, 1998; Talen, 2001). In this paper, accessibility is measured as the utilization-scaled travel cost of a community to the nearby sites in a service network consisting of multiple geographically distributed service sites.

One primary challenge in analyzing service accessibility over a large geographic area and a multiple-year period is estimation and characterization of a large number of time-varying accessibility curves/functions. For example, in this paper we measure service accessibility in Georgia at the census tract level over 16 years; this results in 1624 time-varying random functions. To prevail over this challenge, we propose using a clustering method to reduce the information content of geographically and temporally varying data to meaningful summary trends that reveal the prevalent changes in service accessibility within a given geographic space (e.g. state). This study focuses on the distribution of financial services but the service accessibility framework applies generally to other services both public (e.g. education, parks) and private (e.g. food stores).

Due to an increasing number of applications with a very large number of random (time-varying) functions, data reduction methods such as clustering play an important role in Functional Data Analysis (FDA). The literature for functional data clustering is divided into
Figure 1: Sampled curves of accessibility measurements from each clusters. Grey lines are 200 observed curves randomly sampled from each cluster and colored curve are the estimated global/cluster patterns.

hard clustering methods (Hastie et al., 2000; Bar-Joseph et al., 2002; and Serban, 2008, 2009) and model-based clustering (James and Sugar, 2003; Fraley and Raftery, 2002; and Wakefield et al., 2002). Although there are many competitive approaches to clustering functional data, they are generally limited to the assumption of independence between curves. However, there are many case studies including our motivating application where this assumption does not hold - the service accessibility functions are spatially interdependent since each function corresponds to a census tract or neighborhood within a geographic space. Figure 1 presents a sample of accessibility random functions grouped according to their cluster membership. Each random function corresponds to the time-varying travel cost to financial service sites for one community. Our objective is to cluster the random functions by shape regardless of
scale. In our case study, we analyze the data on the log-scale, and therefore, re-scale each function by simply subtracting the function-specific mean.

The underlying statistical problem is to cluster multiple functions observed with error:

\[ Y_{ij} = f_{s_j}(t_{ij}) + \sigma \varepsilon_{ij}, \quad j = 1, \ldots, n, \]  

where \( s_j \) for \( j = 1, \ldots, n \) are coordinates of spatial units in a \( d \)-dimensional spatial domain (in our empirical studies \( d = 2 \)), \( f_{s_j}(t) \) is the random time-varying function indexed by the spatial unit \( s_j \) observed at the time points \( (t_{1,j}, \ldots, t_{m,j}) \) and \( \varepsilon_{ij}'s \) are spatially correlated random errors. The cluster information will enter in the model via \( f_{s_j}(t) \), that is different clusters will have different \( f \)'s. For simplicity of the presentation, we assume that the observation time points \( t_{ij}, i = 1, \ldots, m \) are the same across all random functions. For Georgia, the number of spatial locations is \( n = 1624 \) and the number of time points available as of 2010 is \( m = 16 \).

In this paper, the clustering approach is model-based. That is, the complete data are \((Y_j, Z_j), \ j = 1, \ldots, n\) where \( Y_j = (Y_{1j}, \ldots, Y_{mj}) \) are realizations from the \( j \)th random function. Given the number of clusters is \( C \), \( Z_j = (Z_{j1}, \ldots, Z_{jC})'s \) specify the cluster membership of the \( j \)th function \((Z_{jk} = 1, \text{ when that spatial location } s_j \text{ is in cluster } k\)). In unsupervised clustering, \( Z_j, \ j = 1, \ldots, n \) are latent variables.

Recent research in clustering functional data has considered spatial interdependence in the cluster membership \( Z_j, j = 1, \ldots, n \) (Blekas et al., 2007) or within-cluster dependence, i.e., spatial dependence among the locations within the same cluster (Booth et al., 2008; Shi and Wang, 2008). In contrast, the clustering method introduced in this paper models the spatial dependence in the joint distribution of \((Y_j, Z_j), \ j = 1, \ldots, n\) by assuming spatial dependence on the cluster membership \( Z_j, j = 1, \ldots, n \) and on the conditional distribution of \( Y_j|Z_j, j = 1, \ldots, n \). Spatial dependence in the joint distribution \((Y_j, Z_j)\) allows borrowing information across curves corresponding to nearby locations yet maintaining local resolution which will lead to enhanced estimation accuracy of the cluster patterns and of the cluster membership.

Importantly, clustering is commonly used as a tool for summarizing a large number of functional profiles, and therefore, computational efficiency is crucial in clustering spatially
interdependent functional data since the spatial covariance matrix tends to be very large. Consequently, we improve upon the computational efficiency of the estimation algorithm by employing a low-rank approximation to reduce the size of the spatial covariance matrix and by using Monte Carlo approximations in the imputation of the latent variables. We refer to our modeling procedure as the *Functional-Spatial Clustering Model (FSCM)*.

The article is organized as follows. Section 2 introduces the clustering model decomposition. Section 3 provides the estimation approach for the clustering trends and the cluster memberships along with the description of the selection of the smoothing parameters and the number of clusters. We illustrate our clustering method with simulated data in Section 4 investigating the accuracy of the cluster trends and cluster membership under a range of model scenarios with varying noise levels and spatial correlation structures. In Section 5 we apply FSCM to summarize the time-varying trends in financial service accessibility for Georgia. Section 6 concludes the paper. Some technical details are deferred to the Supplemental Material. We complement the accessibility study with a series of additional visual displays included in the Supplemental Material.

