35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Online Agglomerative Hierarchical Clustering of Neural Fiber Tracts

Ali Demir[1], Ashraf Mohamed[2], and H. Ertan Çetingül[3]

Abstract— **We consider the problem of clustering neural fiber pathways, produced from diffusion MRI data via tractography, into different bundles. Existing clustering methods often suffer from the burden of computing pairwise fiber (dis)similarities, which escalates quadratically as the number of fiber pathways increases. To address this challenge, we adopt the scenario of clustering data streams into the fiber clustering framework. Specifically, we propose to use an online hierarchical clustering method, which yields a framework similar to doing clustering while simultaneously performing tractography. We evaluate the proposed method through experiments on phantom and real diffusion MRI data. Experiments on phantom data evaluate the sensitivity of our method to initialization and show its superior performance compared with alternative methods. Experiments on real data demonstrate the accuracy in clustering selected white matter fiber tracts into anatomically consistent bundles.**

I. Introduction

Diffusion magnetic resonance imaging (dMRI) can describe the white matter (WM) architecture in vivo by quantifying variations in water diffusion patterns as images of the MR signal attenuation along a set of directions. These diffusion weighted images provide insights into the ensemble average propagator [1], a function that reflects the underlying local fiber orientations in white matter. This unique property of dMRI has generated an influx in the development of tools for inferring the structural brain connectivity from dMRI data, because such tools can advance research on several neurological diseases and psychiatric disorders [2].

Quantification of changes in brain connectivity (i.e., the WM architecture) often includes *diffusion estimation and representation* (e.g., diffusion tensors [3], orientation distribution functions [4], [5], etc.), *tractography* [6], and *statistical analysis* based on the extracted fiber pathways [7]. In clinical applications, it is a common practice to focus on selected WM fiber bundles (i.e., collections of pathways), whose architecture is anticipated to be affected by development, degeneration or disease, by means of manually placing regions of interest (ROIs). This cumbersome and error-prone step is usually necessary to eliminate spurious pathways and/or select a collection of spatially similar ones. In that regard, automatically grouping similar pathways into bundles that are consistent with the underlying neuroanatomy is important for fast, reliable, and reproducible analysis of brain connectivity.

Prior work on WM fiber clustering can be roughly categorized into two groups: i. Methods that perform heavy preprocessing to cluster a large number of fiber pathways (e.g., resulting from the whole-brain tractography), and consider accuracy in clustering, rather than runtime, as the sole performance criterion; ii. Simple batch methods that tackle the problem of clustering a small number of pathways. For instance, in the former, there are sophisticated (and often data-dependent) frameworks, which use different grouping strategies, as well as anatomical atlas guidance [8]–[10]. The latter group contains methods largely based on spectral clustering [10], manifold learning [11], affinity propagation [12], hierarchical clustering [13], and Bregman soft clustering [14]. Some of these methods suffer from the burden of computing pairwise fiber (dis)similarities, which escalates quadratically as the number of pathways increases linearly. However, to our knowledge, *online clustering* (e.g., clustering streams of pathways) has not hitherto been considered as a solution to the problem of clustering large number of fibers.

In this work, we propose to use an *online hierarchical clustering* algorithm, which initially accepts a small set of fiber pathways to generate a user-specified number of clusters represented with parametric models. Then the algorithm either assigns a label to a recently extracted fiber pathway, or put that pathway in a reservoir for labeling following the update of the cluster parameters. We evaluate our method through experiments on phantom and real dMRI datasets.

II. Methods

We employ the generic online clustering framework described in [15], where the data streams (i.e., the set of fiber pathways $\{X_n\}_{n=1}^{N}$, where $X_n := (x_{n,1}, x_{n,2}, \dots)$ and $x_{n,m} \in \mathbb{R}^3$) provided at a particular time point can be analyzed using different baseline clustering methods. Here, we utilize the *hierarchical clustering* algorithm as the baseline, for which the number of clusters is specified beforehand.

