BAYESIAN CLUSTERING OF SPATIAL FUNCTIONAL DATA WITH APPLICATION TO A HUMAN MOBILITY STUDY DURING COVID-19

BY BOHAI ZHANG¹, HUIYAN SANG²,* AND ZHAO TANG LUO²,† AND HUI HUANG³

¹School of Statistics and Data Science, LPMC, and KLMFASR, Nankai University, Tianjin, China, bohai_zhang@nankai.edu.cn
²Department of Statistics, Texas A&M University, College Station, Texas, USA, *huiyan@stat.tamu.edu, †zt.luo@tamu.edu
³School of Mathematics, Sun Yat-sen University, Guangzhou, China, huangh89@mail.sysu.edu.cn

The coronavirus (COVID-19) global pandemic has made a significant impact on people’s social activities. Cell phone mobility data provide unique and rich information on studying this impact. The motivating dataset of this study is the daily leaving-home index data at Harris County in Texas provided by SafeGraph. To study changes in daily leaving-home index and how they relate to public policy and socio-demographic variables, we propose a new Bayesian wavelet model for modeling and clustering spatial functional data, where domain partitioning is achieved by operating on the spanning trees. The resulting clusters can have arbitrary shapes and are spatially contiguous in the input domain. An efficient tailored reversible jump Markov chain Monte Carlo algorithm is proposed to implement the model. The method is applied to the spatial functional data of the daily percentages of people who left home. We focus on the time period covering both lock-down and phased re-opening in Texas during the COVID-19 pandemic, and study the changing behaviors of those functional curves. By linking the clustering results with the socio-demographic information, we identify several covariates of census blocks that have a noticeable impact on the clustering patterns of people’s mobility behaviors.

1. Introduction. Since the outbreak of the coronavirus disease 2019 (COVID-19, e.g. Jiang et al., 2020; Guan et al., 2020), there has been a variety of responses from governments at various levels and individuals, including non-pharmacological policy interventions such as social distancing orders for disease spread control, and phased re-opening orders to minimize the adverse economic consequences (Hsiang et al., 2020; Viner et al., 2020; Lewnard and Lo, 2020). In the United States, both policies and individual behavioral responses have been changing with the rapidly evolving situation of the pandemic (Warren and Skillman, 2020; Gao et al., 2020). Understanding variation in human mobility patterns and responses to COVID-19 policy interventions, across space and over time, is a fundamental issue in public health policies and social behavior studies.

Our motivating dataset is the daily leaving-home index (LHI) data provided by SafeGraph, a company that collects and analyzes GPS trajectory data from a representative sample of mobile devices in the U.S., including cell phones and other wearable GPS devices. Specifically, for each census block group (CBG), a time series of daily percentages of people leaving home is derived from the GPS trajectory data, where the home location for each individual is determined by the primary night time location using a rolling 6-week window. We focus on Harris County in Texas as the study area, and February 17 to June 23 in 2020 as the study time window. Harris County consists of 2144 census block groups in total, most of which belong to the greater Houston metropolitan area. According to the 2015–2018 American Community

Keywords and phrases: Bayesian wavelet model, mobility behavior, reversible jump Markov chain Monte Carlo, spanning tree, spatial functional data.
Survey Data (ACS) provided by the U.S. Census Bureau, Harris County has a total population of 4,602,523, making it the third-most populous county in the United States; and the total number of sampled devices by SafeGraph in Harris County is approximately 460,000, around 10% of the total population. On March 24, the Harris County judge issued a stay-at-home order prohibiting gatherings and providing that residents stay home other than to perform Essential Services. Texas Governor later issued a state-wide phased re-opening order with Phase I re-opening starting from April 30, phase II from May 18, and then Phase III from June 3. The diverse demographics and dynamic policy interventions during COVID-19 make Harris County a good candidate to study COVID-19 policy compliance and mobility patterns.

The objective of this study is to understand how human mobility patterns vary in space and time during COVID-19, and how they are related to social demographic variables such as race, age, income, and education. Figure 1 shows time-averaged LHI values at the CBG level in Harris County. Clearly, LHI data exhibit large-scale spatial patterns, with higher values in the eastern region and the northwestern corner and smaller values in the southwestern part. Local spatial patterns are also noticeable: A few CBGs have substantially higher or lower values than their neighboring CBGs. We treat the LHI time series at each CBG as spatial functional data and propose a Bayesian clustering model to investigate the clustering patterns of temporally dynamic mobility patterns in Harris County. In Section 5, we will show that our model has the capacity of finding big clusters of reasonable homogeneity, as well as detecting specific small clusters that behave very differently from their neighbors.

We further conducted subregional analysis to explain the clustering mobility patterns using several key socio-demographic variables from the ACS data. We found that the ethnic compositions and education levels of census blocks appear to have a large-scale impact on the clustering patterns of the LHI series, resulting in big clusters in which certain ethnic groups have a dominant proportion; the age compositions seem to have a more local impact that results in a few isolated small clusters. As the pandemic is ongoing, policy interventions are still being considered and crafted. The methodology and findings of this work may offer important tools and insights for future public policy making and policy compliance.

Various spatial functional data clustering methods have been proposed in the literature. Romano, Verde and Cozza (2011) and Giraldo, Delicado and Mateu (2012) defined dissimilarities among spatial functions based on spatial variograms and cross-variograms and then applied heuristics or geometric procedures based clustering algorithms such as k-means. Jiang and Serban (2012) proposed a mixture model and introduced spatial dependence by modeling the cluster memberships using an auto-regressive Markov random field. Hu et al.
BAYESIAN CLUSTERING OF SPATIAL FUNCTIONAL DATA

(2022) proposed a Bayesian mixture of finite mixtures model based on the similarity measure for clustering spatial income distributions. While these methods represent advancements in identifying spatial clusters compared with the conventional functional clustering methods assuming i.i.d. curves (Jacques and Preda, 2014), there are several limitations that need to be addressed. First, for many spatial application problems, spatially contiguous clustering configurations are desired due to easy interpretation. While capable of borrowing spatial information for clustering, many existing methods may not fully guarantee contiguity in the resulting clusters such as the mixture model based approaches in Jiang and Serban (2012) and Hu et al. (2022). Moreover, model-based clustering often relies on mixture models that require specifying the number of clusters (e.g., Banfield and Raftery, 1993; Yeung et al., 2001; Fraley and Raftery, 2007), which is often difficult in practice. Finally, clusters of spatial functional data may have highly irregular cluster boundaries. Nevertheless, many Bayesian model-based space partitioning priors such as Voronoi tessellation methods (e.g., Denison and Holmes, 2001; Holmes et al., 2005; Kim, Mallick and Holmes, 2005; Zhang et al., 2016; Payne et al., 2020) and binary tree-based partitioning methods (Chipman, George and McCulloch, 1998; Denison, Mallick and Smith, 1998; Konomi, Sang and Mallick, 2014) are limited to convex-polygon-shaped boundaries or rectangular-shaped boundaries.

More recently, model-based spanning tree partitioning methods have been proposed in the literature (Teixeira, Assunção and Loschi, 2019). Li and Sang (2019) proposed a fused lasso regularization method for the spatially varying coefficient model, which uses a fixed minimum spanning tree (MST) as the “spatial order” to encourage spatial homogeneity between the coefficients at adjacent locations. By cutting edges of the MST, clusters of homogeneous coefficients are created with very flexible shapes (e.g., shapes are not necessarily convex or having boundaries parallel to the input-space axes). Luo, Sang and Mallick (2021) proposed a Bayesian random spanning tree (RST) partitioning method to find the true spatial order by treating different spanning trees for each covariate as unknown parameters. A unique property of the RST partitioning model is that the sample space of partitions induced from the Bayesian random spanning tree models accommodates all possible contiguous partitions. As such, the resulting cluster configurations can have very flexible shapes and sizes. However, both work are concerning only spatial scalar responses.