## 2 Functional Spatial Clustering Model

In model-based clustering, the underlying assumption is that the complete data are \((Y_j, Z_j)\) for \(j = 1, \ldots, n\) where \(Y_j = \{Y_{ij}\}_{i=1,\ldots,m}\) are realizations from a multivariate distribution with mean vector \(\mu_j\) and covariance matrix \(\Sigma\) whereas the cluster membership \(Z_j = (Z_{j1}, \ldots, Z_{JC})\) follows a multinomial distribution with the proportion parameters \(\pi_{jk} = P(Z_{jk} = 1), k = 1, \ldots, C\) (Fraley and Raftery, 2002). Therefore, we need to specify the distribution of the latent variables \(Z_j\) and the conditional distribution \(Y_j|Z_j\) for \(j = 1, \ldots, n\), which in turn, specifies the distribution of the complete data. *In the existing approaches to model-based clustering, \(Y_j|Z_j, j = 1, \ldots, n\) are assumed conditionally independent and the clustering membership \(Z_j, j = 1, \ldots, n\) are also assumed independent. In this paper, we relax these assumptions to incorporate spatial dependence.* In Section 2.1, we describe a locally-dependent Markov model for the latent variable. In Section 2.2, we introduce a functional model for the conditional distribution \(Y_j|Z_j\).
2.1 Markov Model for the Cluster Membership

We assume that the clustering configuration $Z = (Z_1, \ldots, Z_n)$ is a realization from a locally-dependent Markov random field (MRF) which is a stochastic process with the Markov property. Under the Markov assumption, the probability that the $s_j$th spatial unit belongs to the $k$th cluster denoted $\pi_{jk}$ depends on the states of its nearest neighbors where a state is defined by the cluster membership. Following the current literature on MRF modelling, we model the probability mass function $P(Z_{jk} = 1|Z_{\partial j})$ for $j = 1, \ldots, n$ and $k = 1, \ldots, C$ ($C$ is the number of clusters) using the Gibbs distribution where $\partial j$ is a prescribed neighborhood of the $s_j$th spatial unit. In this paper, we use the K-nearest neighbors method to define the neighborhood structure. One limitation of this approach is that the set of neighbors is arbitrarily chosen to consist of the five nearest neighbors; however, the extent of the neighbors may be specified following insights from the application under study.

This distribution originates from statistical physics where it is used to model the states of atoms and molecules, and later on, was adopted by statisticians to model Markov Random Fields. The probability mass function for Gibbs distribution is defined as

$$P(Z_{jk} = 1|Z_{\partial j}) = \pi_{jk} = \frac{1}{N_j(\theta)} \exp(U_{jk}(\theta)) \tag{2}$$

where $U_{jk}(\theta) = \theta \sum_{i \in \partial j} Z_{ik}$ is called the energy function. Large values of $U_{jk}(\theta)$ correspond to spatial patterns with large spatially connected sub-areas belonging to the same cluster. Small values of $U_{jk}(\theta)$ correspond to spatial patterns that do not display any sort of spatial organization. $N_j(\theta)$ is a normalizing constant called the partition function. The probability mass function depends on $\theta$ called the interaction parameter. The larger $\theta$ is, the more extensive the spatial dependence of the cluster membership is. The value $\theta = 0$ corresponds to a multinomial distribution with equal class probabilities and no spatial dependence.

The assumptions on the cluster membership in the approach introduced in this paper are similar to Blekas et al. (2007) although the modeling approach is different. In our model, we assume a Gibbs distribution prior on $Z_{jk}$ whereas Blekas et al. (2007) imposes the spatial constrains on $\pi_{jk}$, i.e., $p(\pi_{jk}) = \frac{1}{N_j(\theta)} \exp(V_{jk}(\theta))$, where $V_{jk}(\theta)$ is a clique potential function defined in a similar way as $U_{jk}(\theta)$ above. Therefore, the Blekas et al. (2007) approach models
the spatial dependence in the cluster membership through $\pi_{jk}$ instead of $Z_{jk}$ as in our model.

In addition to allowing for dependence in the latent variable specifying the cluster membership, we also allow for spatial dependence in the conditional distribution of $Y_j | Z_j$ described in the next section. To the best of our knowledge, in all relevant work, $Y_j | Z_j$ are assumed conditionally independent for computational feasibility although this is one of the most contested assumptions (see Besag, 1986 and the following discussions; Archer and Titterington 2002; and the references therein).

2.2 Functional Model for Conditional Distribution

Given the cluster membership $Z_j = (Z_{j1}, \ldots, Z_{jC}), j = 1, \ldots, n$ where $Z_{jk} = 1$ if $Y_{ij}$ belongs to $k$th cluster and 0 otherwise, we assume that $Y_j | Z_j$ follows a multivariate distribution with a functional representation

$$Y_j(t_i) | (Z_{jk} = 1) = \mu(t_i) + \mu_k(t_i) + \epsilon_{ij},$$

(3)

where $\mu(t)$ is the global trend, $\mu_k(t)$ is the cluster effect and $\epsilon_{ij}$ is the random error which is commonly assumed iid distributed with $N(0, \sigma^2_\epsilon)$. Similar to functional ANOVA, we constrain the sum of time-dependent cluster effects to be equal to zero - $\sum_{k=1}^C \mu_k(t) = 0$. Following this functional representation, we decompose the temporal trend and the cluster patterns according to

$$\mu(t) = \sum_{l=1}^{L_0} \alpha_l \phi_l(t) \text{ and } \mu_k(t) = \sum_{l=1}^{L_k} \beta_{k,l} \phi_l(t)$$

(4)

where the $\phi_l(t), l = 1, \ldots, L_k$ are temporal basis functions in $L_2(T)$. Common choices of temporal basis functions are B-splines or P-splines (Ruppert, Wand and Carroll 2003). In our implementation, we use B-spline basis. We discuss selection of the number of basis $L_k, k = 0, 1, \ldots, C$ in Section 3.3.

One extension of our functional model (3) is to allow $Y_j | Z_j$ be spatially dependent by assuming spatially correlated random errors. The spatial correlation in the random errors
\( \epsilon_{ij} \)'s can be modeled using the covariance function

\[
C(s_j - s_{j'}, v, \sigma^2_s, \sigma^2_\epsilon) = \begin{cases} 
\sigma^2_s \rho(s_j - s_{j'}; v), & \|s_j - s_{j'}\| > 0; \\
\sigma^2_s + \sigma^2_\epsilon, & \|s_j - s_{j'}\| = 0.
\end{cases}
\] (5)

where \( \rho(s_j - s_{j'}; v) \) is the spatial correlation function such as Exponential, Gaussian or Matern and \( \sigma^2_\epsilon \) accounts for the nugget effect (Cressie, 1993).