A. Hierarchical Clustering

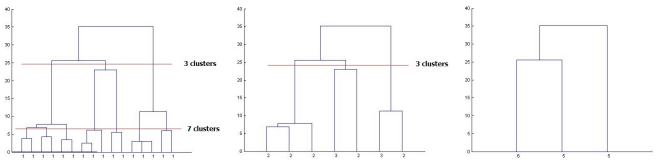
Hierarchical clustering (HC) is a greedy grouping strategy, which is either *agglomerative* or *divisive* [16]. Agglomerative HC is a bottom-up approach, where each *item* (in our case, a fiber pathway) is initiated as a different cluster center and the clusters are merged while moving up the hierarchy [17]. Divisive HC, on the other hand, initially assigns all the items to a single cluster and then splits it while moving down the hierarchy. The resulting hierarchical graph is called a *dendrogram* (see Fig. 1), which is used to choose the level

This work was supported in part by the NIH (R01 EB008432). A. Demir performed this work while at Siemens Corporation, Corporate Technology, Princeton, NJ 08540, USA.

[1]A. Demir (corresponding author) is with the Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey (e-mail: ademir at sabanciuniv.edu).

[2]A. Mohamed is with Imaging and Therapy Systems Division, Siemens Japan K.K., Tokyo, Japan (e-mail: ashraf.mohamed at siemens.com).

[3]H.E. Çetingül is with Imaging and Computer Vision TF, Siemens Corporation, Corporate Technology, Princeton, NJ 08540, USA (e-mail: hasan.cetingul at siemens.com).



(a) (b) (c)

Fig. 1. Dendrograms (of synthetically generated items) with different agglomerative merging levels. The y-axis is the dissimilarity between the cluster centers: (a) Full dendrogram, where each item is assigned to a separate cluster, (b) Dendrogram where individual clusters are merged into 7 groups, (c) Dendrogram where the previous 7 clusters are merged into 3 groups.

of division/merging based on the dissimilarities between the items. In our framework, we employ agglomerative HC and quantify the dissimilarity between two fiber pathways $X_i = (x_{i,1}, x_{i,2}, \dots)$ and $X_j = (x_{j,1}, x_{j,2}, \dots)$ by the *symmetrized Chamfer distance* (in voxels) [18], which is computed as

$$d(X_i, X_j) = \frac{1}{2}\big(d_{\text{Chamfer}}(X_i, X_j) + d_{\text{Chamfer}}(X_j, X_i)\big), \quad (1)$$

$$\text{with } d_{\text{Chamfer}}(X_j, X_i) = \frac{1}{|X_j|}\sum_{x_{j,n}\in X_j} \min_{x_{i,m}\in X_i} \|x_{j,n} - x_{i,m}\|_2.$$

B. Online Clustering

As described in [15], the online clustering algorithm employed here is designed to work with different baseline methods (e.g., k-centers, affinity propagation) for which the definition of a "cluster center" is required. Since we consider the agglomerative HC method as the baseline with a user-specified number of clusters, and the classical HC formulation is not based on the concept of a cluster center, we explicitly identify the cluster centers using the minimum within-bundle distance criterion, and we initialize the model parameters of the clusters the resulting centers represent.

The model of a cluster (with label l) has four parameters: the cluster center $X_c^{(l)}$, the number of fiber pathways n_l associated with cluster l, the sum of the dissimilarities $D_l = \sum_{i\in\mathcal{F}^{(l)}} d(X_i^{(l)}, X_c^{(l)})$ between all of the $n_l = |\mathcal{F}^{(l)}|$ pathways associated with cluster l; and the sum of the squared dissimilarities $S_l = \sum_{i\in\mathcal{F}^{(l)}} d(X_i^{(l)}, X_c^{(l)})^2$. Following the initialization of these parameters for each cluster, a new fiber pathway, X, is compared to the current cluster centers. If the pathway X is similar to one of the cluster centers, say $X_c^{(l)}$, (i.e., $d(X, X_c^{(l)}) < \xi$ for some threshold $\xi > 0$), then the label l is instantly assigned to X and the parameters n_l, D_l, and S_l are updated accordingly. Otherwise, the pathway is added into a set $\mathcal{R}$ called the *reservoir*.