Our main methodological contribution is to extend the RST-based partitioning approach for functional data clustering analysis. Specifically, we consider a wavelet-based representation of functional data, and impose both sparsity and spatially clustering inducing priors for wavelet coefficients. Construction of the spanning trees relies on an undirected connected graph, \( G \). Previous work often rely on the spatial Euclidean distance for graph constructions (Li and Sang, 2019; Luo, Sang and Mallick, 2021). In this work, we will discuss how to specify it using both spatial information as well as covariate information via distance metric learning. By imposing proper priors on the spanning trees and its associated operations, we further propose a tailored efficient reversible jump Markov chain Monte Carlo (RJ-MCMC) algorithm (Green, 1995) to infer the number of clusters. Then, the “optimal” estimate of the space partition structure can be readily obtained based on the empirical pairwise probability matrix of clustering (Dahl, 2006) or the maximum a posteriori (MAP) principle. Finally, we remark that although we focus on the clustering of spatial functional data in our application, the method is applicable to the clustering of functional data on a wide range of complex geometric domains such as networks, thanks to the flexibility of the graph-based approach.

The rest of the paper is organized as follows. In Section 2, we introduce the wavelet-based clustering model for functional data, as well as the clustering strategies through the RST-based approach. Bayesian inference details are provided in Section 3, which consist of both the within-cluster and space-partition inferences; specification of the hyper-parameter graph \( G \) is also discussed in this section. Section 4 gives simulation studies of our proposed model,
where the numerical results demonstrate the effectiveness of our model in identifying the true clusters and yielding reasonable parameter-estimation and prediction results. Then, we apply our model to the LHI dataset in Section 5, aiming to study the mobility behaviors of people in Harris County of Houston during the COVID-19 pandemic. Finally, Section 6 gives a brief summary and discussion of this research. Supplementary Material contains the computation and implementation details of the MCMC algorithm, as well as the supplementary results of simulation studies and the real-data analysis.

2. Methodology.

2.1. Wavelet functional representations. When modeling functional observations, the wavelet-based methods are popular due to their “universality” of representing a large class of function spaces, such as Besov or Triebel spaces (e.g., Daubechies, 1992). Since the wavelet bases are orthogonal and compactly supported, the additive normal white noises affect the wavelet coefficients equally, and the true signals can be extracted by thresholding. Therefore, the wavelet representation of functions is usually parsimonious with sparse coefficients, which enjoys computational efficiency for large datasets. The wavelet shrinkage estimation with thresholding rules became popular due to its optimal properties for a broad range of functional spaces (Donoho and Johnstone, 1994, 1995).

We first introduce the wavelet model for a single functional series. Let \( y_s(t) \) denote a functional datum observed at location \( s \) and time point \( t \), where \( s \in \mathcal{D} \subseteq \mathbb{R}^d \), for \( d = 1, 2, \ldots \), and \( t \in [0, T] \), for \( T > 0 \). Here, location \( s \) may lie on a complex domain such as spatial regions with irregular boundaries, peninsulas, and interior holes, networks representing relations among observations such as road networks or social networks, or covariate spaces. Suppose we observe data at \( n \) locations, denoted as \( \mathcal{S} \equiv (s_1, s_2, \ldots, s_n)' \), and at each location \( s_i \), the data are collected at \( m \) equally spaced time points, \( \{t_1, t_2, \ldots, t_m\} \). Here, we assume that \( m = 2^J \) for an integer \( J \geq 0 \). For our real data problem, we have \( n = 2144 \) spatial locations and \( m = 128 \) daily time points. We model \( y_s(t) \) as,

\[
y_s(t) = f_s(t) + \epsilon_s(t),
\]

where \( \{f_s(t)\}'s \) are unknown functions and \( \{\epsilon_s(t)\}'s \) are the measurement errors which are assumed to be independently distributed normal random variables, denoted as \( \mathcal{N}(0, \sigma^2(s)) \). For the true signal \( f_s(t) \), we use wavelet basis functions to represent it as,

\[
f_s(t) = \phi_{00}(t)\beta_{00}(s) + \sum_{j=1}^{J} \sum_{k=1}^{2^{j-1}} \phi_{jk}(t)\beta_{jk}(s),
\]

where \( \beta_{00}(s) \) is the (mean-level) scaling coefficient, \( \{\beta_{jk}(s) : j = 1, \ldots, J, k = 1, \ldots, 2^{j-1}\}'s \) are referred to as the detail coefficients, and \( J = \log_2(m) \) is the finest resolution of the wavelet representation. Each wavelet function, \( \phi_{jk}(t) \), is obtained based on a scale and location transform of a mother wavelet function, and \( j \) and \( k \) are the corresponding scale and location indices, respectively.

Following Bayesian wavelet shrinkage methods (e.g., Abramovich, Sapatinas and Silverman, 1998; Clyde and George, 2000; Ray and Mallick, 2006), we impose a mixture prior on the detail coefficients to achieve sparsity,

\[
\beta_{jk}(s) \sim \pi_j(s)\mathcal{N}(0, \lambda_j(s)\sigma^2(s)) + (1 - \pi_j(s))\delta_0,
\]

where at location \( s \), \( \pi_j(s) \) is the selection probability for obtaining nonzero wavelet coefficients at the \( j \)-th resolution, \( \lambda_j(s) \) is a scaling parameter (modeling the signal-to-noise ratio), and \( \delta_0 \) is the Kronecker delta function with a point mass at zero. As such, the model allows
the prior assumption on the sparsity level to not only vary at different wavelet resolutions but also in space. If we introduce a latent indicator variable $\gamma_{jk}(s)$ that equals 1 with probability $\pi_j(s)$, the selection prior in (3) can be written as,

$$
\beta_{jk}(s)|\gamma_{jk}(s), \lambda_j(s) \sim N(0, \gamma_{jk}(s)\lambda_j(s)\sigma^2(s)).
$$

Note that for the scaling coefficient $\beta_{00}(s)$, its prior distribution also has the same form to (3) and (4), but $\pi_0(s)$ and $\lambda_0(s)$ are equal to one.

For the variance parameters, $\{\sigma^2(s)\}$ and $\{\lambda_j(s)\}$, we also assume that they are spatially varying and impose conjugate inverse gamma priors,

$$
\sigma^2(s) \sim IG(a(s), b(s)), \lambda_j(s) \sim IG(c_j(s), d_j(s)).
$$

Then, by combining the data model in (1), the process model in (2), and priors in (4) and (5), we have the following Bayesian hierarchical wavelet model,

$$
y_s(t)|\{\beta_{jk}(s)\}, \sigma^2(s) \sim N\left(\phi_{00}(t)\beta_{00}(s) + \sum_{j=1}^{J} \sum_{k=1}^{2^{j-1}} \phi_{jk}(t)\beta_{jk}(s), \sigma^2(s)\right),
$$

$$
\beta_{jk}(s)|\gamma_{jk}(s), \lambda_j(s), \sigma^2(s) \sim N\left(0, \gamma_{jk}(s)\lambda_j(s)\sigma^2(s)\right),
$$

$$
\gamma_{jk}(s)|\pi_j(s) \sim Ber(\pi_j(s)), \sigma^2(s) \sim IG(a(s), b(s)), \lambda_j(s) \sim IG(c_j(s), d_j(s)),
$$

where Ber($p$) denotes a Bernoulli distribution with parameter $p$.

2.2. Random spanning tree priors for partitions of latent model parameters. The hierarchical model given by (6) involves high-dimensional parameters that are varying with $s$ and hence, treating those parameters to be unknown would lead to the over-fitting problem. Cluster inducing prior models have been imposed on wavelet parameters in the previous literature (e.g., see Ray and Mallick, 2006; Suarez et al., 2016). However, many methods are built upon Dirichlet processes that are more suitable for i.i.d. curves. In spatial settings, the latent functions, $\{f_s(t)\}$, usually behave similarly at “neighboring” locations, which implies a spatially clustering effect for their coefficients in the domain $\mathcal{D}$. Detecting and distinguishing this clustering effect is crucial to study the heterogeneous impact of covariates (or latent variables) on the response variable of interest.