An equivalent formulation is to decompose the spatial correlation structure specified by \( \epsilon_{ij} \) into a spatial random effect \( \tau(s) \) and independent random errors \( \epsilon_{ij} \sim N(0, \sigma^2_\epsilon) \). The functional model (3) with spatially correlated errors then becomes

\[
Y_{ij} | (Z_{jk} = 1, \tau) = \mu(t_i) + \mu_k(t_i) + \tau(s_j) + \epsilon_{ij} \] (6)

where \( \tau(s) = \sum_r^n \gamma_r \psi_r(s) \) is the spatial random effect with \( \gamma_r \sim N(0, \sigma^2_\gamma) \) and \( \{\psi_r(s), r = 1, \ldots\} \) is the empirical smoothing basis such that \( \Psi \Psi' = \Omega \) where \( \{\Omega\}_{j,j'} = \rho(s_j - s_{j'}; v) \). In Section 3, we discuss a computationally efficient method to obtain the empirical smoothing basis \( \Psi \) using a reduced rank Cholesky decomposition.

### 3 Model Estimation and Selection

Following the model formulation in (6), we estimate the model parameters, and impute the missing clustering membership and spatial random effects by maximizing the likelihood

\[
L_O(\alpha, \beta_1, \ldots, \beta_C, \sigma^2_s, \sigma^2_\epsilon, \theta) = C \sum_{k_1=1}^{C} \pi_{k_1} \cdots \sum_{k_n=1}^{C} \pi_{nk_n} f(Y_1, \ldots, Y_n; \alpha, \beta_k, \sigma^2_s, \sigma^2_\epsilon) \] (7)

where \( \alpha = (\alpha_1, \ldots, \alpha_{L_0}) \) and \( \beta_k = (\beta_{k1}, \ldots, \beta_{kL_k}) \) are the coefficient vectors of the global trend and of the \( k \)th cluster pattern in (4), respectively. Since both the cluster membership \( Z \) and the random effects \( \gamma_r, r = 1, \ldots, n_S \) are missing, we use a modified version of the EM estimation algorithm. EM converges to the maximum of (7) by iteratively computing (E-step) and maximizing (M-step) the expectation of the likelihood of the complete data.
defined as

$$E[L_C(\alpha, \beta_1, \ldots, \beta_C, \sigma^2_c, \sigma^2_s, \theta)|Y] = E[f(Y|\gamma, Z; \alpha, \beta_1, \ldots, \beta_C, \sigma^2_c)f(\gamma; \sigma^2_s)f(Z; \theta)]. \quad (8)$$

However, there are several computational challenges in the application of the EM algorithm in the context our estimation approach as discussed next.

### 3.1 Computational Challenge I: Spatial Dependence

In the E-step, in order to evaluate $E[L_C(\alpha, \beta_1, \ldots, \beta_C, \sigma^2_c, \sigma^2_s, \theta)|Y]$, we impute the missing cluster membership by its conditional expectation $E[Z|Y]$ which can be calculated through the prior distribution $P(Z_{jk} = 1)$ and the likelihood function $f(Y_j|Z_{jk} = 1)$ as follows

$$E[Z_{jk}|Y] = P(Z_{jk} = 1|Y) \propto f(Y_j|Z_{jk} = 1)P(Z_{jk} = 1); \quad j = 1, \ldots, n; \quad k = 1, \ldots, C.$$  

However, the calculation of the prior distribution $P(Z_{jk} = 1)$ is computationally challenging due to the spatial dependence in $Z_1, \ldots, Z_n$. The computation of the likelihood $f(Y_j|Z_{jk} = 1)$ is also not trivial because of the conditionally spatial dependence in $Y_j|Z_j$. We overcome these challenges by applying two approximations: a pseudo-likelihood imputation for $Z_1, \ldots, Z_n$; and Monte-Carlo approximations in the imputation of the latent variables as described below.

Due to the spatial dependence in $Z_1, \ldots, Z_n$, we only have the joint distribution $P(Z_1, \ldots, Z_n)$. Therefore, it is not straightforward to derive the marginal distribution $P(Z_{jk} = 1)$. In the HMRF (Hidden Markov Random Fields) literature (Besag, 1986, Archer and Titterington, 2002), this difficulty is overcome by assuming local dependence for each spatial unit $s_j$, i.e., the likelihood of $Z_j$ will only depend on its neighbors $Z_{\partial j}$. Thus the joint distribution of $Z_1, \ldots, Z_n$ can be approximated by a pseudo-likelihood function,

$$f(Z_1, \ldots, Z_n; \theta) \approx \prod_{j=1}^{n} f(Z_j|Z_{\partial j}; \theta). \quad (9)$$

where $f(Z_j|Z_{\partial j}; \theta)$ is the Gibbs distribution (2).

Additionally, it is not straightforward to derive $f(Y_j|Z_{jk} = 1)$ (or $f(Y_j|Z_j)$) because
$Y_j|Z_j$'s are conditionally (spatial) dependent. However, $Y_j|(Z_j, \gamma)$ are conditionally independent, and thus we instead impute $Z_{jk}$ as follows,

$$\hat{z}_{jk} = E[Z_{jk}|Y_1, \ldots, Y_n] = E[E[Z_{jk}|Y_1, \ldots, Y_n, \gamma]] \approx \frac{1}{M} \sum_{m=1}^{M} E[Z_{jk}|Y_1, \ldots, Y_n, \gamma^{(m)}]$$

where $\gamma^{(1)}, \ldots, \gamma^{(M)}$ are samples drawn from the conditional distribution of $\gamma|Y$.

Furthermore, when the spatial region is densely sampled - $n_S$ (dimension of $\gamma$) is large-Monte Carlo sampling needs to be performed over a high dimensional space. When the distribution of $\gamma|Y$ has thin tails, we can replace the probability distribution over $\gamma|Y$ with a point estimate $\hat{\gamma} = E[\gamma|Y]$ representing the distribution’s mode because Monte Carlo samples $\hat{\gamma}^{(m)}$, $m = 1, \ldots, M$ will have values very close to the mode, $\hat{\gamma}$. Then

$$\hat{z}_{jk} \approx \frac{1}{M} \sum_{m=1}^{M} E[Z_{jk}|Y_1, \ldots, Y_n, \gamma^{(m)}] \approx \frac{1}{M} \sum_{m=1}^{M} E[Z_{jk}|Y_1, \ldots, Y_n, \hat{\gamma}] = E[z_{jk}|Y_1, \ldots, Y_n, \hat{\gamma}]$$

Subsequently, a further approximation to $\hat{z}_{jk}$ is

$$\hat{z}_{jk} = E[z_{jk}|Y_1, \ldots, Y_n, \hat{\gamma}]$$

Detailed EM algorithm and the integration of these approximations into the estimation approach are deferred to the Supplemental Material 1.

