

Published in final edited form as:

Technometrics. 2012 May 1; 54(2): 129–133. doi:10.1080/00401706.2011.649820.

Discussion of the paper “Clustering Random Curves Under Spatial Interdependence with Application to Service Accessibility” by Jiang and Serban

Jiaping Wang^a, Haipeng Shen^b, and Hongtu Zhu^{a,c}

^aDepartment of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

^bStatistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

^cBiomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

We first congratulate Drs. Jiang and Serban on their interesting and important contribution to the literature of clustering functional data. The paper provides a comprehensive framework of clustering *spatially dependent* curves, namely the *functional-spatial clustering models* (FSCM), by nicely integrating important techniques, including the EM algorithm, Monte-Carlo approximation, dimension reduction, and model selection. The authors relaxed a common independence assumption, that is, curves are assumed to be independent in space. The authors demonstrate that when clustering a collection of curves distributed over space, releasing such an independence assumption represents a big improvement over existing clustering methods for functional data in the literature. We appreciate the opportunity to comment on several aspects of this nice work.

The motivating application for the authors is clustering service accessibility within a geographically distributed service network. Our discussions, however, are motivated by the analysis of massive functional imaging data, such as functional magnetic resonance imaging (fMRI), which are commonly observed over both time and space. There are some common features between the two contexts, such as spatial-temporal dependent functions, but significant structural differences still exist. Such differences form the basis for our specific comments reported in Sections 2 to 4, regarding (i) spatial dependence and neighborhood, (ii) spatial and temporal smoothing, and (iii) computational efficiency.

We first give some brief background on fMRI data. fMRI experiments record images of brain activities over time. An fMRI dataset is four-dimensional, consisting of a three-dimensional spatial image being observed over time. Each 3D fMRI image contains a certain number of two-dimensional slices, and each slice is made up of individual cuboid elements called *voxels*. The time series at each voxel can be viewed as a temporal function, which is distributed spatially across the voxels on the brain. For example, Figure 1(a) shows a particular 2D fMRI image slice, where the red and blue regions are either activated or deactivated by certain experimental stimuli; the goals of fMRI analysis are usually to detect the activation/deactivation regions. More details on fMRI data and related statistical

techniques can be found in, for example, Yue *et al.* (2009), Wang *et al.* (2011), and Lee *et al.* (2011).

1 Spatial Dependence and Neighborhood

Many existing functional clustering methods ignore dependence in spatial data. Specifically, as is often the case in many studies of spatial data, one may observe spatially contiguous regions with similar effects, as seen in the red and blue areas of Figure 1(a). From this point of view, the authors made a big contribution by using neighboring information at each pixel (or voxel) to account for spatial dependence. A challenging issue of using neighboring information is to determine an “appropriate” neighborhood size at each pixel (or voxel). For example, as shown in Figure 1(b), the pixels A and B belong to two different homogeneous regions or clusters, which are separated by sharp and irregular boundaries.

FSCM uses the K-nearest neighbors (KNN) method to incorporate the neighborhood information. However, this nearest neighbor component of FSCM will easily blur the boundary between jumping surfaces, a drawback common to several spatial smoothing methods in the current literature as discussed in Yue *et al.* (2010), including the standard Gaussian spatial smoothing (i.e., kernel smoothing with a Gaussian kernel). An example is provided in our simulation study to be reported in Section 5, and illustrated in Figure 2(d), where we applied the standard Gaussian spatial smoothing to the whole region and the results showed the *blurred* boundary. KNN has a similar boundary-blurring problem as the Gaussian smoothing, which leads to the unsatisfactory clustering results depicted in Figure 2(e).

More intuitively, we consider the pixel (or voxel) A in Figure 1(b) and a “+” shaped local neighborhood centered around it, which includes four nearest neighbors: A1, A2, A3, and A4. Suppose the four nearest pixels in this neighborhood structure are correctly clustered, that is, three of the four pixels are in the yellow cluster and the other one is in the blue cluster. Then the probability that the pixel A belongs to the yellow cluster, based on the Gibbs distribution in FSCM, will be larger than the probability that it belongs to the blue cluster. As a result, it can be easily clustered into the wrong region as there are more yellow pixels around it, although it is in the blue region. The same problem will occur for the pixel B; furthermore, many other symmetric neighborhood structures, such as circles or squares, would have similar problems. Furthermore, the probability mass function for the Gibbs distribution will not be consistent. If the neighborhood size, denoted as K by the authors, becomes larger, then the possibility of clustering A into the wrong region also becomes larger. Hence, the choice of the neighborhood size K can be a critical issue as a large value of K makes the boundaries between the distinct classes less distinct.