Hence, we assume that the varying wavelet coefficients are the same within a cluster but allow them to differ across clusters. The clusters are obtained by partitioning the study domain $\mathcal{D}$ into $k_p$ disjoint subregions, denoted as $\{\mathcal{D}_i: i = 1, \ldots, k_p\}$. The goal is to learn this underlying partition structure correctly and to estimate the cluster-wise wavelet coefficients as well as variance parameters.

As reviewed in the Introduction, many traditional Bayesian model-based clustering methods can’t guarantee spatial contiguity in clustering; and among those exceptions that guarantee spatial contiguity, considerable constraints are imposed on the shape of clusters. In this work, we consider a Bayesian spatial partitioning prior based on random spanning trees that can lead to very flexible shapes and sizes of clusters.

To introduce the spatial clustering method, we first define spatially contiguous clusters of points. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote an undirected graph representing the prior knowledge about the relational or neighborhood information among locations on $\mathcal{D}$, where $\mathcal{V} = \{s_1, \ldots, s_n \in \mathbb{R}^d, n < \infty\}$ is the set of vertices, and $\mathcal{E}$ is the edge set. A graph $\mathcal{G}$ is said to be connected, if for any two vertices there exists a path between them. A subgraph $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ of $\mathcal{G}$ is called a connected component of $\mathcal{G}$, if it is connected and there is no path between any vertex in $\mathcal{V}_s$ and any vertex in the difference set $\mathcal{V}\setminus\mathcal{V}_s$. It is then natural to define spatially contiguous clusters with respect to $\mathcal{G}$ as all the connected components; and a spatially contiguous
partition of $\mathcal{V}$ as the collection of disjoint connected components such that the union of vertices is $\mathcal{V}$. Very recently, Luo, Sang and Mallick (2021) proved that there exists an equivalent but much more compact representation of spatially contiguous partitions using spanning tree graphs, which are defined as all connected subgraphs of $\mathcal{G}$ without any cycles. Specifically, for any given possible spatially contiguous partition defined as above, there exists at least one spanning tree of $\mathcal{G}$, denoted as $\mathcal{T}$, such that the given partition can be induced by removing a subset of edges of $\mathcal{T}$. Figure 2 demonstrates an example of creating a contiguous partition with irregular cluster shapes by removing edges from a spanning tree of a Delaunay triangulation graph.

Following a similar spirit as in Luo, Sang and Mallick (2021), we consider a very flexible Bayesian spanning-tree spatial random partitioning model by assigning prior models for spanning trees as well as the removed edge set given a spanning tree. Specifically, let $\mathcal{P}$ denote a partition of vertices into $k_p$ clusters that is driven by $\mathcal{T}$, and let $\mathbf{w} = \{w_{ij}\}$ be a random vector of the edge weights associated with $\mathcal{G}$; we impose the following priors on $\mathbf{w}$:

$$\mathcal{T} = \text{MST}(\mathbf{w}), \quad w_{ij} \overset{i.i.d.}{\sim} \text{Unif}(0, 1),$$

$$p(\mathcal{P}|\mathcal{T}, k_p) \propto 1, \quad p(k_p) \propto (1 - c)^{k_p},$$

(7)

where $\text{MST}(\mathbf{w})$ is the MST of the graph $\mathcal{G}$ using edge weights $\mathbf{w}$, which is defined as the spanning tree that has a minimal sum of edge weights. We assign uniform priors on edge weights to induce a prior model on the resulting spanning tree. The prior of the space partition implies that given $\mathcal{T}$ and $k_p$, we impose equal probabilities on selecting $(k_p - 1)$ out of $(n - 1)$ edges of $\mathcal{T}$ to obtain a space partition $\mathcal{P}$. The prior on the number of clusters in (7) follows Knorrheld and Raser (2000), which is proportional to a geometric distribution. Here, the hyper-parameter $c \in [0, 1)$ is chosen to penalize large values of $k_p$, i.e., model complexity. Clearly, $c = 0$ leads to a discrete uniform prior on $k_p$, and $c \rightarrow 1$ imposes a large penalty on large values of $k_p$. If we set $k_p \leq k_{\text{max}}$ to truncate the possible values of $k_p$, it is easy to obtain that $E(k_p) = 1/c - k_{\text{max}}(1 - c)^{k_{\text{max}}}/(1 - (1 - c)^{k_{\text{max}}})$ when $c \in (0, 1)$, and $E(k_p) = (1 + k_{\text{max}})/2$ when $c = 0$.

In summary, under the random spanning tree prior model, a random spatially contiguous partition is generated by:

- constructing a graph $\mathcal{G}$ (discussed in Section 3.3) connecting all observed spatial locations.
- for each edge in $\mathcal{G}$, independently generating an edge weight from Unif$(0, 1)$ and then constructing the MST, denoted as $\mathcal{T}$, that minimizes the total edge weights.
- generating the number of clusters, $k_p$, from a (truncated) geometric distribution in (7), and then randomly selecting and removing $(k_p - 1)$ edges from the $(n - 1)$ edges of $\mathcal{T}$. 

**FIG 2.** (a) A Delaunay Triangulation graph with edges longer than 0.2 removed; (b) An example of 5 connected components of the graph in (a); (c) An example of 5 induced clusters by removing the set of 4 red dashed edges from a spanning tree of the graph in (a). Different clusters are marked by different colors.
Specifically, directly obtained from their closed-form full conditional distributions using Gibbs samplers. \( \beta \), \( \lambda \), \( S \) of the within-cluster parameters, conditional on a partition of the location set \( \mathcal{D} \), will be shown in Section 3.

The orthogonality property of \( X \) where \( X \) denotes the Kronecker product, and the measurement-error vector \( \epsilon \) of error vectors, \( \epsilon \). Then, for any observed curve at \( s \in \mathcal{D} \), denoted as \( y_s^{(i)}(t) \), the Bayesian hierarchical model is,

\[
\begin{align*}
   y_s^{(i)}(t) | \bar{\beta}_i, \sigma^2_i & \sim N \left( \phi(t) \bar{\beta}_{i,00} + \sum_{j=1}^{J} \sum_{k=1}^{2^{j-1}} \phi_{jk}(t) \bar{\beta}_{i,jk}, \sigma^2_i \right), \\
   \bar{\beta}_i | \{ \gamma_{i,jk} \}, \{ \lambda_{i,j} \}, \sigma^2_i & \sim N(0, \sigma^2_i V_i), \\
   \gamma_{i,jk} | \pi_{i,j} & \sim \text{Ber}(\pi_{i,j}), \quad \sigma^2_i \sim \text{IG}(c_{i,j}, d_{i,j}), \quad \lambda_{i,j} \sim \text{IG}(c_{i,j}, d_{i,j}),
\end{align*}
\]

(8)

where \( V_i \) is a diagonal matrix of size \( m \), with \( \{ \gamma_{i,jk} \} \) as its diagonal entries.

Let \( S_i = \{ s_{i,1}, \ldots, s_{i,n_i} \} \) be the location set of the \( i \)-th cluster, which has \( n_i \) observations such that \( \sum_{i=1}^{k_p} n_i = n \). Let \( Y(s) = (y_s(t_1), \ldots, y_s(t_m))^\prime \) denote the vector of \( m \) temporal observations at location \( s \), and let \( Y_i = (Y(s_{i,1})^\prime, \ldots, Y(s_{i,n_i})^\prime)^\prime \) denote the vector of the \( mn_i \) observations for the \( i \)-th cluster. We similarly define the corresponding measurement-error vectors, \( \epsilon(s) \) and \( \epsilon_i \). Then, for the \( i \)-th cluster, the data model is,

\[
Y_i = X_{c_i} \bar{\beta}_i + \epsilon_i,
\]

where \( X_{c_i} = 1_{n_i} \otimes X \), \( X \in \mathbb{R}^{m \times m} \) is obtained from the (inverse) discrete wavelet transform (Shensa, 1992) of \( Y_i \), \( \otimes \) denotes the Kronecker product, and the measurement-error vector \( \epsilon_i \) follows \( N(0, \sigma^2_i I_{mn_i}) \). Note that \( X \) is an orthogonal matrix such that \( X'X = XX' = I_m \). The orthogonality property of \( X \) can be utilized to simplify the computations for the posterior sampling of model parameters, which will be shown in Section 3.