### 3.2 Computational Challenge II: Dimension Reduction

In our approach, the conditional spatial dependence $Cov[Y_{ij}, Y_{ij'}|Z] = \sigma^2_s \rho(s_j - s_{j'}; v) + \sigma^2_c$ is modeled using a random spatial effect decomposed using a spatial smoothing basis of functions $\{\psi_r(s), r = 1, 2, \ldots\}$ and random effects $\gamma_r$, $r = 1, \ldots, n$ with mean zero and variance $\sigma^2_s$. The spatial smoothing basis is constructed by Cholesky decomposition of the spatial correlation matrix $\Omega = \Psi \Psi$ where $\Omega = \{\rho(s_j - s_{j'}; v)\}_{j,j'=1,\ldots,n}$. However, a full Cholesky decomposition results in a $n$-by-$n$ matrix $\Psi = \{\psi_r(s_j)\}_{j,r=1,\ldots,n}$. When the spatial domain is densely observed, operation on $\Psi$ is computationally expensive. In order to reduce the computational cost, we replace it with a matrix $\tilde{\Psi}$ of size $n \times n_S$ with $n_S \ll n$, such
that the norm of the difference $\Omega - \hat{\Psi}'\hat{\Psi}$ is less than a prespecified threshold $\eta$. This can be achieved via incomplete Cholesky decomposition (Bach and Jordan, 2002). In the motivating case study, we set $\eta = 0.001$ and the rank is reduced from $n = 1624$ to $n_S = 766$. Note that the low-rank approximations obtained from incomplete Cholesky decomposition is mainly to improve the computational efficiency, not for regularization.

### 3.3 Model Selection

The clustering model described in the previous section depends on a series of tuning parameters which are assumed fixed: the number of clusters $C$ and the number of basis functions $L_k, k = 0, 1, \ldots, C$. Since our clustering algorithm is model-based, the problem of identifying the number of clusters and the number of basis functions is equivalent to a model selection problem where each combination $(C; L_k, k = 0, 1, \ldots, C)$ corresponds to a different model.

Common variable selection methods such as the Akaike information criterion (AIC), and Bayesian information criterion (BIC) have been employed for estimating the number of clusters (Fraley and Raftery, 2002). Both criteria select the number of clusters and the number of basis functions by minimizing the information score function

$$-2 \log L(\alpha, \beta_1, \ldots, \beta_C, \sigma_s^2, \sigma_e^2, \theta) + 2J(C, L_0, L_1, \ldots, L_C)$$

where $\log L(\cdot)$ is the log likelihood which measures the lack of fit.

When the model under consideration contains random effects, it is not straightforward what likelihood function to use for the selection criteria. Vaida and Blanchard (2005) discussed this issue by defining two variations of AIC - marginal AIC (mAIC) and conditional AIC (cAIC) for mixed-effects model selection. Following their arguments, if only the fixed effects contain information about the number of clusters, the AIC with the marginal likelihood function should be used. Whereas if the model selection involves both the fixed effects and random effects, AIC with the conditional likelihood should be used.

Our model formulation is different from Vaida and Blanchard (2005) in that there are two multivariate random variables to condition on $Y$, the latent variable $Z$ and the random effects $\gamma$ specifying the random spatial effects. We therefore use the joint likelihood in (8)
and not the marginal or the conditional likelihood to select the number of clusters. The second term $2J(C, L_0, L_1, \ldots, L_C)$ is the penalty term that measures the complexity of the model. For AIC, the penalty is $2J = 2d$ and for BIC, it is $2J = (\log I)d$ where $d$ is the number of parameters. The final estimate for $(\hat{C}, \hat{L}_k, k = 1, \ldots, C)$ is such that it minimizes the BIC or AIC score.

We note that the minimization is over a large number of parameters. For simplicity we assume $L_k = L$ for all $k = 0, 1, \ldots, C$ although the assumption of similar smoothness level across cluster trends is restrictive but computationally efficient.

Remark: One has to bear in mind, that all model selection approaches have their limitations (e.g. AIC favors larger models) and that model selection may be accompanied by model assessment. In clustering analysis, interpretability of the temporal cluster trajectories could be used to assess the number of clusters starting with a model selected using BIC or AIC as described above.

## 4 Simulation

In this simulation study, our primary objective is to assess the prediction accuracy of the cluster membership and of the cluster trends under a series of spatial interdependence structures and varying noise levels. We compare our method (FSCM) with three other model-based clustering methods:

1. Functional Clustering (FClust) assuming independence in both the cluster membership and the random errors.

2. Functional Clustering under Markov Constrains: (FCMC) assuming Markov dependence in the cluster membership $Z$ as in equation (2) and independence in the conditional distribution $Y|Z$;

3. Functional Clustering under Conditional Dependence (FCCD) assuming independence in the cluster membership but spatial dependence in the conditional distribution $Y|Z$ as in equation (6);

Methods 1-3 are special cases of FCSM. In the simulation study, a comprehensive R implementation including FCSM and the three special cases was used.
4.1 Simulation Set-up

We generated synthetic data from the functional model

\[ Y_j(t_i) = \mu(t_i) + \mu_z(t_i) + \epsilon_{ij}, \quad \text{with} \quad t_i = (i - 1)/(m - 1), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n \quad (10) \]

where \( \epsilon_{ij} \)'s are spatially correlated according to the covariance structure in (5). In this study, we consider a small number of time points \( m = 10 \) and three clusters with the following patterns

\[ \mu_1(t) = \exp(t) \cos(t), \quad \mu_2(t) = \cos\left(\frac{5\pi}{2}t\right), \quad \mu_3(t) = -(\mu_1(t) + \mu_2(t)). \]

For simplicity of the presentation, in this simulation study, we set \( \mu(t) = 0 \). For the spatial covariance function (5), we use Matérn covariance matrix of order \( \frac{1}{2} \) and range parameter 1, the default setting in the function \( \text{stationary.cov} \) from the ’fields’ package in R statistical software. Figure 2 presents the true simulated clusters along with the estimated clustering patterns for one simulation realization.