Conditions to trigger a model update: Since the online clustering of fiber pathways evolve through time, it is anticipated that, in some conditions, the parameters of the cluster models require an update. Here, we define two such conditions. The first condition is the limited size of $\mathcal{R}$, i.e., taking action when the reservoir is full. The second condition is based on an information theoretic approach called the *Page-Hinkley (PH) change point detection test* [19], which can detect a drift in the streaming fibers by considering a sequence of *relevancy information* for these fibers.

The relevancy information measure is computed using a Gaussian distribution model with the first order statistics of the within-cluster dissimilarities. Let μ_l and σ_l be the mean and standard deviation of the fiber dissimilarities within cluster l. The relevancy information between a streaming fiber X and the cluster center $X_c^{(l)}$ is measured, assuming a Gaussian distribution $N(\mu_l, \sigma_l)$, as $f(X, X_c^{(l)}) = \Pr(d(X, X_c^{(l)}), N(\mu_l, \sigma_l))$. For the streaming fiber X analyzed at time $t = 1$ -whether a label is assigned to it or $X \in \mathcal{R}$, the relevancy information is computed as $p_1 = \max_l f(X, X_c^{(l)})$. As time evolves, a sequence p_t is generated and tested for a drift using the PH test. In particular, let $\overline{p}_{1:T} = \frac{1}{T}\sum_{t=1}^{T} p_t$ and $m_{1:T} = \sum_{t=1}^{T}(p_t - \overline{p}_{1:t} + \delta)$ with a tolerance $\delta > 0$. The PH test involves computing $M_T = \max\{|m_{1:t}|; t = 1, 2, \dots, T\}$, and checking the condition $PH_T = (M_T - m_{1:T}) > \lambda$, where $\lambda > 0$. If this condition holds, i.e., the stream is drifting, then a model update is triggered. Note that p_t is initialized at the beginning of the streaming process and reset each time the model is updated.

Model Update: When a model update is triggered (i.e., when $\mathcal{R}$ is full or the PH test detects a drift), the model parameters are updated by applying an additional HC on the union set $\mathcal{H} = \mathcal{R} \cup \mathcal{C}$, where $\mathcal{C} = \{X_c^{(l)}\}_{l=1}^{L}$ is the set of cluster centers before the update. As a result, the new cluster centers $\{\bar{X}_c^{(l)}\}_{l=1}^{L}$ are obtained and the parameters are updated as follows: Let m_l be the number of fibers in $\mathcal{R}$ associated to the updated cluster l. Also let $X_{1,c}^{(l)}, X_{2,c}^{(l)}, \dots, X_{K,c}^{(l)}, K \le L$ denote the cluster centers in $\mathcal{C}$ associated to the updated cluster l, and $n_{k,l}$ be the number of the fibers previously associated to $X_{k,c}^{(l)}$. Then the number of fibers n_l associated to the cluster l is equal to $m_l + \sum_{k=1}^{K} n_{k,l}$. Since the fibers that are previously assigned to a cluster are not considered, D_l and S_l are updated using a random perturbation model, as described in [15]. As a result, the sum of the squared distances for the new cluster l is computed as

$$S_l = \sum_{k=1}^{K}\big(n_{k,l} d(X_{k,c}^{(l)}, \bar{X}_c^{(l)})^2 + S_k\big) + \sum_{i=1}^{m_l} d(X_{i,r}^{(l)}, \bar{X}_c^{(l)})^2, \quad (2)$$

where S_k is the sum of squared distances from $X_{k,c}^{(l)}$ to the fibers associated to the cluster l before the model update, and $\{X_{i,r}^{(l)}\}_i$ is the set of fibers in $\mathcal{R}$ associated to the updated cluster l. Likewise, the sum of the distances is computed as

$$D_l = \sum_{k=1}^{K}\big(n_{k,l} d(X_{k,c}^{(l)}, \bar{X}_c^{(l)}) + D_k\big) + \sum_{i=1}^{m_l} d(X_{i,r}^{(l)}, \bar{X}_c^{(l)}), \quad (3)$$

where D_k is the sum of the distances from $X_{k,c}^{(l)}$ to the fibers in cluster l before the update.