2.3. Prior models for parameters given a partition. We now discuss the Bayesian functional clustering model conditional on a spanning tree. Given a partition of \( k_p \) clusters, all the functional observations in the \( i \)-th cluster share the same wavelet-coefficients vector, denoted as \( \bar{\beta}_i = (\bar{\beta}_{i,00}, \bar{\beta}_{i,11}, \ldots, \bar{\beta}_{i,J2^{J-1}})^\prime \in \mathbb{R}^m \), for \( i = 1, 2, \ldots, k_p \). The measurement-error variances, \( \{ \sigma^2(s) \} \), may show heteroscedasticity across clusters, and we account for this by specifying a cluster-specific parameter \( \sigma^2_i \) to the \( i \)-th cluster. Similarly, the scale parameters \( \{ \lambda_{i,j} \} \) and the latent variables \( \{ \gamma_{i,jk} \} \) are also assumed to be cluster-specific to allow sufficient flexibility. Then, for any observed curve at \( s \in \mathcal{D} \), denoted as \( y_s^{(i)}(t) \), the Bayesian hierarchical model is,

\[
\begin{align*}
   y_s^{(i)}(t) | \bar{\beta}_i, \sigma^2_i & \sim N \left( \phi(t) \bar{\beta}_{i,00} + \sum_{j=1}^{J} \sum_{k=1}^{2^{j-1}} \phi_{jk}(t) \bar{\beta}_{i,jk}, \sigma^2_i \right), \\
   \bar{\beta}_i | \{ \gamma_{i,jk} \}, \{ \lambda_{i,j} \}, \sigma^2_i & \sim N(0, \sigma^2_i V_i), \\
   \gamma_{i,jk} | \pi_{i,j} & \sim \text{Ber}(\pi_{i,j}), \quad \sigma^2_i \sim \text{IG}(c_{i,j}, d_{i,j}), \quad \lambda_{i,j} \sim \text{IG}(c_{i,j}, d_{i,j}),
\end{align*}
\]

where \( V_i \) is a diagonal matrix of size \( m \), with \( \{ \gamma_{i,jk} \} \) as its diagonal entries.

3. Bayesian Inference of the Model Parameters.

3.1. Inference of the within-cluster parameters. We first discuss posterior inference of the within-cluster parameters, conditional on a partition of the location set \( S \). For notation simplicity, let \( \lambda_i \equiv (\lambda_{i,00}, \lambda_{i,11}, \ldots, \lambda_{i,J})^\prime \), \( \pi_i \equiv (\pi_{i,0}, \pi_{i,1}, \ldots, \pi_{i,J})^\prime \), and \( \gamma_i \equiv (\gamma_{i,00}, \gamma_{i,11}, \ldots, \gamma_{i,J2^{J-1}})^\prime \), for \( i = 1, \ldots, k_p \); recall that \( \gamma_{i,00} \) and \( \pi_{i,0} \) associated with the scaling coefficient \( \beta_{i,00} \) are always equal to one. The posterior samples of \( \{ \bar{\beta}_i, \sigma^2_i \} \) can be directly obtained from their closed-form full conditional distributions using Gibbs samplers. Specifically,

\[
p(\bar{\beta}_i | Y_i, \sigma^2_i, \lambda_i, \gamma_i) \sim N(\mu_{\bar{\beta}_i, |}, \Sigma_{\bar{\beta}_i, |}),
\]

(9)
where $\mu_{\beta_i} = (V_i^{-1} + n_i I_m)^{-1}X'_c Y_i$, and $\Sigma_{\beta_i} = \hat{\sigma}_i^2 (V_i^{-1} + n_i I_m)^{-1}$; here we use the fact that $X'_c X'_c = n_i I_m$, and recall that $V_i = \text{diag}\{\gamma_{i,jk}\}$. The posterior distribution, $p(\sigma_i^2 | Y_i, \lambda_i, \gamma_i)$, is obtained based on the marginalized likelihood, $p(Y_i | \sigma_i^2, \lambda_i, \gamma_i)$, which follows $N(0, \sigma_i^2 (X_c V_i X'_c + I_{mn}))$. Then, by imposing a conjugate prior, $p(\sigma_i^2 | a_i, b_i) \sim \text{IG}(a_i + mn_i, b_i + Y'_i (X_c V_i X'_c + I_{mn})^{-1} Y_i)$.

Since conditional on the latent indicator variables, $\{\gamma_{i,jk}\}$, the data model of $y(s)$ is a mixture of a multivariate normal distribution and a Dirac delta distribution at zero, the conditional posterior distributions of $\{\lambda_{i,j}\}$ are obtained by combining the data model and prior model in (8). In specific, for $\{\lambda_{i,j} : j \geq 1\}$ (i.e., the scaling parameters associated with the detail coefficients),

\begin{equation}
(11) \quad p(\lambda_{i,j} | \lambda_{i,-j}, Y_i, \bar{\beta}_i, \sigma_i^2, \gamma_i) \sim \text{IG} \left( \sum_{k=1}^{2^j-1} \gamma_{i,jk} + c_{i,j}, \sum_{k=1}^{2^j-1} \gamma_{i,jk}^2 + d_{i,j} \right),
\end{equation}

where $\lambda_{i,-j}$ denotes the vector of scale parameters $\lambda_i$ without $\lambda_{i,j}$. For $\lambda_{0,i,j}$, we similarly obtain that, $p(\lambda_{i,0} | \lambda_{i,-0}, Y_i, \bar{\beta}_i, \sigma_i^2, \gamma_i) \sim \text{IG}(1 + c_{i,0}, \gamma_{i,0}^2 + d_{i,0})$.

For the vector of hyper-parameters $\pi_i$, which basically controls the sparsity of the detail coefficients at each wavelet resolution for the $i$-th cluster, Ray and Mallick (2006) proposed to estimate them by maximizing a marginal log-likelihood; then, by fixing $\{\pi_i\}$ at their estimates, the remaining parameters are sampled from their posterior distributions, resulting in an empirical Bayes approach. Alternatively, since the information about $\pi_i$ is contained in the indicator variables $\gamma_i$, we can integrate out $\pi_i$ to obtain the marginal prior of $\gamma_i$ (e.g., Guan and Stephens, 2011; Zhou and Guan, 2019). Specifically, if we put a non-informative beta prior, $p(\pi_{i,j}) \sim \text{Beta}(1,1)$, then,

\begin{equation}
(12) \quad p(\gamma_i) = \int_{\pi} \prod_{j=1}^{J} \prod_{k=1}^{2^j-1} p(\gamma_{i,jk} | \pi_{i,j}) \times p(\pi_{i,j}) d\pi_i
\end{equation}

\begin{equation}
= \int_{\pi} \prod_{j=1}^{J} \prod_{k=1}^{2^j-1} (\pi_{i,j})^{\gamma_{i,jk}} (1 - \pi_{i,j})^{1-\gamma_{i,jk}} d\pi_i
\end{equation}

\begin{equation}
= \prod_{j=1}^{J} \frac{\Gamma \left( \sum_{k=1}^{2^j-1} \gamma_{i,jk} + 1 \right) \Gamma \left( 2^j - 1 + 1 - \sum_{k=1}^{2^j-1} \gamma_{i,jk} \right)}{\Gamma(2^j - 2)}. \Gamma(2^j - 2)
\end{equation}