The above example suggests that we should be careful to choose K along with the right neighborhood structure. It is also critical to choose an appropriate distance measure for clustering as shown in this example: the Euclidian distance in KNN can’t determine an appropriate neighborhood structure. Instead, suppose we can define some alternative distance measure that can remove the yellow pixels from the neighborhood structure around A, then KNN can cluster A into the right region based on the remaining pixels in the

neighborhood. As one example, Hall *et al.* (2010) discussed how to choose K and its neighborhood structure by first assuming some local distribution (Poisson or Binomial), and then determining the distance measure based on the distribution. However, they did not consider the spatial-temporal nature of functional imaging data. According to our extensive experience, the neighborhood size K should depend on the spatial location of each pixel. That is, K should vary across pixels or voxels, and thus it should be determined spatially across the whole region. See Li *et al.* (2011) for more discussions.

2 Spatial and Temporal Smoothing

As most existing works in the literature, FSCM uses a B-spline or P-spline smoothing method in the temporal domain to smooth the random curves. The suitable models are selected by either the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). Spatial smoothing structure is also important when the authors assume spatial dependence. The authors smooth the random curves by involving the neighborhood random curves, which may increase the accuracy of the estimation and subsequently the clustering performance.

One also has to be careful with the model selection procedure for the following two reasons. First, Bayes factors depend on prior beliefs about the expected distribution of the parameter values, and there is no guarantee that the Bayes factor implied by BIC will be close to the one calculated from a prior distribution that an observer would actually regard as appropriate. Secondly, to obtain the Bayes factors that follow from BIC, investigators would have to vary their prior distributions according to the marginal distributions of the variables and the nature of the hypothesis. One potential consequence of this is that some of the random curves can be smoothed inappropriately. For example, Figure 2(b) plots the true temporal functions used in the three spatial regions of our simulation study (Section 5). As one can see, the differences among these three curves are very small, and there are also some jumping signals. The B-spline or P-spline with AIC/BIC can easily ignore their differences due to the wrong model selection, which then increases both false positive and negative rates. One alternative approach to resolve this problem is to simultaneously incorporate spatial and temporal smoothing, and then use some adaptive procedure to determine the related parameters, as proposed in Li *et al.* (2011).

3 Computational Efficiency

The authors explicitly model spatial dependence among spatially connected locations via Markov random fields (MRF) (Besag, 1986). However, as discussed in Zhu *et al.* (2007), calculating the normalizing factor of a MRF and estimating spatial correlation for a moderate number of spatial locations in a 2D space or 3D volume can be very computationally intensive. Moreover, it can be restrictive to assume a specific type of correlation structure across the whole 2D space or 3D volume.

For example, a typical fMRI dataset has tens of thousands of voxels, i.e. spatial locations, and hundreds of time points. Meanwhile, the assumption of the same correlation structure across the whole brain is very unrealistic, as correlation structures vary across the whole brain volume due to its physiology features. The large number of time points also

dramatically increase computational time in the dimension reduction procedure. Eventually, these problems will make it very challenging to directly apply the Monte-Carlo approximation implemented in FSCM to fMRI data.

Alternatively, one may consider some frequency domain approach to use some transformation, such as the Fourier transformation or the wavelet transformation, to transform the time series into the frequency domain. Such transformation can reduce the temporal correlation, and decrease the associated computation efforts, see the example in Wang *et al.* (2011).

4 A Simulation Example

We illustrate the points discussed above through a simulation example. Specifically, we compare the authors' method with ours that is currently under development.

Our method consists of two primary steps: (i) smoothing and (ii) clustering. The smoothing step implements a multiscale adaptive procedure to denoise the random curves adaptively and hierarchically, by adaptively creating a sequence of nested ellipsoids at each spatial location to capture the location-specific spatial dependence within its neighborhood. This smoothing technique can capture the functional segregation and integration of different spatial regions and prepare for the follow-up clustering. In the clustering step, we implement an EM algorithm on the wavelet transformations of the smoothed random curves. Note that the wavelet transform of stochastic time series is asymptotically Gaussian. The number of the clusters is automatically determined using a technique similar to that of Chen and Khalili (2009).

We consider toy-example fMRI simulation studies where the spatial domain only consists of a single slice with 32×32 voxels, and the time domain includes 128 equally spaced time points in $[0, 1]$. The true activation image is composed of three activated regions, where each region consists of 4 contiguous circles, and each circle is of diameter 8 pixels, as indicated using different colors in Figure 2(a). Denote the three activation regions from left to right as R_1 , R_2 and R_3 , respectively, and the inactive region as R_4 .