**Markov Model for the Cluster Membership.** The spatial units \( s_1, \ldots, s_n \) with \( n = 812 \) are randomly sampled from the centroids of the census tracts in Georgia. The cluster membership \( (Z_1, \ldots, Z_n) \) is generated using the Gibbs sampler introduced by Geman and Geman (1984). Following HMRF methodology, the cluster membership of a site \( s_j \) is sampled from a multinomial distribution with proportion parameters \( (\pi_{j1}, \ldots, \pi_{jC}) \) of the Gibbs distribution defined in Equation 2.

**Simulation settings.** We investigate the estimation accuracy of the cluster membership by varying three factors in a \( 2^3 \) full-factorial design.

1. Spatial Dependence of \( Z \) controlled by the hyperparameter \( \theta \) of the Gibbs distribution in equation (2). Levels \( \theta = 0.2 \) and \( \theta = 0.5 \) correspond to weak and strong spatial dependence in the cluster membership, respectively.

2. Conditional Spatial Dependence controlled by \( \sigma_s^2 \). Levels \( \sigma_s^2 = 1 \) and \( \sigma_s^2 = 5 \) correspond to weak and strong conditional dependence in \( Y|Z \), respectively.

3. Noise level controlled by \( \sigma_r^2 \). Levels \( \sigma_r^2 = 1 \) and \( \sigma_r^2 = 5 \) correspond to low and high noise levels, respectively.
4.2 Results

Because we have the true clustering membership in our simulation example, we can assess the accuracy of the clustering membership prediction for the method introduced in this paper and the alternative methods using a clustering accuracy error. In the simulation study, the clustering methods assume that the number of clusters is fixed to $C = 3$ and not estimated from the data. Since our clustering method is model-based, we convert cluster probabilities (soft clustering) to hard clustering using the most likely membership class based on the predicted membership probabilities. We measure the clustering error using the Rand index (Rand, 1971), which is the fraction of all misclustered pairs of curves. This index is invariant to label switching either among true or estimated labels. Let $C = \{f_1, \ldots, f_S\}$ denote the
set of true curves, \( \hat{C} = \{ \hat{f}_1, \ldots, \hat{f}_S \} \) denote the set of estimated curves, and \( T \) and \( \hat{T} \) denote the true and estimated clustering maps, respectively. The Rand index is defined by

\[
R(C, \hat{C}) = \sum_{r<s} I(T_k(f_r, f_s) \neq \hat{T}_k(f_r, f_s)) / \binom{N}{2}.
\]

In Table 1 we present the clustering accuracy instead of the clustering error, which is measured as \( 1 - R(C, \hat{C}) \). Therefore, the clustering accuracy measure is close to 1 when there are only few misclustered curves. Although our primary focus is on enhanced estimation of the clustering membership, we also investigate the accuracy of the clustering patterns (Table 2) computed by standardized root mean squared error (RMSE)

\[
RMSE = \sqrt{\frac{||\mu_k(t) - \hat{\mu}_k(t)||^2}{||\mu_k(t)||^2}}
\]

along with the accuracy of a series of parameters in our model including \( \theta \), \( \sigma_s \), and \( \sigma_\epsilon \).

The study is replicated for 100 runs. In Tables 1 and 2, we report the mean performance results with standard errors. We applied AIC for selecting the number of clusters \( C \) ranging from 2 to 10 clusters on all eight simulation settings. Table 3 shows the percentage of replications for which the number of clusters is correctly identified over the 100 runs. Tables 4 to 6 present the accuracy of the primary parameters in our clustering model, the hyper-parameter \( \theta \) of the Gibbs distribution specifying the distribution of the cluster membership \( Z \), the parameters \( \sigma_s^2 \) controlling the level of conditional spatial dependence in \( Y|Z \) and \( \sigma_\epsilon^2 \) controlling the noise level.

We summarize our findings as follows:

- (Table 1) The accuracy of the cluster membership estimates is highest for the model introduced in this paper but we note that the high accuracy is due to accounting for conditional spatial dependence rather that the spatial dependence in the cluster membership - the accuracy of the cluster membership for FSCM vs. FCCD is \( 0.80 \pm 0.02 \), and respectively, \( 0.79 \pm 0.03 \) whereas for FCMC and FClust is \( 0.69 \pm 0.03 \), and respectively, \( 0.66 \pm 0.044 \) under the setting \( \theta = 0.5 \), \( \sigma_s^2 = 1 \) and \( \sigma_\epsilon^2 = 5 \), for example. The accuracy of the estimates for FSCM over FCCD improves under strong conditional spatial dependence and high noise level. Moreover, the clustering accuracy is low under low signal-to-noise ratio (\( \sigma_\epsilon = 5 \)) but
Table 1: Average percent clustering accuracy $100(1-\mathcal{R}(\mathcal{C}, \hat{\mathcal{C}}))$. Standard errors are given in the parentheses.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$\sigma^2_z$</th>
<th>$\sigma^2_{\varepsilon}$</th>
<th>FSCM</th>
<th>FCCD</th>
<th>FCMC</th>
<th>FClust</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1</td>
<td>1</td>
<td>0.96(0.01)</td>
<td>0.96(0.03)</td>
<td>0.69(0.03)</td>
<td>0.72(0.05)</td>
</tr>
<tr>
<td>0.2</td>
<td>1</td>
<td>5</td>
<td>0.80(0.03)</td>
<td>0.79(0.03)</td>
<td>0.59(0.02)</td>
<td>0.66(0.04)</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>1</td>
<td>0.67(0.10)</td>
<td>0.66(0.09)</td>
<td>0.59(0.02)</td>
<td>0.59(0.02)</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>5</td>
<td>0.76(0.04)</td>
<td>0.67(0.05)</td>
<td>0.59(0.01)</td>
<td>0.57(0.01)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.97(0.01)</td>
<td>0.96(0.01)</td>
<td>0.72(0.06)</td>
<td>0.72(0.06)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>5</td>
<td>0.80(0.02)</td>
<td>0.79(0.03)</td>
<td>0.69(0.03)</td>
<td>0.66(0.04)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>1</td>
<td>0.67(0.09)</td>
<td>0.67(0.10)</td>
<td>0.59(0.02)</td>
<td>0.58(0.02)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>5</td>
<td>0.75(0.05)</td>
<td>0.67(0.05)</td>
<td>0.59(0.02)</td>
<td>0.58(0.02)</td>
</tr>
</tbody>
</table>