After we obtain the posterior samples of $\{\gamma_{i,jk}\}$, the posterior samples of $\{\pi_{i,j}\}$ can be easily recovered from the following beta distribution,

\begin{equation}
(13) \quad p(\pi_{i,j} | Y_i, \gamma_i) \sim \text{Beta} \left( \sum_{k=1}^{2^j-1} \gamma_{i,jk} + 1, \sum_{k=1}^{2^j-1} (1 - \gamma_{i,jk}) + 1 \right).
\end{equation}

Last, to sample from the posterior distributions of the indicator variables $\{\gamma_{i,jk}\}$, we further integrate out $\sigma_i^2$ from $p(Y_i | \sigma_i^2, \lambda_i, \gamma_i)$, resulting in

\begin{equation}
(14) \quad p(Y_i | \lambda_i, \gamma_i) \propto \left( 1 + \frac{1}{b_i} Y'_i (X_c V_i X'_c + I_{mn})^{-1} Y_i \right)^{-\frac{a_i + mn_i}{2}},
\end{equation}
which follows a multivariate t distribution with mean 0, covariance matrix $\frac{b_n}{a_i}(X_iV_iX_i' + I_{mn,i})$, and degrees of freedom $a_i$. The posterior probability of $\gamma_{i,jk}$ can be written as an odds ratio,
\begin{equation}
 p(\gamma_{i,jk} = 1|\gamma_{i,-jk}, Y_i, \lambda_i) = \frac{O_{i,jk}}{1 + O_{i,jk}},
\end{equation}
where
\begin{equation}
 O_{i,jk} = \frac{p(Y_i|\gamma_{i,jk} = 1, \gamma_{i,-jk}, \lambda_i)p(\gamma_{i,jk} = 1, \gamma_{i,-jk})}{p(Y_i|\gamma_{i,jk} = 0, \gamma_{i,-jk}, \lambda_i)p(\gamma_{i,jk} = 0, \gamma_{i,-jk})},
\end{equation}
and $\gamma_{i,-jk}$ denotes the vector of the $\gamma$-parameters without $\gamma_{i,jk}$. Note that the prior probabilities, $p(\gamma_{i,jk} = 1, \gamma_{i,-jk})$ and $p(\gamma_{i,jk} = 0, \gamma_{i,-jk})$, are obtained from (12). The posterior samples of $\{\gamma_{i,jk}\}$ are then obtained from $\text{Ber}(O_{i,jk}/(1 + O_{i,jk}))$ using Gibbs samplers.

We remark that it is computationally expensive to evaluate the odds ratio and update each $\gamma_{i,jk}$ using standard methods when $m$ is large, as the marginalized log-likelihood function, $p(Y_i'|\sigma^2, \lambda_i, \gamma_i)$, involves evaluating the quadratic term, $Y_i'|X_iV_iX_i' + I_{mn,i})^{-1}Y_i$. Instead, we design a highly efficient iterative algorithm tailored to this specific model. The method takes advantage of the orthogonality of the wavelet basis, the low rank structure of $(X_iV_iX_i' + I_{mn,i})$ due to clustering, and updating the quadratic term by re-using the quantities from the previous iteration (see Supplementary Material, Section S1 for details).

3.2. Inference of the space-partition parameters. The partition parameters in $\theta_T$ involve the spanning tree $T$, the number of removed edges (minus one) $k_p$, and the partition of vertices in $T$ (or equivalently, the removed edge set in $T$). We update these parameters using the reversible jump Markov chain Monte Carlo (RJ-MCMC) algorithm in Green (1995), which allows transitions in the parameter space of $\theta_T$. Suppose we have completed $(r - 1)$ iterations and obtained $\{\theta_T^{(\ell)} : \ell = 1, \ldots, r - 1\}$; following the idea in Luo, Sang and Mallick (2021), at the $r$-th RJ-MCMC iteration, we consider four possible moves for updating $\theta_T$:

1. Birth move: Adding a new cluster to $\mathcal{P}^{(r-1)}$ based on the current spanning tree $T^{(r-1)}$ by randomly splitting an existing cluster, and hence $k_p^{(r)} = k_p^{(r-1)} + 1$. Splitting an existing cluster is achieved by randomly removing a within-cluster edge of $T^{(r-1)}$, which has a probability of $1/(n - k_p^{(r-1)})$;
2. Death move: Removing a cluster from $\mathcal{P}^{(r-1)}$ based on $T^{(r-1)}$ by randomly merging an existing cluster into an adjacent cluster, and hence $k_p^{(r)} = k_p^{(r-1)} - 1$. Removing a cluster is achieved by randomly adding a between-cluster edge of $T^{(r-1)}$, which has a probability of $1/(k_p^{(r-1)} - 1)$;
3. Change move: Performing a birth move first and then a death move so that the number of clusters stays unchanged, and hence $k_p^{(r)} = k_p^{(r-1)}$;
4. Hyper move: Updating the parameter $T$. Specifically, $T^{(r)}$ is obtained by randomly constructing a spanning tree that is compatible with $\mathcal{P}^{(r-1)}$ based on $G$.

The acceptance ratio of the proposed move involves the calculation of the likelihood ratio, an expensive step in standard RJ-MCMC algorithms. We use a collapsed marginal likelihood that integrates out local cluster-specific parameters to facilitate the mixing when sampling partitions. Also notice that only a subset of curves changes the cluster assignments during the birth, death, and change moves. As a result, the corresponding likelihood ratio only involves a small number of data. To obtain initial estimates of both the within-cluster parameters and the space-partition parameters, we borrow information across time for each curve to warm start both the edge weights of the underlying graph $G$ and the initial spanning tree $T$. More details on our RJ-MCMC algorithm are given in Supplementary Material, Section S1.
We remark that the labels of clusters may switch during RJ-MCMC, but the resulting partitions are unaffected by label switching. Moreover, the samplings of all other cluster-specific parameters and spanning trees only rely on partitions of locations and hence are unaffected by label switching either. We also remark that the partition of functions is the ultimate inference goal as opposed to the spanning trees and the removed edge sets that lead to partitions. Therefore, we only need to focus on summarizing the posterior results of the partition in the numerical studies in Sections 4 and 5. Alternatively, if someone is interested in summarizing the posterior results of functional curves at certain locations instead of for each cluster, we recommend using the functional median (Sun and Genton, 2011) or functional mode (Rousseeuw, Ruts and Tukey, 1999; Hyndman and Shang, 2010) based on the notion of functional depth for obtaining a single estimated curve. Details on summarizing the posterior estimates are given in Supplementary Material, Section S2.

3.3. Construction of $G$. The connected undirected graph $G$ is a hyperparameter of the RST-based clustering algorithm. This graph plays two very important roles in the method. First, it reflects the prior assumption on the contiguity constraint of clustering; the more edges connecting long-distance vertices, the weaker the spatial proximity constraint on clustering. Second, it determines the prior space of random spanning trees; a denser graph will result in a larger spanning tree space that needs to be explored in MCMC. In practice, when functional series are obtained from spatial locations and strong spatial contiguity is preferred, Euclidean distances of $S$ can be naturally used to construct $G$, either by the Delaunay triangulation (e.g., Lee and Schachter, 1980) or the k-nearest neighbors (K-NN) algorithm. In particular, one can select different $K$ values in the K-NN graph according to the strength of spatial contiguity assumption. Alternatively, in the scenario that covariate information is available and needs to be incorporated for clustering, it is desirable to use it or combine it with spatial information to define a weighted distance between the functional series, so that locations in the resulting clusters are “close” in the transformed covariate space.