III. Experimental Results

We test our method on two phantom and one real datasets. For HC, we use the hclust function (initialization with average linkage; model update with single linkage) in MATLAB® PRTools library. We heuristically set $\xi = \{4, 6, 10\}$ for `Phantom1`, `Phantom2`, and real data, respectively, and keep other settings at default given in [15].

TABLE I

PERFORMANCE OF THE CLUSTERING METHODS: MEAN AND STANDARD DEVIATION OF THE CLUSTERING ACCURACY AND COMPUTATION TIME.

	Phantom1		Phantom2	
Method\Performance	Acc. (%)	Time (s)	Acc. (%)	Time (s)
k-centers	98.9±6.9	238.9	99.9±0	376.4
k-AP	99.9±0	463.6	99.9±0	551.7
HC	100±0	604.0	100±0	713.9
online k-centers	60.1±8.9	14.4	83.1±8.2	3.2
online k-AP	72.1±11.5	3.5	72.1±8.5	10.4
online HC	100±0	2.69	92.9±9.1	3.2

A. *Experiments on Phantom Data*

The proposed clustering strategy is first tested on the biological phantom in [20], constructed from excised rat spinal cords and designed to have two crossing fiber bundles. The diffusion weighted images of the phantom, herein referred to as `Phantom1`, were acquired with a 40×9 image matrix (40 slices with an isotropic spatial resolution of 2.5 mm) and a diffusion sensitization at $b=1{,}300$ s/mm^2 applied along a set of 90 gradient directions with 10 baseline (b_0) images. Our algorithm is initialized with an homogeneous set of 250 fibers (equal number of fibers from each bundle) and $|\mathcal{R}|$ is set to 50. Fig. 2 shows clustering of 5,000 fiber pathways (2,500 pathways per bundle, computed using [18]) into two bundles through time. On the same data, we also compare our method with alternative techniques such as k-centers and affinity propagation (k-AP, assuming exactly k clusters) by repeating the experiment 100 times with shuffled stream orders. Table I shows the performance of these algorithms and their online versions in terms of accuracy in clustering and computation time. Our first observation is that, as one expects, online clustering is significantly faster than batch clustering because for the batch methods, computation of the full dissimilarity matrix is required before clustering, and this computation takes 238.5 seconds for `Phantom1`. We also observe that online HC outperforms both online k-centers and k-AP in terms of accuracy and computation time, because k-centers and k-AP are more likely to consider the previous cluster centers as outliers at the model update step, whereas HC merges clusters without considering their centers.

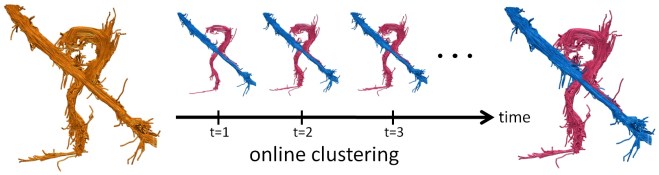


Fig. 2. Online hierarchical clustering of the fiber pathways produced by the method proposed in [18]: Input pathways (left), clustering streams of pathways (middle), and the final result (right).

We conduct a second experiment on the Neurospin MR phantom [21] dataset provided for the MICCAI 2009 Fiber Cup [22]. The diffusion MR data of this phantom (`Phantom2`) were acquired using the following imaging parameters: The image matrix is of size 64×64×3 with isotropic spatial resolution of 3 mm. Two repetitions of 65 images are acquired (1 baseline and 64 diffusion weighted images at $b = 1{,}500$ s/mm^2), with TR/TE = 5,000/94 ms. Fig. 3 shows the fiber pathways selected from four tracts (including three partially overlapping tracts) and their clustering into bundles using online HC. Table I shows the accuracy in clustering and computation time for `Phantom2` (for the batch methods, computation of the full dissimilarity matrix takes 376.3 seconds), for which online HC, similar to the previous experiment, outperforms k-centers and k-AP.

(a) (b) (c)

Fig. 3. (a) Illustration of the Neurospin MR phantom, (b) Extracted fiber pathways from four selected bundles using streamline tractography, (c) Clustering result given by online HC.