To simulate the toy fMRI data, we consider the following generative model for pixel d and time t :

$$Y_j(d, t) = S_j(d, t) + Z_j(d, t), \quad d \in R_j, \quad j = 1, \dots, 4, \quad (1)$$

where $S_j(d, t)$ is the true temporal signal at pixel d , and $Z_j(d, t) \sim N(0, \sigma_j^2)$ is the corresponding error, both of which depend on the region that pixel d belongs to, as described in more details below. Within the activated region R_j , $j = 1, 2, 3$, the true temporal signal is set to be the difference between two exponential functions, as in the following expression,

$$S_j(d, t) = 1 + [\exp(-|t - u_{j1}|/T_{j1}) - \exp(-|t - u_{j2}|/T_{j2})],$$

where

$$(u_{11}, T_{11}) = (0.4, 0.008), (u_{21}, T_{21}) = (0.41, 0.01), (u_{31}, T_{31}) = (0.405, 0.01), (u_{12}, T_{12}) = (0.41, 0.08), (u_{22}, T_{22}) = (0.42, 0.01), (u_{32}, T_{32}) = (0.415, 0.01).$$

The three functions are plotted in Figure 2(b). As one can see, the functions are very similar with each other, and have sharp drops, which actually will cause problems for FSCM as we show below. For the inactive region, we set the temporal signal to be $S_4(d, t) = 0$.

In addition to the region-specific temporal mean signal, we also set up the error variances σ_j^2 to depend on the region so that each region has a different signal to noise ratio (SNR). More specifically, for each j , we consider any pixel d within R_j , treat the $S_j(d, t)$ across time as the data, and calculate their sample variance, denoted as Var_j . Calculation shows that $\text{Var}_1 = 0.0066$, $\text{Var}_2 = 0.0056$, $\text{Var}_3 = 0.0045$, and $\text{Var}_4 = 0$. We then set $\sigma^2 = \max\{\text{Var}_j; j = 1, 2, 3, 4\} = 0.0066$ and $\sigma_j^2 = \sigma^2 / \text{SNR}_j$, where $\text{SNR}_1 = 0.7$, $\text{SNR}_2 = 0.8$, $\text{SNR}_3 = 0.7$ and $\text{SNR}_4 = 0.8$. The above setup suggests that the SNRs in the three activation regions are approximately 0.7, 0.68, and 0.48 respectively, which are all very small.

We applied FSCM to our simulation by setting the number of clusters as 4 and $K=5$, i.e. KNN uses five nearest neighbors. We repeated the simulations several times, and a typical clustering result is provided in Figure 2(e). The corresponding clustering rates (number of the locations in i th cluster/total number of locations) are 0.23, 0.20, 0.29, and 0.28 for the four clusters, respectively. The true clustering rates are, respectively, 0.1836, 0.1836, 0.1836, and 0.4492 for the different regions. It is clear that FSCM does not work well in this simulation study. A typical result from our method is given in Figure 2(f). We can see that only a few locations are mis-clustered.

We speculate that the above simulation setup is challenging for FSCM due to the rather small differences among the temporal signal functions and the sharp drops in their functional forms, shown in Figure 2(b), as well as the rather small signal-to-noise ratios. To empirically confirm our speculations, we considered another simulation setup with the same spatial activation map, but the three temporal signal functions are very different from each other and very smooth. Specifically, we assume that

$$S_1(d, t) = e^t, \quad S_2(d, t) = t^2, \quad S_3(d, t) = t^3,$$

where $t \in \{-1, -15/16, \dots, 0, \dots, 15/16, 1\}$, and $S_4(d, t) = 0$. We then simulate the fMRI data according to the generative model (1) with $Z_j(d, t) \sim N(0, 0.1)$. See Figure 3(a) for plots of the true temporal signals $S_j(d, t)$, $j = 1, 2, 3$. In this case, the temporal signal functions in the four regions, especially in the three active ones, differ significantly from each other, and the SNRs are larger: 1.81 in R_1 , 1.05 in R_2 , and 0.70 in R_3 . Again, we set the number of clusters equal to 4 and $K = 5$, and apply FSCM to obtain the results shown in Figure 3(b), which indicate that FSCM works very well in this case. The results for our method are not shown here as they are very similar to those obtained by FSCM.

5 Concluding Remarks

Jiang and Serban have developed a nice framework to cluster *spatially dependent* random curves. However, the issues of spatial boundary, smoothing, and computational efficiency suggest that the problem under investigation is challenging and deserves extra care. Specifically, the neighborhood size should be carefully chosen; spatial smoothing should be incorporated to increase the accuracy of the estimation by borrowing strength within a neighborhood; finally, when the spatial and temporal sampling rates are high, computation becomes very intensive for the MRF prior. Clustering of spatial-temporal functional data with dependence is an interesting research problem that needs further investigations.

