Supplemental Material for “Nonparametric Bayes Models of Fiber Curves Connecting Brain Regions”

Zhengwu Zhang, Maxime Descoteaux and David B. Dunson

October 18, 2018

1 Posterior sampling for product kernel mixture model

Posterior sampling procedures for model (6) are given as follows:

1. Update the cluster allocation of S_i for each fiber curve from the conditional posterior with

$$\Pr(S_i = h|-) = \frac{\pi_h \prod_{m=1}^{M} \mathcal{K}_m(\boldsymbol{c}_i^{(m)}; \theta_h^{(m)})}{\sum_{l=1}^{K} \pi_l \prod_{m=1}^{M} \mathcal{K}_m(\boldsymbol{c}_i^{(m)}; \theta_l^{(m)})}.$$

2. Update the weights on each component from the conjugate conditional posterior

$$(\boldsymbol{\pi}|-) \sim Dir\bigg(\frac{\alpha}{K} + n_1, ..., \frac{\alpha}{K} + n_K\bigg),$$

   where n_h is the number of observations with cluster h.

3. Update the parameter $\theta_h^{(m)}$ for $m = 1, ..., M$ and $h = 1, ..., k$ from

$$(\theta_h^{(m)}|-) \propto P_0^m(\theta_h^{(m)}) \prod_{i:S_i=h} \mathcal{K}_m(\boldsymbol{c}_i^{(m)}; \theta_h^{(m)}),$$

where P_0^m is a conjugate prior to $\mathcal{K}_m(\boldsymbol{c}_i^{(m)}; \theta_h^{(m)})$ for each component m; in particular, we use Gaussian-Inverse Wishart priors.

2 Jointly model fiber curves and connection strength

Model (7) does not incorporate information on the strength of connection w_j between the two ROIs within individual j. However, it is straightforward to generalize the model to include this additional information by letting

$$G_j \sim \sum_{h=1}^{\infty} \pi_h^* \delta_{\{G_h^*(\cdot), \psi_h\}},$$

where now the hth component of G_j includes not only the mixing measure G_h^* characterizing the distribution of fiber curves in that component but also parameters ψ_h within a kernel $\mathcal{K}_w(\cdot;\psi_h)$ for the measure of connection strength. The resulting joint model characterizes flexible dependence in the connection strength and fiber curves through shared dependence on the individual's cluster allocation. For continuous measures of connection strength w_j, we can simply use a Gaussian kernel. However, we will focus on w_j equal to the number of connections between the regions of interest, so that $\mathcal{K}_w(w;\psi)$ is a parametric distribution with support on the non-negative integers. To induce this kernel, we can apply the approach of Canale and Dunson (2011) and simply 'round' a Gaussian kernel with unknown mean and variance, with negative values mapped to 0, values in (0,1) mapped to 1, values in (1,2) mapped to 2 and so on. Posterior sampling can proceed via a slight modification of the sampler in Section 4.2.

Posterior inference for jointly modeling: The connection strength w_j is defined to be the count of fibers. To model this count variable independently, one can rely on the rounded kernel mixture model proposed by Canale and Dunson (2011). Here we present a simple Gibbs sampling algorithm for this joint modeling. Some modification of the Gibbs sampler in Section 4.2 is needed. Again, let ζ_j, for $j = 1, ..., J$, be the membership indicator of individuals and let ξ_{ji}, for $i = 1, ..., n_j$, be the membership indicator of fiber curves for the jth subject. Sampling proceeds via the following steps:

1. Sample the membership indicator for the jth individual ($j = 1, ..., J$) from a multinomial distribution with probabilities

$$P(\zeta_j = h|-) \propto \pi_h^* p(w_j|\psi_h) \prod_{i=1}^{n_j} \sum_{l=1}^{L} w_{lh}^* \prod_{m=1}^{M} \mathcal{K}_m(c_{ji}^{(m)}|\theta_{lh}^{(m)^*}),$$

   where $p(w_j|\psi_h) = \Phi(a_{w_j+1}|\psi_h) - \Phi(a_{w_j}|\psi_h)$.

2. Sample the membership indicator ξ_{ji}, for $j = 1, ..., J$ and $i = 1, ..., n_j$, from the following multinomial distribution:

$$P(\xi_{ji} = l|-) \propto w_{l\zeta_j}^* \prod_{m=1}^{M} \mathcal{K}(c_{ji}^{(m)}|\theta_{l\zeta_j}^{(m)^*}).$$

3. Sample π_h^* by the following procedure: First sample u_h according to $(u_h^*|-) \sim$ beta$(1 + m_h, \alpha + \sum_{s=h+1}^{K} m_s), h = 1, ..., K-1,$ and $u_K^* = 1$, where m_h is the number of subjects assigned to cluster h. π_h^* is constructed as

$$\pi_h^* = u_h^* \prod_{s=1}^{h-1}(1 - u_s^*).$$

4. Generate the auxiliary variable y_j^*, $j = 1, ..., J$, for the connection strength part: Sample $t_j \sim U(\Phi(a_{w_j};\psi_{\zeta_j}), \Phi(a_{w_j+1};\psi_{\zeta_j}))$ and let $y_j^* = \Phi^{-1}(t_j;\psi_{\zeta_j})$.

5. Sample w_{lh}^* by the following procedure: Sample $(v_{lh}^*|-) \sim$ beta$(1+n_{lh}, \beta+\sum_{s=l+1}^{L} n_{sh}), l = 1, ..., L-1, h = 1, ..., K$ and $v_{LK}^* = 1$, where n_{lh} is the number of observations assigned to

atom l of distribution h. w_{lh}^* is constructed as

$$w_{lh}^* = v_{lh}^* \prod_{s=1}^{l-1} (1 - v_{sh}^*).$$

6. Sample the parameter ψ_h for $h = 1, ..., K$ according to:

$$P(\psi_h | \cdot) \propto P_0^{\psi}(\psi_h) \prod_{\{j|\zeta_j = h\}} N(y_j^*; \psi_h),$$

   where $P_0^{\psi}(\cdot)$ is a Gaussian-gamma conjugate prior distribution for ψ. If there is no observation for the hth cluster, we sample ψ_h from P_0^{ψ}.