B. *Experiments on Real Data*

The proposed algorithm is also evaluated on a human brain MRI dataset provided for the Pittsburgh Brain Connectivity Challenge (Spring 2009). We consider the diffusion weighted images (of the subject `brain0`), which were acquired with a 128×128 image matrix, (68 slices with an isotropic spatial resolution of 2 mm), and a diffusion sensitization at $b=1{,}500$ s/mm^2 applied along a set of 256 gradient directions with 29 baseline images. First, three different WM fiber bundles (the corpus callosum, cingulum, and fornix) in Fig. 4 are reconstructed via streamline tractography [23] initiated from a spherical seed ROI. Then the resulting 2,940 pathways are fed, as streams, into the online HC method. The algorithm is initialized with a homogeneous set of 400 fibers and $|\mathcal{R}|$ is set to 80. Fig. 4 shows the fiber pathways and their successful clustering into three anatomically distinct bundles. Another experiment is performed with a relatively large hand-drawn ROI on an axial slice and 4,262 fiber pathways (Figs. 5(a)-5(b)) passing through the ROI are selected for clustering. We again observe that the final clustering (in Fig. 5(c)) given by online HC is in accordance with the neuroanatomy.

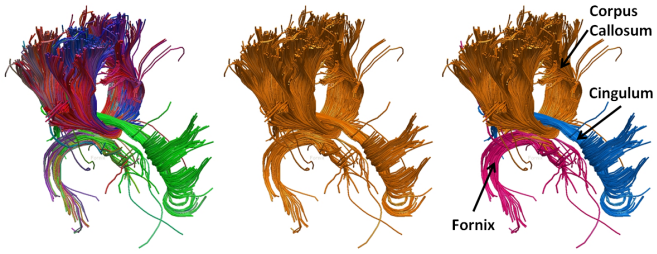


Fig. 4. Online clustering of selected WM fiber pathways (experiment 1): (a-b) Fiber pathways forming the corpus callosum, cingulum, and fornix illustrated using (a) directional coloring, and (b) solid coloring; (c) Clustering of the fiber pathways into three bundles.

Fig. 5. Online clustering of selected WM fiber pathways (experiment 2): (a-b) Fiber pathways forming the corpus callosum, cingulum, internal capsule, superior longitudinal fasciculus, and some U-fibers illustrated using (a) directional coloring, and (b) solid coloring; (c) Clustering of the fiber pathways.

IV. Conclusion and Future Work

We presented an online agglomerative hierarchical clustering method, which yields a framework similar to doing "unified tractography and clustering." We demonstrated the performance of our method through experiments on phantom and real dMRI datasets. Compared to conventional HC, considering sets of fiber pathways as data streams reduces the overall computation time, because a relatively small pairwise dissimilarity matrix is computed/analyzed at initialization and only when the reservoir is full, and the cluster label of an item is assigned at the time of its creation. Future work will focus on online clustering of large-scale data (e.g., whole-brain pathways), and of fibers produced by probabilistic tractography (e.g., see [24]) to generate WM atlases.

References

[1] C. Lenglet, J. Campbell, M. Descoteaux, G. Haro, P. Savadjiev, D. Wassermann, A. Anwander, R. Deriche, G. Pike, G. Sapiro, K. Siddiqi, and P. Thompson, "Mathematical methods for diffusion MRI processing," *NeuroImage*, vol. 45, pp. S111–S122, 2009.

[2] D. Jones, Ed., *Diffusion MRI: Theory, methods, and applications*. Oxford University Press, 2011.

[3] P. J. Basser, "Inferring microstructural features and physiological state of tissues from diffusion weighted images," *NMR in Biomedicine*, vol. 8, pp. 333–344, 1995.

[4] D. Tuch, "Q-ball imaging," *Magnetic Resonance in Medicine*, vol. 52, no. 6, pp. 1358–1372, 2004.

[5] I. Aganj, C. Lenglet, G. Sapiro, E. Yacoub, K. Ugurbil, and N. Harel, "Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle," *Magnetic Resonance in Medicine*, vol. 64, no. 2, pp. 554–566, 2010.

[6] M. Descoteaux, R. Deriche, T. R. Knösche, and A. Anwander, "Deterministic and probabilistic tractography based on complex fibre orientation distributions." *IEEE Trans. Med. Imaging*, vol. 28, no. 2, pp. 269–86, Feb. 2009.