Acknowledgments

The authors want to thank the editor for his careful review and insightful comments which have greatly improved the quality of the paper. Wang and Zhu's work was supported in part by NIH grants UL1-RR025747-01, P01CA142538-01, MH086633, and AG033387. Shen's work was supported in part by NIH grant 1RC1DA029425-01 and NSF grants CMMI-0800575, DMS-1106912.

References

- Besag J. On the Statistical Analysis of Dirty Pictures (with discussions). *Journal of the Royal Statistical Society, Series B.* 1986; 48:259–302.
- Chen J, Khalili A. Order Selection in Finite Mixture Models with a Non-smooth Penalty. *Journal of the American Statistical Association.* 2009; 104:187–196.
- Hall P, Park BU, Samworth RJ. Choice of Neighbor Order in Nearest-Neighbor Classification. *The Annals of Statistics.* 2008; 36:2135–2152.
- Li Y, Zhu H, Shen D, Lin W, Gilmore JH, Ibrahim J. MARM: Multiscale Adaptive Regression Models for Neuroimaging Data. *Journal of the Royal Statistical Society, Series B.* 2011; 78:559–578.
- Lee S, Shen H, Truong Y, Lewis M, Huang X. Independent Component Analysis Involving Auto-correlated Sources with an Application to Functional Magnetic Resonance Imaging. *Journal of the American Statistical Association.* 2011; 106:1009–1024.
- Wand, MP.; Jones, MC. *Kernel Smoothing.* Chapman and Hall; London: 1995.
- Wang, J.; Zhu, H.; Fan, J.; Giovanello, K.; Lin, W. Adaptively and Spatially Estimating the Hemodynamic Response Functions in fMRI. *Medical Image Computing and Computer Assisted Intervention (MICCAI) Conference; LNCS; 2011.* p. 269-276.
- Yue Y, Loh JM, Lindquist MA. Adaptive Spatial Smoothing of fMRI Images. *Statistics and Its Interface.* 2010; 3:1–11.
- Zhu HT, Gu MG, Peterson BG. Maximum Likelihood from Spatial Random Effects Models via the Stochastic Approximation Expectation Maximization Algorithm. *Statistics and Computing.* 2007; 15:163–177.