Table 2: Average standardized root mean squared error (RMSE) for the clustering patterns estimation accuracy. Standard errors are given in the parentheses.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$\sigma^2_z$</th>
<th>$\sigma^2_{\varepsilon}$</th>
<th>FSCM</th>
<th>FCCD</th>
<th>FCMC</th>
<th>FClust</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1</td>
<td>1</td>
<td>0.04(0.005)</td>
<td>0.04(0.02)</td>
<td>0.34(0.11)</td>
<td>0.35(0.12)</td>
</tr>
<tr>
<td>0.2</td>
<td>1</td>
<td>5</td>
<td>0.13(0.02)</td>
<td>0.16(0.04)</td>
<td>1.12(0.25)</td>
<td>0.50(0.14)</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>1</td>
<td>0.37(0.19)</td>
<td>0.41(0.19)</td>
<td>1.08(0.24)</td>
<td>1.10(0.23)</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>5</td>
<td>0.19(0.09)</td>
<td>0.39(0.11)</td>
<td>1.13(0.23)</td>
<td>1.23(0.22)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.04(0.005)</td>
<td>0.04(0.005)</td>
<td>0.33(0.11)</td>
<td>0.34(0.13)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>5</td>
<td>0.13(0.02)</td>
<td>0.15(0.05)</td>
<td>0.32(0.12)</td>
<td>0.47(0.14)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>1</td>
<td>0.37(0.19)</td>
<td>0.38(0.19)</td>
<td>1.12(0.26)</td>
<td>1.13(0.27)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>5</td>
<td>0.20(0.12)</td>
<td>0.39(0.11)</td>
<td>1.12(0.21)</td>
<td>1.21(0.22)</td>
</tr>
</tbody>
</table>

Table 3: 100% Percentage of replications that identified correct number of clusters.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$\sigma^2_z$</th>
<th>$\sigma^2_{\varepsilon}$</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1</td>
<td>1</td>
<td>97</td>
<td>98</td>
</tr>
<tr>
<td>0.2</td>
<td>1</td>
<td>5</td>
<td>93</td>
<td>89</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>1</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>5</td>
<td>96</td>
<td>84</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>5</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>1</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>5</td>
<td>87</td>
<td>85</td>
</tr>
</tbody>
</table>
Table 4: Average estimation of dependence in cluster membership $\theta$. Standard errors are given in the parentheses.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$\sigma^2_s$</th>
<th>$\sigma^2_z$</th>
<th>FSCM</th>
<th>FCMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1</td>
<td>1</td>
<td>0.10(0.04)</td>
<td>0.14(0.07)</td>
</tr>
<tr>
<td>0.2</td>
<td>1</td>
<td>5</td>
<td>0.09(0.05)</td>
<td>0.15(0.07)</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>1</td>
<td>0.10(0.05)</td>
<td>0.15(0.07)</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>5</td>
<td>0.11(0.06)</td>
<td>0.16(0.07)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.23(0.03)</td>
<td>0.18(0.06)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>5</td>
<td>0.20(0.05)</td>
<td>0.20(0.06)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>1</td>
<td>0.14(0.05)</td>
<td>0.16(0.07)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>5</td>
<td>0.20(0.07)</td>
<td>0.18(0.08)</td>
</tr>
</tbody>
</table>

Table 5: Average estimation of conditional dependence $\sigma^2_s$. Standard errors are given in the parentheses.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$\sigma^2_s$</th>
<th>$\sigma^2_z$</th>
<th>FSCM</th>
<th>FCCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1</td>
<td>1</td>
<td>1.11(0.26)</td>
<td>1.22(0.80)</td>
</tr>
<tr>
<td>0.2</td>
<td>1</td>
<td>5</td>
<td>1.00(0.19)</td>
<td>0.99(0.20)</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>1</td>
<td>35.8(25.5)</td>
<td>34.0(24.2)</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>5</td>
<td>3.61(0.60)</td>
<td>3.30(0.66)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1.12(0.25)</td>
<td>1.16(0.31)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>5</td>
<td>1.01(0.20)</td>
<td>1.02(0.19)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>1</td>
<td>34.9(25.2)</td>
<td>33.8(22.5)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>5</td>
<td>3.51(0.61)</td>
<td>3.18(0.66)</td>
</tr>
</tbody>
</table>

Table 6: Average estimation of the noise level $\sigma^2_z$. Standard errors are given in the parentheses.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$\sigma^2_s$</th>
<th>$\sigma^2_z$</th>
<th>FSCM</th>
<th>FCCD</th>
<th>FCMD</th>
<th>FClust</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1</td>
<td>1</td>
<td>1.00(0.02)</td>
<td>1.01(0.03)</td>
<td>2.38(0.11)</td>
<td>1.52(0.10)</td>
</tr>
<tr>
<td>0.2</td>
<td>1</td>
<td>5</td>
<td>4.81(0.07)</td>
<td>4.81(0.08)</td>
<td>5.38(0.30)</td>
<td>5.24(0.13)</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>1</td>
<td>1.34(0.14)</td>
<td>1.35(0.13)</td>
<td>2.34(0.34)</td>
<td>2.33(0.28)</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>5</td>
<td>4.87(0.10)</td>
<td>5.02(0.13)</td>
<td>6.13(0.26)</td>
<td>6.05(0.30)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1.01(0.02)</td>
<td>1.00(0.02)</td>
<td>1.54(0.11)</td>
<td>1.51(0.10)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>5</td>
<td>4.82(0.08)</td>
<td>4.83(0.07)</td>
<td>5.37(0.11)</td>
<td>5.22(0.13)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>1</td>
<td>1.33(0.13)</td>
<td>1.33(0.14)</td>
<td>2.37(0.31)</td>
<td>2.36(0.32)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>5</td>
<td>4.90(0.12)</td>
<td>5.02(0.13)</td>
<td>6.10(0.26)</td>
<td>6.03(0.30)</td>
</tr>
</tbody>
</table>

17
not significantly lower for strong spatial dependence in $Y|Z$ ($\sigma_s = 1$ vs. $\sigma_s = 5$).