Distance metric learning (DML, Shental et al., 2002; Xing et al., 2003; Suárez, García and Herrera, 2018) provides a useful way for obtaining the “optimal” weighted distance (Mahalanobis distances) between the covariates, defined as $\|s_j - s_\ell\|_M = \sqrt{(s_j - s_\ell)'M(s_j - s_\ell)}$, where $M$ is a positive semi-definite matrix. The goal is to amplify the differences of the covariates in dissimilarity sets and reduce their differences in similarity sets on the transformed scale. Finding the optimal weighted distance of the covariates can either be treated as a preprocessing step for a subsequent clustering or classification method, a post-analysis step to identify important covariates that explain some clustering results, or as an unknown parameter to be estimated jointly with clustering and other model parameters following the similar fashion as in Payne et al. (2020).

For our real data problem, for simplicity, we focused on illustrating the use of DML for the post-analysis of the spatial functional data clustering results based on a spatial Euclidean-distance based graph. We also considered using DML to reconstruct $G$ in the pre-processing step, such that the resulting functional clustering results reflect contiguity constraints for both socio-demographic variables and spatial information (Supplementary Material, Section S4). The method proceeds by obtaining some preliminary similarity and dissimilarity sets by using some fast, but often non-spatial and hence less accurate, functional data clustering methods, such as the functional principal component analysis (FPCA, Chiou and Li, 2007; Jacques and Preda, 2013). Then, based on the weighted distance, $\|s_i - s_j\|_M$, we can construct the graph $G$ for our proposed RST-based clustering method.

4. Simulation Studies. In this section, we use simulated datasets to test the performance of our proposed Bayesian functional clustering model based on random spanning
trees (we denote it as Fclust-RST). We consider three different scenarios for the true generated functions: (1) periodic functions with cluster-wise-constant varying magnitudes; (2) shifted Doppler functions with cluster-wise-constant varying location parameters; (3) shifted Doppler functions where the location parameters not only vary across clusters, but also have small variations within each cluster.

We consider a few competitive methods, including FClust (Jacques and Preda, 2013), kCFC (Chiou and Li, 2007), Mclust (Banfield and Raftery, 1993), and funHDDC-wavelet (Erwan, 2019). FClust performs the clustering using the distributions of functional principal component scores, while kCFC performs k-centers clustering based on the $L^2$-distance between the functional curves represented by the truncated Karhunen-Loève expansion; Mclust is designed for clustering the multivariate data and hence does not account for the functional form, which can serve as a baseline method; funHDDC-wavelet first transforms the data to the wavelet domain and then fits a Gaussian mixture to the wavelet coefficients. One shared limitation among these four methods is that they cannot automatically determine the number of clusters, and certain model-selection approaches are needed to determine $k_p$ (e.g., BIC, Schwarz et al., 1978).

For our proposed model, R packages “wavethresh” (Nason, 2016) and “igraph” (Csárdi, 2013) are used to perform the discrete wavelet transform and graph operations, respectively. We adopted the warm initialization strategy using initial parameter estimates from each curve to obtain the initial clusters; details on the warm-started initial values and the specifications of the hyperparameters are given in Supplementary Material, Section S1.3.

Here, we mainly focus on presenting the simulation results of the second scenario, where we generated data from the shifted Doppler functions (e.g., Abramovich, Sapatinas and Silverman, 1998; Ray and Mallick, 2006),

$$f_u(t; t_0) = -0.025 + 0.6\sqrt{t(1-t)} \sin(2.10\pi/(t - t_0)),$$

which has a location parameter $t_0$ taking values in $[0, 1]$. We uniformly generated $n = 300$ spatial locations in $D = (0, 1) \times (0, 1)$ and then created a connected undirected graph, $G$, based on the Delaunay triangulation of those locations. Each edge of $G$ was assigned with a random weight from Unif(0, 1), and a minimum spanning tree $T_0$ was obtained from $G$. We randomly deleted four edges of $T_0$ to create $k_p = 5$ true clusters (see the left panel of Figure 3). At spatial locations of the $i$-th cluster, we generated the true functional curves based on (16) at $m = 128$ equally spaced time points, where the location parameter $t_{0,i} = 0.1 + (i - 1) \times 0.2$, for $i = 1, 2, \ldots, 5$. For each cluster, the true signals were added with normal white noises whose variances are $\sigma_i^2 = 0.004 + (i - 1) \times 0.001$, for $i = 1, 2, \ldots, 5$.

We ran the MCMC algorithm for 20,000 iterations, with a burn-in period of 5,000 iterations and thinning by storing the sampling results every fifth iteration. In total, we collected 3,000 posterior samples. With the warm-started initial clusters, the MCMC chain converged quickly after a few hundred iterations. We checked the trace plots of the posterior log-likelihood values as well as the scalar variance parameter $\sigma^2(s)$ at a few randomly selected locations, and these MCMC chains were mixed well. The right panel in Figure 3 shows the point estimate of the MST and its induced clusters by Dahl’s method (see Supplementary Material, Section S2 for details). We can see that the inferred clusters exactly match the true ones. The corresponding predictive curves by Dahl’s method are shown in Figure 4, which capture the oscillatory patterns of the true signals very well. Last, we checked the posterior estimates of the variance parameters, $\{\sigma_i^2\}$, and obtained the 95% credible intervals of $\{\sigma^2(s)\}$ at five randomly selected locations as shown in Figure 5. These credible intervals contain the true values of $\{\sigma^2(s)\}$, indicating that our proposed method can quantify the uncertainties of the functional curves properly.
FIG 3. Left panel: The true minimum spanning tree (with four edges deleted) and the induced five clusters; Right panel: The minimum spanning tree and the inferred clusters of Fclust-RST using Dahl’s method.

FIG 4. The observed shifted Doppler curves (in grey) versus the posterior predictive curves using Dahl’s method (in red) for each cluster, where $t_0$ is constant within a cluster.

FIG 5. The 95% credible intervals (indicated by the horizontal bars) of $\{\sigma^2(s)\}$ at five randomly selected locations for the shifted Doppler example with cluster-wise constant location parameters. The solid circles denote the posterior medians of $\{\sigma^2(s)\}$ and the triangles (in red) denote the corresponding true values.

We calculated the adjusted Rand index (Hubert and Arabie, 1985) of the clustering results for each data simulation scenario, which measures the discrepancy between two classifications (see Table 1). A greater value of the adjusted Rand index indicates a better match with the true partition, and a perfect match leads to the maximum value of one. For this simulation example, the competitive methods, FClust and funHDDC-wavelet, achieve the maximum
value of one, which is greater than 0.720 by kCFC and 0.009 by Mclust. We remark that the good clustering performance of the FClust and funHDDC-wavelet is achieved under the condition that the true number of clusters is given. In practice, we need to pre-specify the number of clusters, $k_p$, or try a few values of $k_p$ for both methods and select the best one using some model-selection criteria such as BIC. This can be impractical when a large number of clusters could exist (e.g., the LHI data considered in Section 5). Furthermore, although not shown here, when $k_p$ is slightly mis-specified to be larger than the true number of clusters, the resulting clusters by FClust and funHDDC-wavelet are discontiguous in space, which may hurt the interpretability of the clusters. In contrast, the inference of the number of clusters is a natural product of our Bayesian method.

Additional simulation results and details can be found in Supplementary Material, Section S3. In particular, for the shifted Doppler examples with cluster-wise-constant/slightly-varying location parameters, we conducted more comprehensive comparisons to test the scenarios where the functional curves have a) lower signal-to-noise ratios and b) unbalanced clusters where some clusters have a very small size. We observe that our proposed method is generally preferred over other methods when the true clusters have unbalanced sizes.

### 5. Analysis of the Leaving-Home Index Dataset.

#### 5.1. Data description.
In this section, we analyze the SafeGraph leaving-home index (LHI) dataset, where leaving-home indices are the daily percentages of people who left home. The LHI data are obtained on the CBG level for personal privacy. We obtained the raw data from the SafeGraph COVID-19 Data Consortium (https://www.safegraph.com/covid-19-data-consortium) and processed the raw data to obtain the CBG level LHI data in Harris County, Texas, United States. There are 2144 CBGs in total, and we chose the data from 2138 CBGs of no missing values. The time range is from February 17, 2020 to June 23, 2020, including 128 time points. This time period includes dates of important policy interventions on social distancing in Texas. Figure 1 shows the time-averaged LHI series at each CBG and clearly, there are some strong spatial patterns of them, with higher values in the eastern region and the northwestern corner and smaller values in the southwestern part. In the meanwhile, there are a few census blocks with substantially higher or lower values than their neighboring blocks. Hence, a good clustering model is desirable to identify big clusters of reasonable homogeneity and to distinguish small clusters that behave differently from their neighbors.