7. Sample the parameters $\theta_{lh}^{(m)^*}$ for $l = 1, ..., L$, $h = 1, ..., K$ and $m = 1, ..., M$ from

$$P(\theta_{lh}^{(m)*} | -) \propto P_0^m(\theta_{lh}^{(m)*}) \left(\prod_{\{i,j|\zeta_j = h, \xi_{ij} = l\}} \mathcal{K}_m(c_{ji}^{(m)} | \theta_{lh}^{(m)*}) \right),$$

   where $P_0^m(\cdot)$ is the conjugate prior for the parameter in $\mathcal{K}_m(\cdot | \theta^{(m)})$. If no observation is assigned to the cluster, we draw $\theta_{lh}^{(m)*}$ from the prior P_0^m.

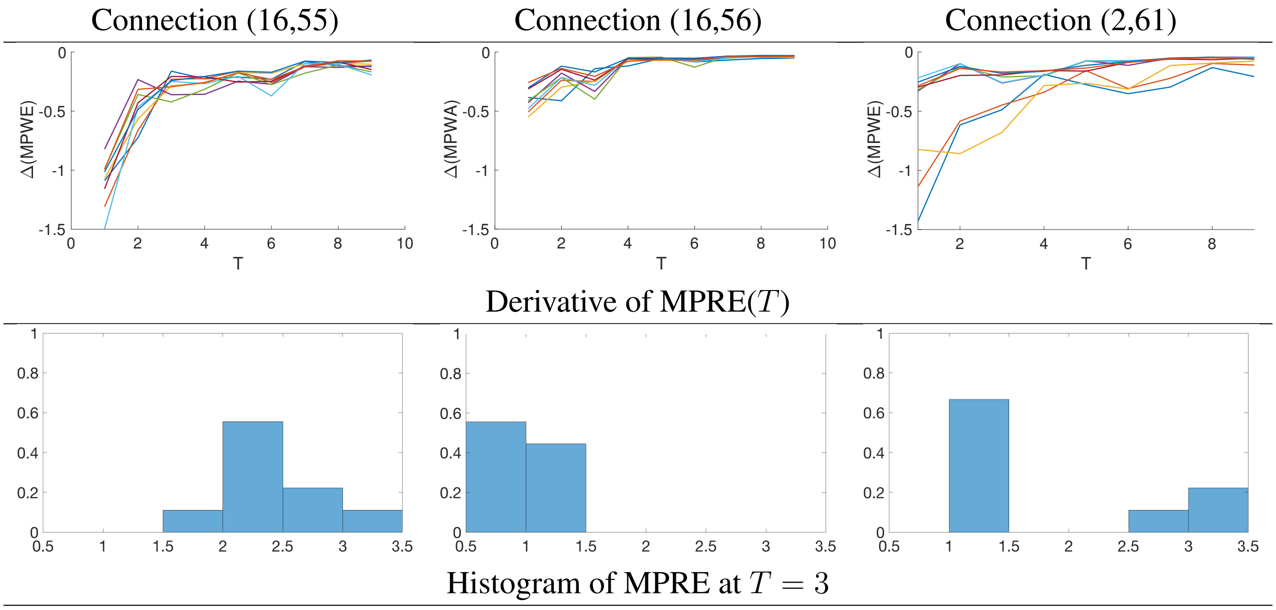
8. Sample the concentration parameters α and β. We choose conjugate priors: $\alpha \sim \text{gamma}(a_\alpha, b_\alpha)$ and $\beta \sim \text{gamma}(a_\beta, b_\beta)$. The posterior samples for α and β are constructed as

$$P(\alpha | -) \sim \text{gamma}(a_\alpha + (K-1), b_\alpha - \sum_{h=1}^{K-1} \log(1 - \mu_h^*)),$$

$$P(\beta -) \sim \text{gamma}(a_\beta + K(L-1), b_\beta - \sum_{l=1}^{L-1} \sum_{h=1}^{K} \log(1 - v_{lh}^*)).$$

3 Selection of T

The number of coefficients T is an important parameter for precisely representing fiber curves. The higher the T the better the representation power. Selection of T is similar to selection of the number of PC components in PCA, where the goodness-of-fit for a given T can be evaluated through the cumulative percentage of total variation explained or some formal tests of hypothesis (Jolliffee, 1986; Choi et al., 2017). However, these procedures can not be directly applied since we have a different setting: we represent shapes of fibers in new subjects using a fixed set of basis functions learned from training data; and there are other components (e.g., rotation and translation) contributing to the fiber recovery (refer to Section 2.3). We borrow some ideas from the PCA (Jolliffee, 1986) components selection and utilize the following criteria to evaluate the choice of T.



Connection (16,55) Connection (16,56) Connection (2,61)