- (Table 2) The error of the cluster patterns is lowest for the less restrictive model introduced in this paper except for the setting $\sigma_s = 1$ vs. $\sigma_s = 5$. A significant improvement in accuracy of FSCM over the model accounting for spatial conditional dependence only is under strong spatial dependence.

- (Table 3) AIC and BIC both correctly identify the number of clusters with a high probability except when $\sigma_s^2$ is much larger than $\sigma_\epsilon^2$. Under such settings, $\sigma_s^2$ is poorly estimated.

- (Table 4) The scaling parameter $\theta$ for the spatial dependence of $Z$ is under-estimated for strong spatial dependence in $Z$ ($\theta = 0.5$). One reason for this bias is that we use only five nearest neighbors. This suggests that a method for selecting the number of neighbors would improve the estimation accuracy of the spatial dependence of $Z$.

- (Table 5) The accuracy of the parameter $\sigma_s^2$ is very low when $\sigma_s^2$ is much larger than $\sigma_\epsilon^2$.

- (Table 6) The estimates for $\sigma_\epsilon^2$ are significantly more accurate for FSCM and FCCD.

5 Classification of Service Accessibility

5.1 Preliminaries

Accessibility is measured as the utilization-scaled travel cost of a community $U$ to the nearby sites in a service network consisting of multiple geographically distributed service sites: $S = \{s_1, \ldots, s_n\}$. We note that the number of service sites will vary from one year to another. For simplicity, we introduce the accessibility measure for a general value $n$. In this paper, we analyze the accessibility to financial services. We acquire the mail addresses of the financial service sites from the Federal Deposit Insurance Corporation (FDIC). In our study we use yearly data from 1994 to 2009 - a total of $m = 16$ time points. We geocoded the site mail addresses using ArcView - a GIS software provided by Environmental Systems Research Institute (ESRI).

One challenge in measuring service accessibility is defining the travel cost of the residents in a community $U$ to the sites in the service network $S$. In the research works so far, the travel cost is calculated as the average or minimum distance between the centroid of the region $U$ and the nearby sites in the service network. (Talen 1996, 1998, 2001; Lovett
et al. 2002) However, communities occupy uneven geographic areas varying in size, and therefore, their simplified representation by their centroids is restrictive. In this research, we instead represent a community by a sample of $B$ locations in the geographic space of the neighborhood $U$, $u_1, \ldots, u_B \in U$, and compute the street-network distances from these sample locations to the service sites. In our implementation, we use $B = 5$.

In this research paper, we measure the accessibility from a community $U$ to the network $S$ as a summary of the street-network distances \( \{d(u_b, s_j)\}_{b=1,\ldots,B; j=1,\ldots,n} \) adjusted for the service utilization. Specifically, we use the travel cost to measure how much a person in a given neighborhood $U$ is required to travel to a service site. We compute the accessibility of a neighborhood to a service network as the average utilization-adjusted travel cost over the sampling locations in community $U$ and in year $t$ using

$$Y(U, t) = \frac{1}{B} \sum_{b=1}^{B} \left( C(u_b, t)^\beta W(u_b, t) \right), \quad (11)$$

where $C(u_b, t)$ is the travel cost at the sample location $u_b$ measured as the average street-network distance to the closest $Q$ service sites available at time $t$ (in our study, $Q = 3$), $W(u_b, t)$ is the utilization weight at location $u_b$ and $\beta$ is a distance utility parameter. We estimate $\beta$ using robust linear regression: $\log(W(u_b, t)) \sim -\log(C(u_b, t))$ and the estimated value is $\beta = 0.18$. In this paper, the utilization weight $W(u_b, t)$ is measured as the population rate divided by the service rate at location $u_b$. The population and service rates are estimated using kernel density estimation.

Dividing the geographic space into contiguous spatial units $U_{s_j}$, $j = 1, \ldots, n$, where each spatial unit corresponds to a neighborhood, the accessibility varies across the geographic space: $Y(U_{s_j}, t) = Y_j(t)$. In this research, census tracts are used as proxy for communities as they are delineated with local input and designed to represent neighborhoods. There are a total of 1624 census tracts in Georgia. The accessibility maps for Georgia and Atlanta for years 1994 and 2009 are included in the Supplemental Materials 2 and 3.

We apply the clustering method introduced in this paper to the service accessibility curves, $Y_j(t)$ $j = 1, \ldots, n$, for Georgia ($n = 1624$). Our objective is to cluster the accessibility curves according to their shape regardless of scale. In other words, we are interested...
in identifying accessibility patterns after removing the state-level overall mean and after adjusting the accessibility curves to have similar scales. The adjusted accessibility patterns reveal how regional deviations from the overall state-level trend have changed over a time period.

In the next section, we discuss our findings based on a series of plots which summarize the spatial and temporal accessibility patterns.

5.2 Discussion

The analysis is performed on log-transformed accessibility measures to make the data approximately normally distributed. When interpreting the figures in this paper, one has to bear in mind that large values for the accessibility measure (high travel cost) correspond to low access to financial services.

We apply the clustering algorithm after rescaling the accessibility curves; our objective is to cluster by shape regardless of scale. We re-scale each log-transformed accessibility curve by subtracting its overall mean and adjusting the scale. Because we log-transform the
data, the function (location)-specific mean in the additive model is the *scaling factor* in the multiplicative model on the original scale. Figure 3 (a) presents the scaling factors computed at the community level as the average over the accessibility measure for the years 1994 to 2009. Comparing the spatial map of the scaling factors in Figure 3(a) to the 2009 accessibility map in Figure 3(b), we find that the scaling factors explain a significant proportion of the spatial variability in the data. Specifically, the spatial maps in these figures are similar. Low utilization-scaled travel cost regions in Figure 3(b) correspond to low scaling factors (shown in blue to green in Figure 3 (a)) and high utilization-scaled travel cost regions in Figure 3(b) correspond to high scaling factors (shown in orange to red in Figure 3 (a)).