Since the LHI data are the proportions taking values in $(0, 1)$, we took the conventional logit transform of them before model fitting. We also included a number of demographic variables at those CBGs in the analysis, including ethnic compositions, age compositions, median household incomes, population densities, proportions of people who are living alone, and proportions of people with a Bachelor’s degree (which measure the average education levels). These demographic variables are potential factors that could affect the LHI values, since they are related to people’s work and life styles, and hence may help to interpret the clustering results. Figure 6 shows the proportions of different ethnic groups in the CBGs, where the ethnic groups are: African American (AA), Asian (AS), Hispanic (HI), and White

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Fclust-MST</th>
<th>FClust</th>
<th>kCFC</th>
<th>Mclust</th>
<th>funHDDC-wavelet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Periodic functions</td>
<td>0.999</td>
<td>1.000</td>
<td>0.591</td>
<td>0.740</td>
<td>1.000</td>
</tr>
<tr>
<td>Doppler signals, constant $t_0$</td>
<td>0.999</td>
<td>1.000</td>
<td>0.720</td>
<td>0.009</td>
<td>1.000</td>
</tr>
<tr>
<td>Doppler signals, varying $t_0$</td>
<td>0.999</td>
<td>1.000</td>
<td>0.738</td>
<td>0.175</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Fig 6. The spatial maps of different ethnic groups of Harris county, which are African American (AA), Asian (AS), Hispanic (HI), and White (WH).

Fig 7. The spatial maps of the averaged median household income (upper-left panel, in the units of thousand dollars), the population density (upper-right panel), the proportion of people who are living alone (lower-left panel), and the proportion of people who have a Bachelor’s degree (lower-right panel).

We observe that different ethnic groups are populated in different regions. For example, the regions where Hispanic people have a large proportion seem to be complementary to those of White people. A figure showing dominating ethnic groups for each CBG is given in Section S4 of Supplementary Material, from which we see more clearly the spatial clustering patterns of different ethnic groups.

The age-composition information is also available, where the age groups are defined to be “Elderly” (people of age 65 or above), “Middle-Aged” (people of age between 30 and 64 inclusive), “Young” (people of age between 18 and 29 inclusive), and “Juvenile” (people of age less than 18). The proportions of different age groups at each CBG are shown in Figure S26 in Supplementary Material. The spatial maps of the remaining demographic variables are shown in Figure 7, including the averaged median household income, the population density, the proportion of people who are living alone, and the proportion of people who have a Bachelor’s degree. The income variable is positively correlated with the education variable, with a correlation coefficient equal to 0.73; it is also correlated with the ethnic-composition variables. The population densities do not vary much across different census blocks, and so
do the proportions of people who are living alone, except for one census block which has a very high value of one.

5.2. Clustering results. We applied our proposed random-spanning-tree based clustering model (Fclust-RST) to cluster the LHI series at 2138 CBGs. The hyperparameter $\mathcal{G}$ was obtained through Delaunay triangulation based on the centroids of the census blocks. We ran the MCMC chains with different specifications of $c$, and the trace plots of log of posterior likelihood values as well as estimates of the scalar parameters (such as $\sigma^2(s)$) were checked to ensure the convergences of the MCMC chains. We found that the point estimates of the clustering by Dahl’s method are similar for different $c$ values. Then, we chose the clustering results of Fclust-RST with $c = 0.30$ for illustration according to the widely applicable information criterion (WAIC, e.g., Watanabe and Opper, 2010; Vehtari, Gelman and Gabry, 2017). More details of the sensitivity studies are given in Supplementary Material, Section S4.

The point estimate of the clustering by Dahl’s method identifies 85 clusters, with 25 bigger clusters of no less than ten census blocks and 60 smaller clusters of less than ten census blocks (see Figure 8). Based on the clustering results, the ethnic-composition variables and the household income/education variables appear to have a large-scale impact on the clustering patterns of the LHI series, which drives the creation of bigger clusters. For example, Clusters 1, 3, 13, 15, 82, and 83 mainly include the CBGs that are located on the north, east, and southeast of downtown Houston. These high-poverty areas have a large percentage of the ethnic groups AA and HI (the overall median value is about 90%), as well as a very low percentage of people with a Bachelor’s degree (the median value is about 6%). In
contrast, Clusters 8 and 85 contain Houston’s wealthiest neighborhood within Interstate 610 inner loop, including River Oaks, West University/Medical Center near Rice University, and Bellaire. These two clusters are most populated by the ethnic group WI (the median value is over 66%), which have both a high percentage of people with a Bachelor’s degree (the median value is over 60%) and a high median household income (the median value is over $110,000). Our clustering results successfully distinguish between two spatially neighboring Clusters 62 and 85. Despite their spatial proximity, the median household income in Cluster 62 is approximately $39,000, compared with $100,000 of Cluster 85; Cluster 62 also has a much lower percentage of people with a Bachelor’s degree (which is around 16%), compared with 64% of Cluster 85. Clusters 26 and 53 contain a number of suburbs including Cypress, Katy and Tomball along State Highway 99 (the Grand Parkway) northwest loop. The residents in these fast growing Houston suburbs are mainly working professionals with higher education and middle-class families with kids; the dominating race is White but overall, these two clusters have a more diverse racial make-up. Clusters 54 and 18 contain other middle-class suburbs, including Humble, Atascocita, and Kingwood, which are close to Houston’s international airport in northeast Houston.

The corresponding predictive curves by Dahl’s method are shown in Figure S30 in Supplementary Material, which match the observed LHI series in each cluster very well. The predicted LHI series generally decrease a few days before the date of stay-at-home order (March, 24th) and gradually increase after Phase I re-opening date (April, 30th), indicating that the policy interventions on social distancing are generally effective. The main differences of those predictive series lie in the starting values in early March and the sharpness of de-
increase before the social distancing order. Since our method aims to detect spatially contiguous clusters, we checked the predictive curves of the spatially neighboring clusters in different regions to examine its clustering capacity. We found that the predictive LHI curves of the spatially neighboring clusters generally behave very differently, indicating that the clustering results of Fclust-RST are reasonable. For example, Figure 9 shows the pairwise comparisons of the predictive curves for Clusters 8, 22, 62, 84, and 85 in the central-west region. We can see that these predictive LHI curves behave differently in certain time periods. Compared to the lower-income and lower-education clusters (Clusters 62 and 84), the higher-income and higher-education clusters (Clusters 8, 22, and 85) have larger LHI values both before the social distancing order and after the Phase I re-opening order; the decreases of the LHI values before enforcing the social distancing order are also sharper. Similar comparison results in other regions are provided in Supplementary Material, Section S4.

Figure 10 shows the predictive LHI curves of a few smaller clusters (of less than ten blocks). We can see that the LHI curves of smaller clusters can behave very differently from those of bigger clusters, many of which did not decrease much during the time period of closing and re-opening social activities (e.g., Clusters 4, 27, 32, 48, 57); the LHI curves in some smaller clusters even increased after the date of closing social events (e.g., Clusters 9, 16, 35). Proximity to a walking friendly built environment such as parks and trails may explain the maintained level of mobile activity in some of the clusters. For example, Clusters 16, 35, 37, and 57 are within walking-distance to Alief Community Park, Moody Park, Edgewood Park, and George Bush Park, respectively. The age compositions may also be important factors that lead to these behaviors, and we found that many of smaller clusters have a large proportion of Juvenile and Young people (such as Clusters 9, 27, 32). In fact, Cluster 9 is a neighborhood near Rice University with many college students. The increasing pattern in April and decreasing pattern after June reflected the changes in students’ primary home addresses from University apartments to their families’ homes, and the reduced business activities due to summer break.