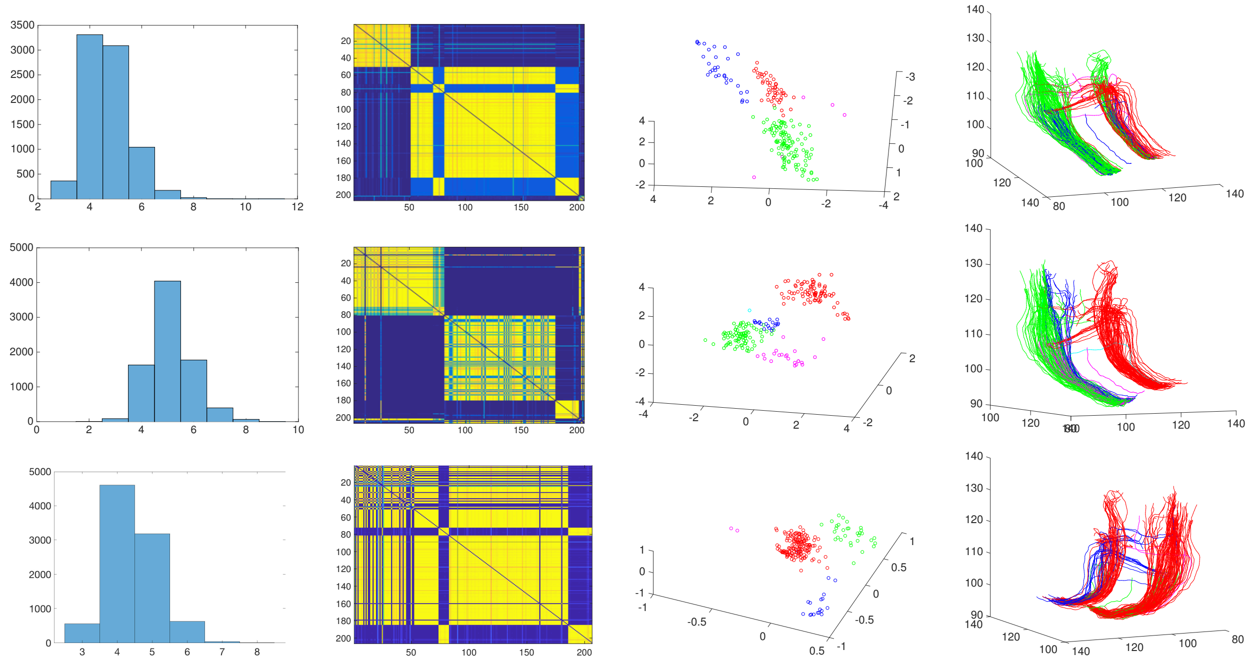
Derivative of MPRE(T)

Histogram of MPRE at $T = 3$

Figure 1: The first row shows the derivative of MPRE at different T value (the MPRE change as T increases to $T+1$). The second row shows the histogram of MPRE at $T = 3$ for the 3 connections. The unit for MPRE is mm.

For a bundle with N fibers, the goodness-of-fit can be evaluated through mean point-wise representation error (MPRE), defined as $(\sum_{j=1}^{N}\sum_{i=1}^{m}\|y_j(s_i) - \hat{y}_j(s_i)\|)/Nm$, where $y_j(s_i)$ represents the j-th fiber's coordinate observed at s_i and $\hat{y}_j$ represents the recovered fiber with T coefficients. MPRE is a function of T (its value will decrease as T increases). We want to select a T such that (1) MPRE is within a reasonable range and (2) it explains most of the variation in the data. We define cumulative percentage of total variation (CPTV) explained by T coefficients as $100 * (1 - \sum_{j=1}^{N}\|y_j - \hat{y}_j\|^2 / \sum_{j=1}^{N}\|y_i - \mu\|^2)$. We conducted MPER analysis for three representative connections (indexed as (16,55), (16,56) and (2,61)) in 9 scans of the test-retest dataset. From Figure 1 first row, we can see that MPRE levels off and does not increase significantly for $T > 3$ in almost all cases. We can see that for $T > 3$ the increase of T does not significant decrease the MPRE anymore in almost all cases. Also, at $T = 3$, we have CPTV = $91.46 \pm 3.71\%$ for (16,55), CPTV = $97.9 \pm 1.0\%$ for (16,56) and CPTV = $97.1 \pm 2.20\%$ for (2,61). Figure 1 second row shows the histogram of MPRE at $T = 3$ for the 9 scans of the 3 connections, indicating that the average fiber recovery error is often less than 2 voxels (each voxel has the size of $2 \times 2 \times 2\ mm^3$, and the unit for MPRE is mm).

These results encourage our choice of $T = 3$, especially when we only consider the global geometry of fiber bundles and deal with hundreds and thousands of streamlines (a lower dimension is important for reducing the computational cost of our model). We set $T = 3$ in all our real data applications.



(a) Posterior dist. of k (b) Pairwise Prob. (c) Clustering of $\{c_i^{(m)}\}$ (d) Clustering of fibers

Figure 2: Posterior summary for the connection between *r_pl* and *l_pcc*. The three rows are results for the shape, translation and rotation component, respectively.

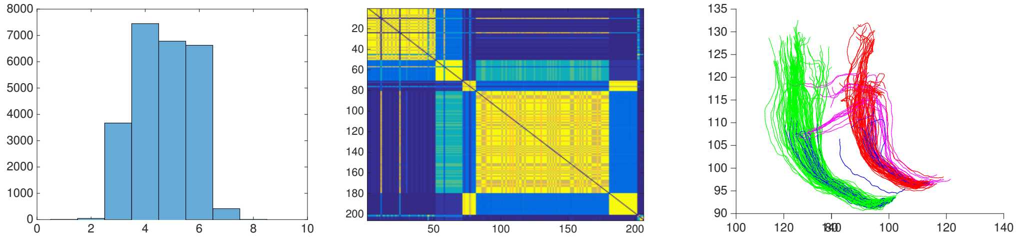
4 Experimental results for modeling a single connection

Figure 2 shows posterior inference results for the connection (*r_pl, l_pcc*). This connection contains 206 fiber curves. In column (c), we plot these multivariate components, and their final clustering configuration. We see that the shape component has three major clusters (shown in red, green and blue). The red ones have L shape; the green and blue ones have similar shape but can be distinguished through their lengths. The translation component seems to have three major clusters. Our algorithm broke one cluster into two, and thus gave us four major clusters. The result based on fusion of three components together is shown in Figure 3. Four clusters are identified in the final clustering results. The green and red clusters are two major long fiber bundles in this connection. They have similar shapes, but can be distinguished by their locations. The magenta cluster has an "L" shape, and the blue cluster contains straight and short fibers.