We employed the model selection criteria in Section 3.3 for identifying the number of clusters in the log-accessibility data. However, as provided in Figure 4, both the AIC and BIC values decrease with the number of clusters ranging from 2 to 10. This suggests that the number of clusters is large but these criteria do not provide an upper threshold for the number of clusters. A potential justification for this result is that there are outlying random functions, which do not belong to one given cluster. Subsequently, we visually assessed the cluster patterns for a range of the number of clusters varying from 2 to 10. A clustering with $C > 5$ results in additional cluster patterns that are not significantly different from the ones in Figure 5. Consequently, we decided to analyze the clustering with $C = 5$ as provided
Figure 5: Temporal trends for the accessibility measure in Georgia.

Figure 5 presents the estimated global trend $\mu(t)$ and cluster-specific effects over 1624 accessibility curves in Georgia. We first highlight that the global time-varying access to financial services $\mu(t)$ has increased from 1994 to 2009 (Figure 5 upper most left panel). (Note that this is equivalent to a global decrease in the utilization-adjusted travel cost within the period under study.) We conclude that throughout Georgia the accessibility to financial services has strengthened over the past 16 years.

The additional cluster patterns, $\mu_k(t), k = 1, \ldots, 5$ in Figure 5 are adjusted for the overall trend $\mu(t)$. For example, the upward trend in cluster 1 is a near-level trend for the cluster when combined with $\mu(t)$. It is only an upward trend relative to the global downward trend seen in the first panel. Therefore, the cluster trends in the five panels are relative to the overall trend, rather than absolute trends. We are interested in these relative trends since our objective is to highlight local patterns adjusted for the overall state trend revealing local variations in access to financial services. Below we therefore discuss the cluster patterns in
relative terms.

Cluster 1 corresponds to an increase in the travel cost (weakening access to financial services) with a shift in the trend around 2005. In contrast to Cluster 1, Cluster 2 consists of the communities with a decreasing access starting with 1999 to 2009; this cluster is the only one which shows a weakening accessability (increased travel cost) in the past 10 years. This cluster includes 356 communities, almost 20% of the total communities. Cluster 3 is the most dynamic with two upward and one downward trends within the 16 year period. This cluster consists of 535 communities. The communities in the last two clusters experience a strengthening of the access to financial services but with slightly different slopes. Possibly, these two clusters could be merged in one.

Figure 6 shows the map of the clustering membership for Georgia and Atlanta. Two consistent patterns are for cluster 2 which primarily maps to rural communities and for cluster 3 with the most dynamic accessibility pattern which maps to urban areas, extensively present in central Atlanta. This points to a decrease in access to financial services in urban areas with more than one upward or downward shift. Importantly, a large number of communities with predominant black and Hispanic population are classified in clusters 1 and 2 whereas communities with high percentage of white population have mixed clustering. Generally, communities corresponding to cluster 1 and 2 (34% of the total communities in Georgia) are potential new markets for financial service providers.

On the other hand, most communities in north Atlanta are in clusters 4 and 5 with a significant increase in access (decrease in travel cost) to financial services. In contrast, communities in south Atlanta experience high dynamics in accessibility generally with a weaker access to financial services throughout the period of 16 years. We note that one discrepancy between north and south Atlanta is the division of white and black population; the percentage of white population is very high in north whereas the percentage of back population is very high in south (see additional demographic plots in the Supplemental Materials 1 and 2 of this paper). Although there is also a discrepancy in the income level between north and south, many communities with low income in north fall in cluster 4 and 5 whereas communities with similar income levels in south primarily fall in clusters 1 and 2. This points to potential opportunities in South Atlanta which will not only suggest new
Figure 6: Cluster map for the accessibility measure in Georgia and Atlanta, GA. Each color in the cluster map corresponds to a different cluster.

markets for financial service providers but also will reduce inequities in access to financial services throughout Atlanta.

6 Conclusions

The spatial-functional clustering method in this paper is a means for summarizing the time-dependent cluster effects of a large number of time-varying random functions which are spatially interdependent. The proposed clustering method offers readily interpretable summaries of the temporal changes in a space-time varying process.

From a methodological point of view, one important aspect of our clustering model is that it allows for spatial dependence in the joint data \((Y, Z)\). Under this assumption, the spatial and temporal trends are accurately estimated by borrowing information across curves in the nearby locations. In our simulation study, we found that accounting for spatial dependence in the conditional distribution of \(Y|Z\) has most impact on the estimation accuracy of the cluster membership and of the cluster patterns when there is significant spatial dependence in the data. In additional simulation studies (not reported here), we found that for simulated
data with independent errors, accounting for the spatial dependence in $Z$ results in more accurate estimation than when the spatial dependence in $Z$ is ignored. Similar simulation insights have been reported by Blekas et al. (2007).

Given the two contributions in enhancing the computational efficiency of the estimation algorithm (Section 3), allowing for spatial dependence in the joint data ($Y, Z$) does not lead to a significant computational effort. Therefore, even when there is reduced spatial dependence in the clustering membership and conditional distribution, there is not a loss in the estimation accuracy with no additional significant computational cost in applying the functional spatial clustering model as opposed to the more restrictive models.

In our motivating application, describing financial service accessibility in Georgia, we find that over a period of 16 years, there is a steady increase in access to financial services (decrease in the utilization-adjusted travel cost). In addition, the summary accessibility trends significantly vary over time. A large percentage of the communities in Georgia with a significant decrease in access to financial services are in areas with predominant black and Hispanic population. This points to potential market opportunities for service providers. Noteworthy, service accessibility is very dynamic in most of central Atlanta with upward and downward changes over the past 16 years. This may be because most communities in Atlanta have experienced high mobility of both people and services.

Other potential applications of the clustering approach in this paper include location-based market segmentation, performance analysis of a service enterprise with multiple service sites where the performance may be measured by the sales divided by the site size, or summarizing product-level sale patterns for a retail service provider.

References