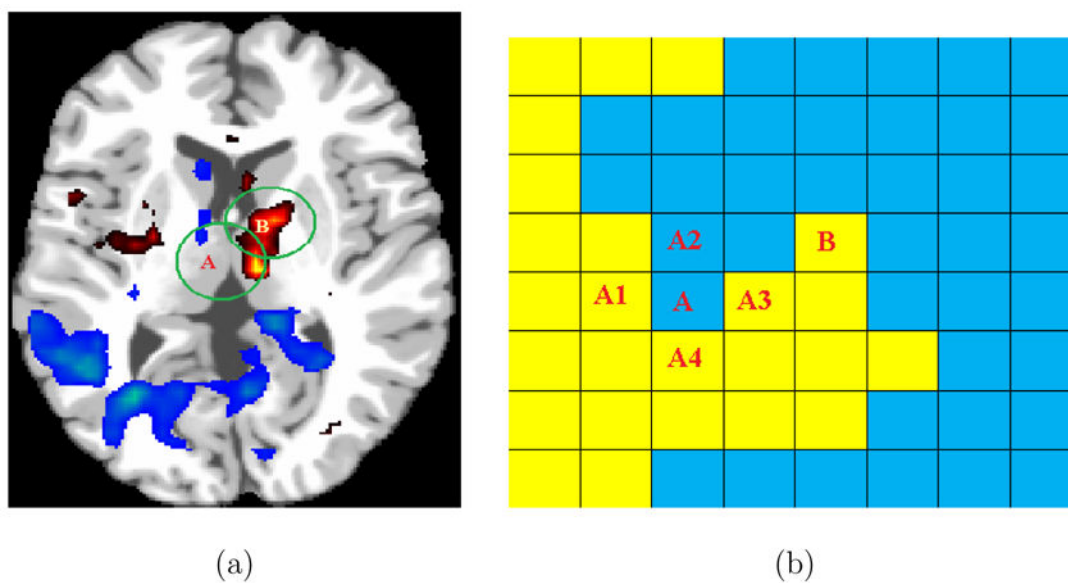
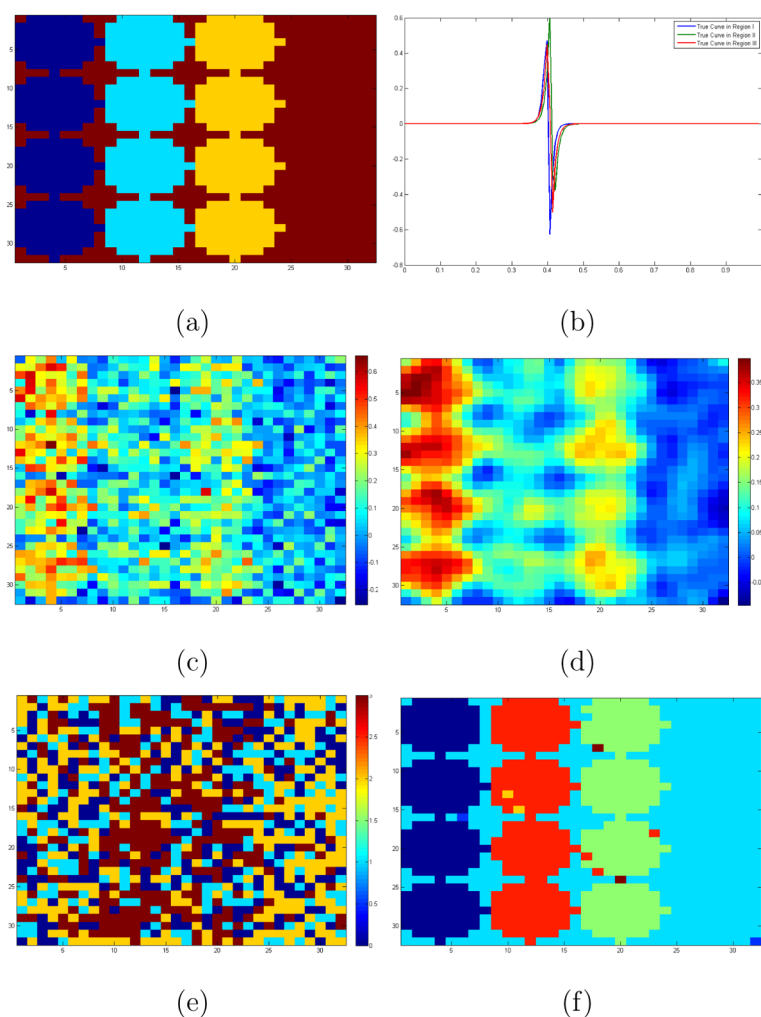


Figure 1. Spatial regions with sharp boundaries: (a) An fMRI slice example; (b) Illustration of spatial neighborhood in a 2-D space. A and B are two locations that are close to the boundary of an activated region. The online version of this figure is in color.

**Figure 2.**

First Simulation Setup and Results: (a) A temporal cut of the true active pattern; (b) The true temporal signal curves within the three different active spatial regions; (c) A temporal cut of the simulated images; (d) Gaussian smoothing of the simulated images; (e) A typical clustering result using FSCM; (f) A typical clustering result from our method (under development).

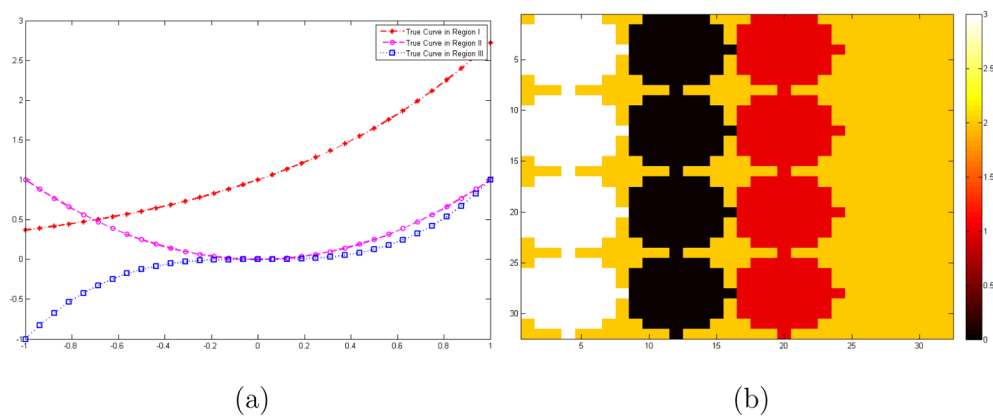


Figure 3. Second Simulation Result: (a) The true temporal signal curves within the three different active spatial regions; (b) A typical clustering result from FSCM. The online version of this figure is in color.