Figure 4 shows posterior inference results for the connection (*r_pl*, *l_pg*). Column (a) shows posterior samples of number of clusters and (b) shows the pairwise probability heat map according to the posterior samples. Figure 4 (c) and (d) show the final clustering results on $\{\boldsymbol{c}_i^{(m)}\}$ and the original fibers.

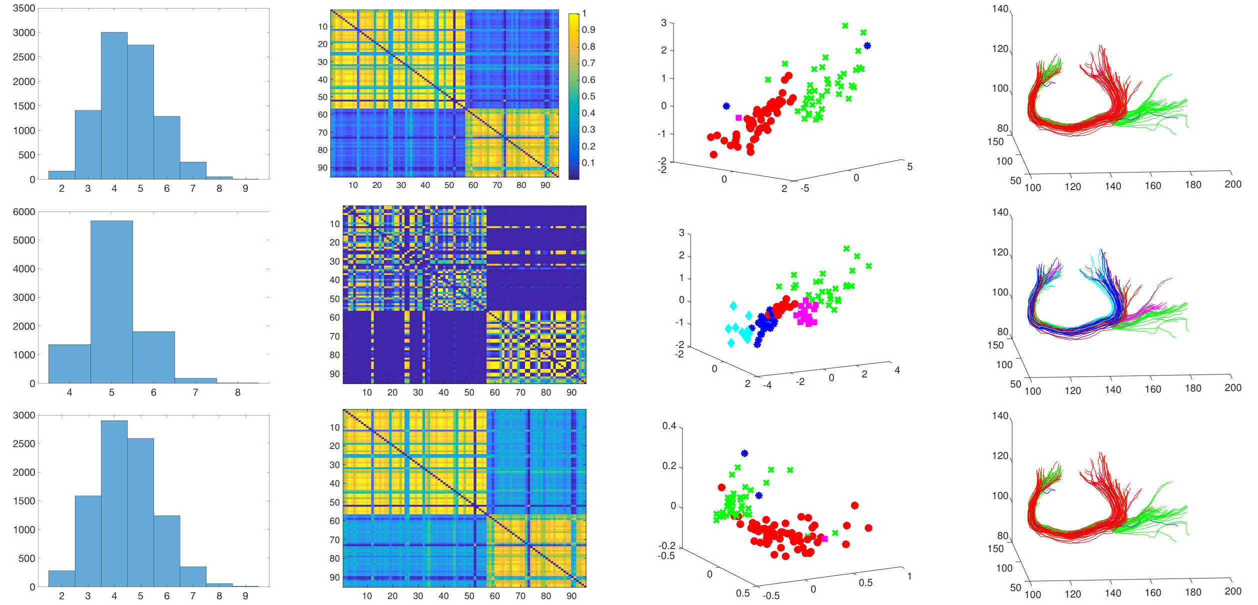
5 Manual clustering of fibers in (*r_pl, l_pg*) and (*r_pl, l_pcc*)

The final manual clustering of fibers in (*r_pl, l_pg*) is shown in Figure 5. Fibers in this connection split into two bundles, and the manual clustering is based on this split. We identified two clusters.



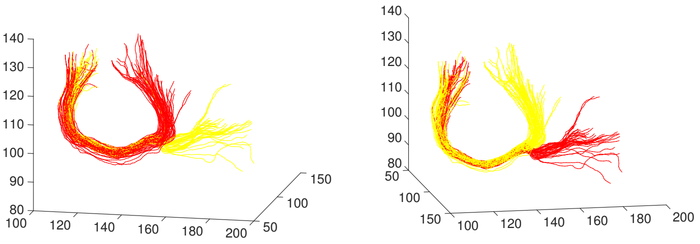
(a) Posterior dist. of k (b) Adjacency matrix (c) Clustering on raw fibers

Figure 3: Joint model result for the connection (*r_pl, l_pcc)*.



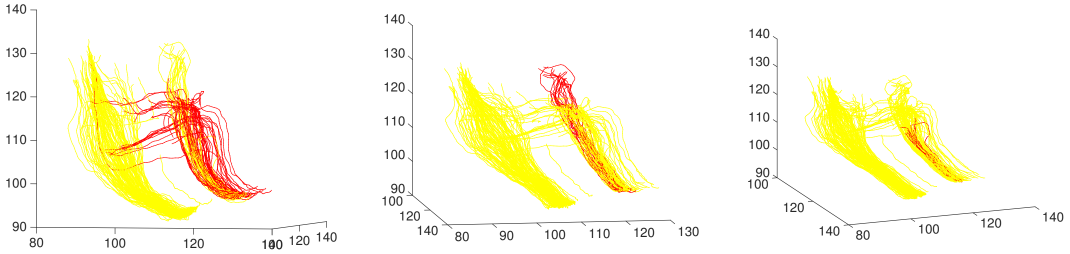
(a) Posterior dist. of k (b) Adjacency matrix (c) Clustering of $\{c_i^{(m)}\}$ (d) Clustering of fibers

Figure 4: Posterior summary for the connection (*r_pl*, *l_pg*). The three rows are results for the shape, translation and rotation component, respectively.

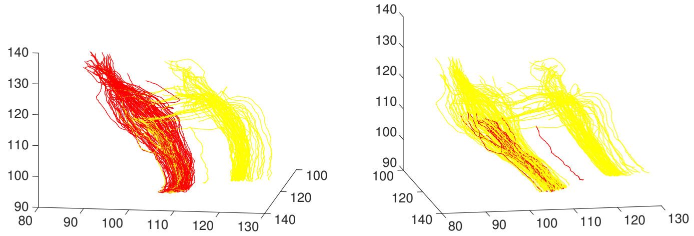


(a) Cluster 1 (b) Cluster 2

Figure 5: Manual clustering of (*r_pl, l_pg*). Red ones are fibers in current cluster and yellow ones are remaining fibers in the connection.