[7] S. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. Nichols, C. Mackay, K. Watkins, O. Ciccarelli, M. Cader, P. Matthews, and T. Behrens, "Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data," *NeuroImage*, vol. 31, no. 4, pp. 1487–1505, 2006.

[8] P. Guevara, C. Poupon, D. Rivière, Y. Cointepas, M. Descoteaux, B. Thirion, and J.-F. Mangin, "Robust clustering of massive tractography datasets," *NeuroImage*, vol. 54, no. 3, pp. 1975 – 1993, 2011.

[9] H. Li, Z. Xue, L. Guo, T. Liu, J. Hunter, and S. Wong, "A hybrid approach to automatic clustering of white matter fibers," *NeuroImage*, vol. 49, no. 2, pp. 1249 – 1258, 2010.

[10] L. O'Donnell and C.-F. Westin, "Automatic tractography segmentation using a high-dimensional white matter atlas," *IEEE Trans. Med. Imaging*, vol. 26, pp. 1562–1575, 2007.

[11] A. Tsai, C.-F. Westin, A. Hero, and A. Willsky, "Fiber tract clustering on manifolds with dual rooted-graphs," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–6.

[12] A. Leemans and D. Jones, "A new approach to fully automated fiber tract clustering using affinity propagatoin," in *Int. Symp. Magn. Reson. Med.*, 2009, p. 856.

[13] D. Wassermann, L. Bloy, E. Kanterakis, R. Verma, and R. Deriche, "Unsupervised white matter fiber clustering and tract probability map generation: Applications of a Gaussian process framework for white matter fibers," *NeuroImage*, vol. 51, pp. 228–241, 2010.

[14] M. Liu, B. Vemuri, S. Amari, and F. Nielsen, "Shape retrieval using hierarchical total Bregman soft clustering," *IEEE Trans. on Pattern Anal. and Machine Intelligence*, vol. PP, no. 99, p. 1, 2012.

[15] X. Zhang, C. Furtlehner, and M. Sebag, "Data streaming with affinity propagation," in *Proc. of the European Conf. on Machine Learning and Knowledge Discovery in Databases - Part II*, 2008, pp. 628–643.

[16] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, 2012.

[17] J. H. Ward, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[18] H. Çetingül, M. Nadar, P. Thompson, G. Sapiro, and C. Lenglet, "Simultaneous ODF estimation and tractography in HARDI," in *Int. Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 86–89.

[19] D. Hinkley, "Inference about the change point from cumulative sum-tests," *Biometrika*, vol. 3, no. 58, pp. 509–523, 1970.

[20] J. S. Campbell, K. Siddiqi, V. V. Rymar, A. F. Sadikot, and G. B. Pike, "Flow-based fiber tracking with diffusion tensor and q-ball data: Validation and comparison to principal diffusion direction techniques," *NeuroImage*, vol. 27, no. 4, pp. 725–736, 2005.

[21] C. Poupon, B. Rieul, I. Kezele, M. Perrin, F. Poupon, and J.-F. Mangin, "New diffusion phantoms dedicated to the study and validation of high-angular-resolution diffusion imaging (HARDI) models," *Magnetic Resonance in Medicine*, vol. 60, no. 6, pp. 1276–1283, 2008.

[22] P. Fillard, M. Descoteaux, A. Goh, S. Gouttard, B. Jeurissen, J. Malcolm, A. Ramirez-Manzanares, M. Reisert, K. Sakaike, F. Tensaouti, T. Yo, J.-F. Mangin, and C. Poupon, "Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom," *NeuroImage*, vol. 56, pp. 220–234, 2011.

[23] S. Mori and P. van Zijl, "Fiber tracking: principles and strategies - a technical review," *NMR in Biomedicine*, vol. 15, pp. 468–480, 2002.

[24] H. Çetingül, L. Dumont, M. Nadar, P. Thompson, G. Sapiro, and C. Lenglet, "Simultaneous ODF estimation and robust probabilistic tractography from HARDI," in *MICCAI 2012 Workshop on Computational Diffusion MRI*, 2012, pp. 13–24.