Further inspection reveals that our clustering method successfully identifies census blocks in which specific facilities are located. For example, Cluster 4 is the place where Exxon-Mobile Chemical locates, which contains many plants on petroleum refinery and processing, while Cluster 65 contains industrial areas of many chemical/energy companies such as Chevron Phillips Chemical, Afton Chemical, Ethyl Corporation, and Calpine Corporation.
TABLE 2
The weights of covariates produced by DML, where the similarity and dissimilarity sets are obtained from the clustering results of our proposed model.

<table>
<thead>
<tr>
<th>Clusters, nk ≥ 10</th>
<th>Elderly</th>
<th>Juvenile</th>
<th>Young</th>
<th>AA</th>
<th>AS</th>
<th>HH</th>
<th>WI</th>
<th>Income</th>
<th>Live.alone</th>
<th>Pop.dens</th>
<th>Edu</th>
<th>Lon</th>
<th>Lat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.000</td>
<td>0.111</td>
<td>0.000</td>
<td>1.232</td>
<td>1.896</td>
<td>1.243</td>
<td>1.184</td>
<td>0.000</td>
<td>0.092</td>
<td>0.882</td>
<td>1.640</td>
<td>1.346</td>
<td>1.239</td>
</tr>
<tr>
<td>Clusters, nk &lt; 10</td>
<td>0.000</td>
<td>0.008</td>
<td>0.000</td>
<td>1.169</td>
<td>0.936</td>
<td>1.453</td>
<td>1.004</td>
<td>0.004</td>
<td>0.452</td>
<td>0.845</td>
<td>1.860</td>
<td>1.399</td>
<td>1.312</td>
</tr>
<tr>
<td>All clusters</td>
<td>0.000</td>
<td>0.135</td>
<td>0.018</td>
<td>1.222</td>
<td>1.116</td>
<td>1.241</td>
<td>1.154</td>
<td>0.000</td>
<td>0.070</td>
<td>1.640</td>
<td>1.359</td>
<td>1.248</td>
<td></td>
</tr>
</tbody>
</table>

The high LHI values of these two clusters are likely due to the fact that these companies were still operating after the social distancing order. Moreover, Cluster 32 is the area of George Bush Intercontinental Airport and Cluster 27 contains William P. Hobby Airport; the non-stop operations of airports likely lead to the flat curves of the LHI series during the pandemic.

We also applied FClust to clustering the LHI series, with the number of clusters specified to be a comparable value of 50 (see Figure S37 in Supplementary Material). However, the resulting clusters are not spatially contiguous, rendering difficulties in interpreting them.

5.3. Post-analysis of clustering results by DML. To further evaluate the impact of covariates on the spatial clustering results, we resort to the distance-metric-learning (DML) method in Xing et al. (2003) as a post-analysis procedure (also see Section 3.3). The covariates considered in DML are described in Section 5.1; longitudes (“Lon”) and latitudes (“Lat”) of centroids of census blocks are also added as covariates. These covariates are standardized by subtracting their respective means and dividing by their respective standard deviations. DML provides a weighted distance metric of covariates that best explains the similarities of the data. Our clustering results are used to define the similarity and dissimilarity conditions of DML. We constrain the weighting matrix $M$ to be a diagonal matrix, so that the square roots of its diagonal entries are the weights to be imposed on the Euclidean distances of covariates. Hence, the magnitudes of weights indicate the relative importance of covariates in measuring the similarity/dissimilarity conditions provided by the clustering results. Table 2 shows the weights of these 13 covariates for both the bigger-clusters case (i.e., $nk ≥ 10$) and the smaller-clusters case (i.e., $nk < 10$). For both cases, the weights of the ethnic compositions have a large magnitude, indicating their importance in measuring the similarities of the data, which coincides our observation that the ethnic compositions play an important role in forming clusters. The education variable also has a large weight, probably due to its positive correlation with the income variable that impacts people’s life and work styles. The large weights of spatial coordinates, Lat and Lon, indicate the need to include spatial information in the clustering to account for the remaining spatial variabilities unexplained by the socio-demographic variables. The effects of the remaining variables are relatively weak. When using all clusters to define the similarity/dissimilarity conditions, the age-composition variables have small nonzero weights, likely due to the effects of some clusters with relatively higher proportions of Juvenile and Young people (e.g., Clusters 9, 27, 32).

6. Conclusion. In summary, we have proposed a Bayesian wavelet model for clustering spatial functional data on a graph, where the space partitioning is achieved by operating on the random spanning trees (RSTs) constructed from data locations. This RST-based clustering approach enjoys the merit of ensuring the contiguity of clusters in the input space. Furthermore, there is no restriction on the shape of clusters due to the nature of random spanning trees in representing graph structures compactly, which provides more flexibilities in capturing the true homogeneous clusters of arbitrary shapes in real applications. Since we assume that the contiguity constraint is defined with respect to the choice of a spatially connected graph, the current method cannot characterize both globally dis-contiguous and locally contiguous clusters. When two or more locally contiguous clusters that are remote in space share very similar functional patterns, our method can only identify them as separate clusters. One
idea to address this limitation is to perform a post-non-spatial clustering analysis of local clusters obtained from our method.

We applied our proposed model to the LHI dataset in Harris County, Texas, in the goal of modeling people’s mobility behaviors during the COVID-19 pandemic. By comparing the clustering results with the available socio-demographic information, we found that the ethnic compositions and income/education levels of census blocks have a global impact on the behaviors of the LHI series, which mainly drives the creation of big clusters. We notice that higher-income and higher-education clusters in general tend to have a sharper decrease in LHI before lock-down period compared to those of lower income and lower-education clusters. This agrees with the impact of income on social distancing responses investigated in Weill et al. (2020) and Jay et al. (2020). Since people in the lower-income neighborhoods are less likely to work from home, it is desirable for the state government to make policies to enable physical distancing of lower-income residents, such as making unemployment insurance available, restricting evictions, and banning essential utility shut-offs during the pandemic (Jay et al., 2020). In addition, since the behaviors of responses to social distancing are spatially heterogeneous, instead of implementing one-size-fits-all closing/reopening policies for all of the regions within a county, policies should account for the diverse nature of different regions. Our model allows us to explore spatially heterogeneous effects of public policy interventions on mobility patterns and identify areas that most benefit from government intervention. This can help policy makers to determine where these interventions are most needed for current and future pandemic preparedness responses.

Although we focus on the leaving home index variable of the SafeGraph data, the proposed method is general and can be applied to other (longitudinal) COVID-19 datasets. For example, the median distance travelled from home data provided by SafeGraph reflect the mobility behavior of residents, which can be readily analyzed by our method. Our method can also be applied to similar mobility datasets from Google Mobility Report (https://www.google.com/covid19/mobility/) and PlaceIQ (Couture et al., 2020).

**Acknowledgements.** Zhang’s research is supported by National Natural Science Foundation China Project, number 11901316, and the Fundamental Research Funds for the Central Universities, Nankai University (63201156), China. It is also partially supported by the Key Laboratory of Pure Mathematics and Combinatorics (LPMC) of Ministry of Education, China and the Key Laboratory for Medical Data Analysis and Statistical Research (KLDASR) of Tianjin, China. The authors thank SafeGraph, Inc. for permitting access to their datasets and Dr. Cici Bauer for guiding us to SafeGraph data. The authors also thank Dr. Ying Sun for valuable comments and suggestions on functional posterior summaries.

**SUPPLEMENTARY MATERIAL**

**supp-Fclust-RST (PDF file)**

This supplementary material contains implementation details of the MCMC algorithm, and additional results of simulation studies and the real-data analysis.

**REFERENCES**