(a) Cluster 1 (b) Cluster 2 (c) Cluster 3



(d) Cluster 4 (e) Cluster 5

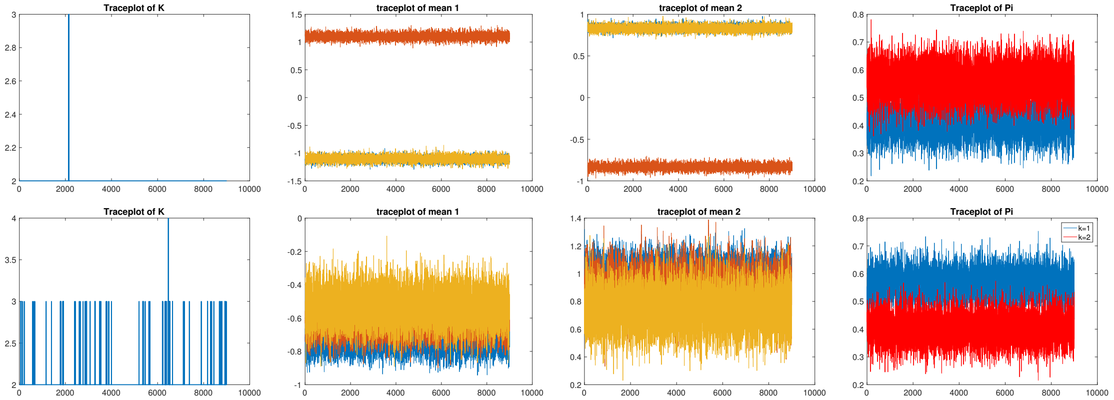
Figure 6: Manual clustering of (*r_pl, l_pcc*).

In cluster 1 there are a few fibers having distinct shapes (change their directions significantly near the end points), but we did not distinguish them.

The final manual clustering of fibers for the other connection (*r_pl, l_pcc*) are shown in Figure 6. We identified five clusters based on their shapes and locations. First, fibers were clustered into two major clusters based on their locations. Red fibers in the first row are from the first major cluster and red fibers in the second row are from the second major cluster. Next we refined this clustering based on their shapes and obtained 3 sub-clusters for the first major cluster and 2 sub-clusters for the other major cluster.

6 MCMC Diagnosis

The following MCMC diagnosis is performed for each run. We first perturbed the prior distribution within a reasonable range, and checked if the final results and inferences change. Next, we



(a) (b) (c) (d)

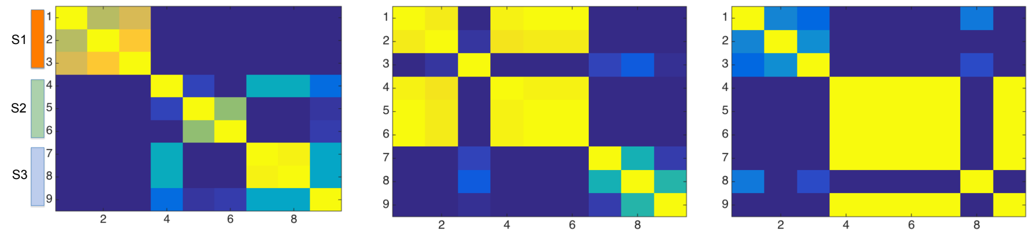
Figure 7: Trace plots of some parameters in the MCMC chains generating Figure 6 (first row) and 8 (second row) in the main paper. (a) k (number of clusters); (b) mean of the shape component for the first cluster (the mean is in $\Re^3$); (c) mean of the shape component for another cluster; and (d) π for the two clusters.

specified different initializations for the MCMC chain. All results remain similar, indicating that the MCMC sampling strategies and Bayesian inferences are robust.

Next, we used standard convergence diagnostic tools to study the MCMC chains. We performed post-processing to fix the label switching issue caused by our discrete mixture models. Trace plots and the Gelman-Rubin R statistics were used for convergence diagnosis. Figure 7 shows the MCMC trace plots of the cluster number k, means and π for data in Figure 7 in the paper. The effective sample size is $> 90\%$ of the total number of samples. The Gelman-Rubin R statistics for all parameters are close to 1 (e.g 1.0005, 1.0004 and so on), indicating no evidence of lack of convergence. Similar studies were also performed (perturbation of the initialization, visual examination of the trace plots and Gelman-Rubin R statistics) to examine and confirm the convergence for other MCMC chains in the Section 5.3.

7 Results on 9 scans of 3 subjects

This section presents the results for analyzing 9 scans of 3 subjects in Figure 8 in the main paper. Pairwise probability heat maps in different scenarios are shown in Figure 8, showing clustering results based on (a) only shape, (b) only translation, and (c) both shape and translation. The 9 scans were ordered by concatenating columns of Figure 8 in the main paper (scans of the same subject are next to each other). From (a) we observe that, if we only use the shape part, the posterior clustering result favors five clusters, e.g., the 3 scans of subject 1 are clustered together; the scan 2 and 3 of subject 2 are clustered together, while the scan 1 is a separate cluster. We then inspected this result more closely: their shape components are plotted in Figure 9. The fibers in the 3 scans of subject 1 have a very different distribution from data of subject 2 and 3 in their shapes; scan 2 and 3 of subject 2 are different from scan 1. We compared the NDP clustering result with the ground truth subject ids and the results are shown in Table 1. We can see that the shape part has the best clustering performance. Combining shape and translation does not improve clustering.



(a) Shape (b) Trans. (c) Shape & Trans.

Figure 8: Pairwise probabilities of clustering for 9 scans of 3 subjects in Figure **??**.

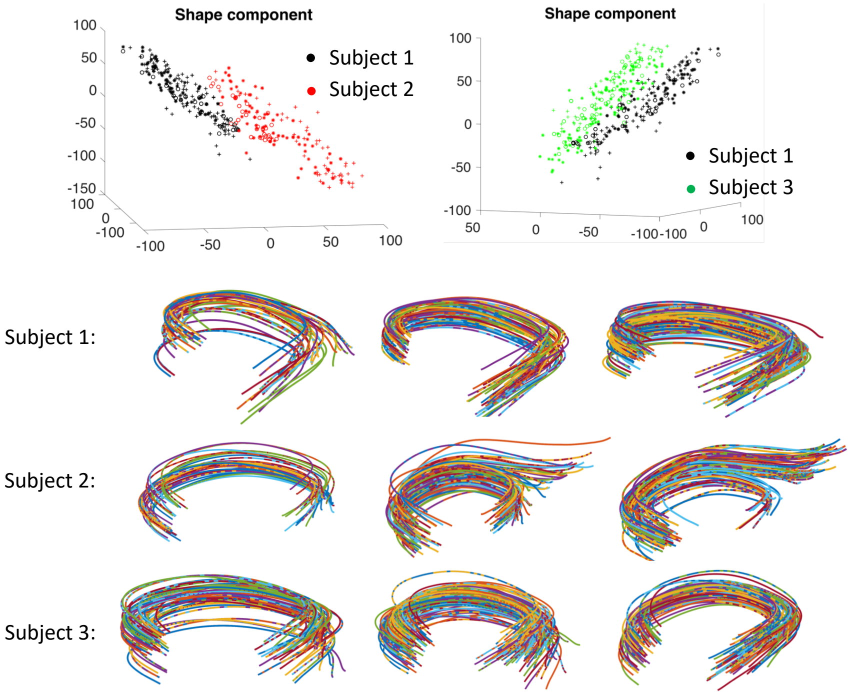
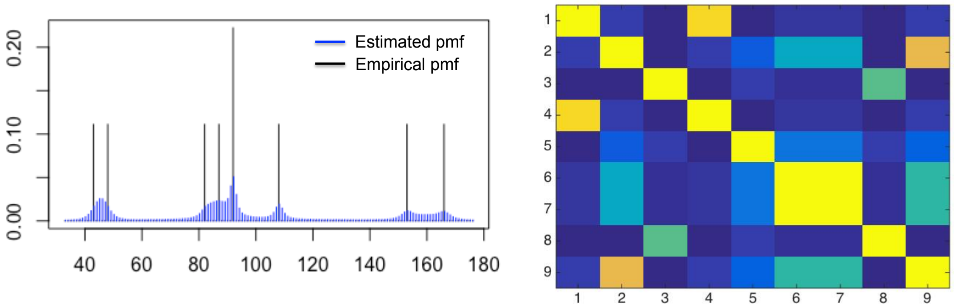


Figure 9: The first row shows the mapped shape components in $\Re^3$ for fiber curves in the three subjects. The recovered fiber curves with used only 9 parameters (3 for shape component, 3 for rotation and 3 for translation).

Table 1: Clustering of subjects using fiber curves connecting (*r_pl*, *l_pg*).

	Shape	Trans.	Shape & Trans.	Count
RI	0.8889	0.7222	0.7222	0.6389
ARI	0.6522	0.3130	0.3130	-0.1818



(a) (b)

Figure 10: (a) Posterior estimation of the pmf for fiber counts; (b) Heat map of the adjacency matrix.

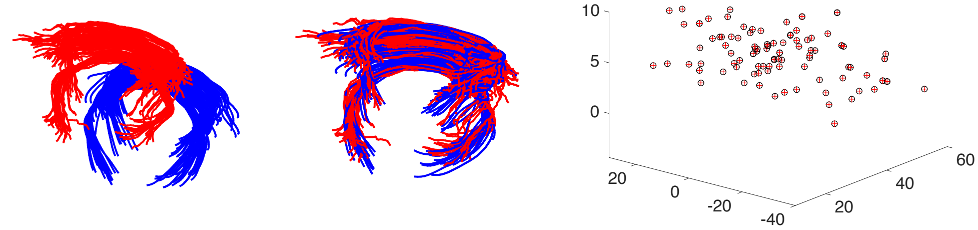
As a comparison, we clustered subjects according to their fiber counts by the rounded kernel mixture model of Canale and Dunson (2011), using their recommended priors, collecting $10,000$ posterior draws, and discarding the first $1,000$. Figure 10 shows the result, with (a) the estimated fiber count pmf for the 9 scans and (b) pairwise probabilities that two elements are clustered together. The estimated pmf illustrates the enormous heterogeneity in the counts, with five peaks in the distribution. From the pairwise probability matrix, scans for the same subject are not reliably clustered together. The Rand index and adjusted Rand index of the final clustering configuration are reported in Table 1. These results illustrate that fiber counts have very high variability and cannot reliably distinguish between subjects.

8 Sensitivity to misalignment

By definition the shape component is invariant to the misalignment but other components are not. To illustrate this, we take 9 scans (from three subjects) in the test retest dataset and permute their tractographic data with random rotations and translations. The random rotation is generated through Euler angles (each of them is drawn from $N(0, 0.2^2)$) and the translation is generated from normal distributions (each of x, y, z coordinates is from $N(6, 4)$). Figure 11 panel (a) illustrates the fiber bundles in a randomly selected connection (31,47) for scan 1 and scan 2 of subject 1 before and after permutation. We then align them back using the algorithm in Garyfallidis et al. (2015), and the aligned bundles are shown in panel (b). In panel (c) of Figure 11, we show the calculated shape components before and after alignment (for the moving blue bundle). We see that the shape component is invariant to the permutation/misalignment. In Table 2, we compare our NDP model's result for different components, where we can see that, the shape component is invariant to the misalignment, but the translation and rotation are not.

9 Subsampling of fibers and more results

Among the 45 connections, some of them contain very rich fiber curves, e.g. with more than 10000 fibers. In order to efficiently fit our NDP model and infer the posterior distribution of fiber curves,



(a) Before alignment of scan 1 (red) and scan 2 (blue) from S1

(b) After alignment of scan 1 and scan 2 from S1

(c) Shape component of scan 2 before and after alignment

Figure 11: Shape component is preserved before and after alignment.

Table 2: Misalginment sensitivity analysis using connection (31,47)

	Before alignment				After alignment			
	Shape	Trans.	Rot.	Shape&Trans.	Shape	Trans.	Rot.	Shape&Trans.
RI	0.8611	0.75	0.6667	0.36	0.8611	0.75	0.5278	0.75
ARI	0.6429	0	0.1724	0	0.6429	0.5	0.1500	0.5

we randomly subsample fiber curves in each connection. To study how many fibers (denoted as n) to keep for each connection for our analyses, we run our NDP model with different n's for 10 randomly selected ROI pairs. Figure 12 shows the RI and ARI of 10 randomly selected ROI pairs. We observe that although with a bigger n, we generally can obtain better results, the decline of the clustering performance is slow with the decreasing of n (when $n \geq 100$).

The selected 45 connections involve 32 different regions. In Table 3, we list the name of these regions according to the Desikan-Killiany atlas (Desikan et al., 2006). Table 4 shows the clustering results for the remaining 27 connections (as a complement of Table 2 in the paper).

10 HCP Data Selection and exploratory analysis

From the Human Connectome Project data, according to (Zhang et al., 2018), we first select a trait of reading (NIH Toolbox Oral Reading Recognition Test: Age-Adjusted Scale Score) that potentially relates to structural connectome. Figure 13 shows the histogram of this trait score. We then order this trait and select two groups of subjects - one group with the highest scores and one group with the lowest scores. The two groups have equal number of subjects - assume each group have m subjects. In the first experiment of Section 5.4, we have $m = 10$ and in the second experiment, we have $m = 50$.

Next, we visually checked the subjects' T1 brain images and found that in general, people with very low reading scores tend to have smaller brains with less complex gyrus and sulcus folding patterns compared with people with high reading scores. Some examples of the T1 images and their reading scores are shown in Figure 14.

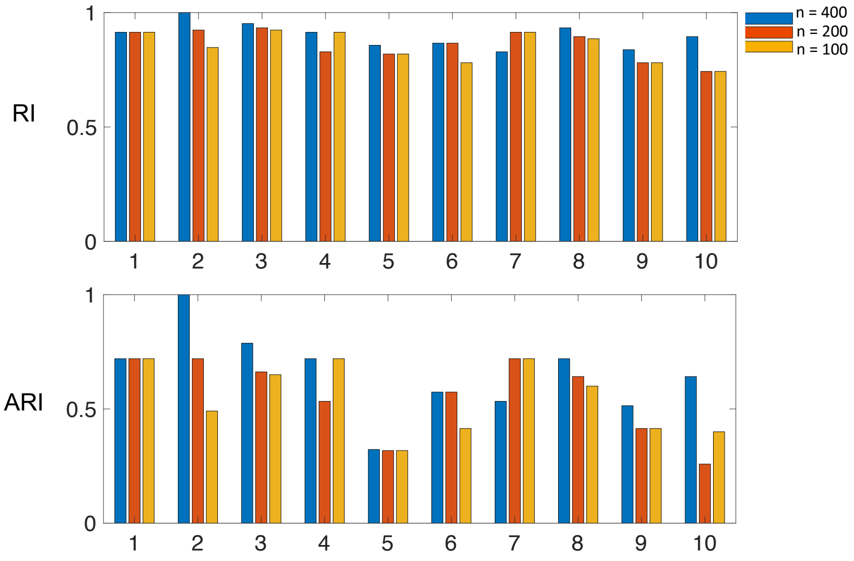


Figure 12: Clustering results (RI and ARI) for 10 randomly selected ROI pairs under different sample size n.

Table 3: Names and indices of selected regions in Desikan-Killiany atlas

Index	Name	Index	Name
2	ctx-lh-caudalanteriorcingulate	37	ctx-rh-caudalmiddlefrontal
3	ctx-lh-caudalmiddlefrontal	43	ctx-rh-isthmuscingulate
7	ctx-lh-inferiorparietal	45	ctx-rh-lateralorbitofrontal
9	ctx-lh-isthmuscingulate	47	ctx-rh-medialorbitofrontal
11	ctx-lh-lateralorbitofrontal	50	ctx-rh-paracentral
13	ctx-lh-medialorbitofrontal	51	ctx-rh-parsopercularis
16	ctx-lh-paracentral	53	ctx-rh-parstriangularis
22	ctx-lh-posteriorcingulate	55	ctx-rh-postcentral
23	ctx-lh-precentral	56	ctx-rh-posteriorcingulate
24	ctx-lh-precuneus	57	ctx-rh-precentral
25	ctx-lh-rostralanteriorcingulate	58	ctx-rh-precuneus
26	ctx-lh-rostralmiddlefrontal	59	ctx-rh-rostralanteriorcingulate
27	ctx-lh-superiorfrontal	60	ctx-rh-rostralmiddlefrontal
28	ctx-lh-superiorparietal	61	ctx-rh-superiorfrontal
31	ctx-lh-frontalpole	62	ctx-rh-superiorparietal
36	ctx-rh-caudalanteriorcingulate	64	ctx-rh-supramarginal

Table 4: Comparison of clustering results of using geometric information and count.

RI/ARI	(22,61)	(23,50)	(23,56)	(23,57)	(23,61)	(24,43)	(24,58)	(24,62)
Shape	0.74/0.40	0.90/0.64	0.91/0.72	0.91/0.72	0.71/0.30	0.91/0.72	0.83/0.53	0.91/0.72
Trans.	0.35/0.05	0.50/0.06	0.55/0.17	0.62/0.13	0.74/0.31	0.81/0.46	0.74/0.40	0.77/0.35
Shape &Trans	0.74/0.40	0.66/0.30	0.80/0.28	0.76/0.22	0.52/0.15	0.75/0.33	0.66/0.30	0.91/0.72
Count	0.58/0.19	0.58/0.19	0.58/0.19	0.58/0.14	0.70/0.15	0.58/0.19	0.79/0.23	0.45/0.01
RI/ARI	(25,61)	(26,36)	(26,47)	(26,59)	(26,60)	(26,61)	(27,36)	(27,37)
Shape	0.91/0.72	0.91/0.72	0.64/0.24	0.55/0.17	0.91/0.72	0.91/0.72	0.83/0.53	0.87/0.57
Trans.	0.65/0.18	0.49/0.16	0.49/0.16	0.68/0.18	0.49/0.16	0.58/0.19	0.83/0.53	0.71/0.28
Shape &Trans	0.70/0.11	0.66/0.30	0.32/-0.00	0.47/0.11	0.66/0.30	0.49/0.16	0.83/0.53	0.70/0.30
Count	0.64/0.24	0.58/0.19	0.52/0.15	0.58/0.19	0.14/0.00	0.55/0.05	0.14/0.00	0.64/0.24
RI/ARI	(27,47)	(27,51)	(27,53)	(27,56)	(27,57)	(27,59)	(27,60)	(27,61)
Shape	0.78/0.41	0.86/0.32	0.90/0.50	0.82/0.43	0.91/0.72	0.74/0.40	0.91/0.72	0.74/0.40
Trans.	0.54/0.07	0.88/0.48	0.69/0.06	0.83/0.53	0.49/0.16	0.45/0.01	0.75/0.37	0.60/0.16
Shape &Trans	0.14/0.00	0.78/0.21	0.71/0.25	0.70/0.30	0.66/0.30	0.75/0.30	0.74/0.40	0.49/0.16
Count	0.64/0.24	0.49/0.16	0.64/0.24	0.14/0.00	0.80/0.11	0.52/0.15	0.14/0.00	0.14/0.00
RI/ARI	(28,43)	(28,58)	(28,62)	(28,64)	(31,47)			
Shape	1.00/1.00	0.91/0.72	0.90/0.64	0.79/0.19	0.91/0.72			
Trans.	0.74/0.40	0.54/0.12	0.78/0.41	0.81/0.18	0.84/0.47			
Shape &Trans	0.75/0.37	0.70/0.30	0.74/0.40	0.73/0.07	0.66/0.30			
Count	0.45/0.01	0.14/0.00	0.31/-0.01	0.49/0.16	0.49/0.16			

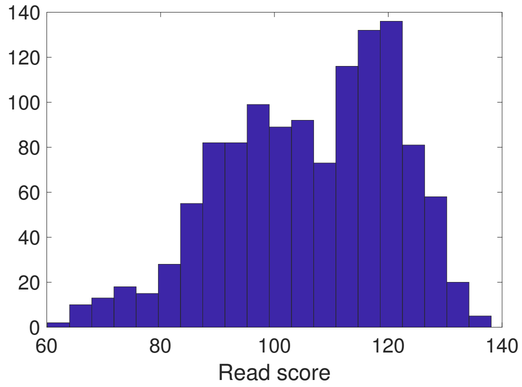
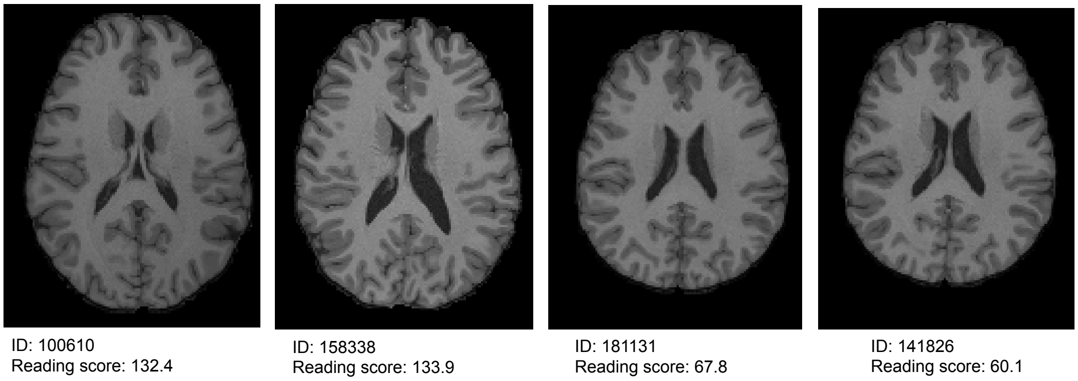


Figure 13: Histogram of English reading scores in the HCP after age adjusting.



ID: 100610
Reading score: 132.4

ID: 158338
Reading score: 133.9

ID: 181131
Reading score: 67.8

ID: 141826
Reading score: 60.1

Figure 14: T1 images of subjects in the HCP dataset with their reading scores.

References

Canale, A. and Dunson, D. B. (2011). Bayesian kernel mixtures for counts. *Journal of the American Statistical Association*, 106(496):1528–1539.

Choi, Y., Taylor, J., Tibshirani, R., et al. (2017). Selecting the number of principal components: Estimation of the true rank of a noisy matrix. *The Annals of Statistics*, 45(6):2590–2617.

Desikan, R. S., Sgonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968 – 980.

Garyfallidis, E., Ocegueda, O., Wassermann, D., and Descoteaux, M. (2015). Robust and efficient linear registration of white-matter fascicles in the space of streamlines. *NeuroImage*, 117:124–140.

Jolliffee, I. T. (1986). *Principal Component Analysis*. Springer.

MacCallum, R. C., Zhang, S., Preacher, K. J., and Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological methods*, 7(1):19.

Zhang, Z., Allen, G., Zhu, H., and Dunson, D. (2018). Relationships between human brain structural connectomes and traits. *bioRxiv:10.1101/256933*.