HHS Public Access

Author manuscript

J Neurosci Methods. Author manuscript; available in PMC 2017 November 13.

Published in final edited form as:

J Neurosci Methods. 2016 September 01; 270: 156–164. doi:10.1016/j.jneumeth.2016.04.024.

Two step Gaussian mixture model approach to characterize white matter disease based on distributional changes

Namhee Kim[a,*], **Moonseong Heo**[b], **Roman Fleysher**[a], **Craig A. Branch**[a,c], and **Michael L Lipton**[a,d,e,f,*]

[a]The Gruss Magnetic Resonance Research Center, Radiology, The Albert Einstein College of Medicine, Bronx, NY, USA

[b]Department of Epidemiology and Population Health, The Albert Einstein College of Medicine, Bronx, NY, USA

[c]Department of Physiology and Biophysics, The Albert Einstein College of Medicine, Bronx, NY, USA

[d]Department of Psychiatry and Behavioral Sciences, The Albert Einstein College of Medicine, Bronx, NY, USA

[e]The Dominick P Purpura Department of Neuroscience, The Albert Einstein College of Medicine, Bronx, NY, USA

[f]Department of Radiology, The Montefiore Medical Center, Bronx, NY, USA

Abstract

Background—Magnetic resonance imaging reveals macro- and microstructural correlates of neurodegeneration, which are often assessed using voxel-by-voxel *t*-tests for comparing mean image intensities measured by fractional anisotropy (FA) between cases and controls or regression analysis for associating mean intensity with putative risk factors. This analytic strategy focusing on mean intensity in individual voxels, however, fails to account for change in distribution of image intensities due to disease.

New method—We propose a method that aims to facilitate simple and clear characterization of underlying distribution. Our method consists of two steps: subject-level (Step 1) and group-level or a specific risk-level density function estimation across subjects (Step 2).

Results—The proposed method was demonstrated with a simulated data set and real FA data sets from two white matter tracts, where the proposed method successfully detected any departure of the FA distribution from the normal state by disease: $p < 0.001$ for simulated data; $p = 0.047$ for the posterior limb of internal capsule; $p = 0.06$ for the posterior thalamic radiation.

Comparison with existing method(s)—The proposed method found significant disease effect ($p < 0.001$) while conventional 2-group *t*-test focused only on mean intensity did not ($p = 0.61$) in a simulation study. While significant age effects were found for each white matter tract

*Corresponding authors at: The Gruss Magnetic Resonance Research Center, Radiology, The Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA. Fax: +1 718 430 3399. Namhee.Kim@einstein.yu.edu (N. Kim), Michael.Lipton@einstein.yu.edu (M.L. Lipton).

from conventional linear model analysis with real FA data, the proposed method further confirmed that aging also triggers distribution-wide change.

Conclusion—Our proposed method is powerful for detection of risk factors associated with any type of microstructural neurodegenerations with brain imaging data.

Keywords

Gaussian mixture model; Density function estimation; Aging effects; Fractional anisotropy; Diffusion tensor imaging

1. Introduction

Voxel-by-voxel *t*-test is the most common analytic approach used to compare images between two groups for assessment of the impact of neurodegenerative diseases on brain structure (Thiebaut de Schotten et al., 2011; Muller et al., 2009; Lipton et al., 2012; Kim et al., 2013). Voxel-wise analyses in general, however, are inherently based on a strong assumption that disease effects are homogenous across subjects. However, this homogeneity assumption is often violated as reported in many studies (Thiebaut de Schotten et al., 2011; Muller et al., 2009; Lipton et al., 2012; Kim et al., 2013; Kou et al., 2010; Rosenbaum and Lipton, 2012; Yan et al., 2013; Silverman, 2009; Kochunov et al., 2007; Walhovd et al., 2005; Benson et al., 2007). White matter (WM) tissue in the brain is highly susceptible to aging and heterogeneous neurodegenerative disease pathology such as Alzheimer disease and traumatic brain injury (TBI) (Thiebaut de Schotten et al., 2011; Muller et al., 2009; Lipton et al., 2012; Kou et al., 2010; Rosenbaum and Lipton, 2012; Yan et al., 2013; Silverman, 2009). These neurodegenerative diseases cause WM axonal damage with consequent disruption of cerebral connectivity leading to cognitive dysfunction (Silverman, 2009; Kochunov et al., 2007; Walhovd et al., 2005). WM microstructural damage can be assessed by measuring fractional anisotropy (FA) obtained from diffusion tensor imaging (DTI). WM axonal damage typically reduces the directional coherence of water diffusivity, manifesting as lower FA. However, the spatial variation of WM pathology among subjects is high (Thiebaut de Schotten et al., 2011; Muller et al., 2009; Lipton et al., 2012; Kim et al., 2013), and thus voxel-by-voxel statistical analysis of FA images may not be appropriate in the assessment of neurodegenerative diseases.

A few studies have addressed such limitation of voxel-wise analysis by instead adopting a histogram approach to characterize and compare shapes of WM FA distributions between patients and healthy subjects (Benson et al., 2007; Lipton et al., 2008; Nave R et al., 2007). However, these comparisons have been limited to two-sample *t*-tests comparing patients' and controls' estimated moment summary statistics, such as mean, variance, skewness, and kurtosis of individual subjects' histograms of brain-wide image distributions (Benson et al., 2007). While this approach better accounts for spatial heterogeneity of disease pathology across subjects than voxel-by-voxel group comparison, it is not easy to summarize the four moments into a salient conclusion concerning group differences or their association with risk factors.

Statistical methods for characterization of distribution by putative risk factors have been studied in general data setting in which one or a few measurements per subject are observed, and published in R packages, GAMLSS, and GAMLSS-MX by Stasinopoulos and Rigby (2007). While GAMLSS-MX adopts nonparametric density function estimation based on mixture models, GAMLSS adopts parametric approach. To our knowledge, however, their methods have not been studied for application to brain imaging data whose structure is highly unique in terms of the number of measurements observed per subject. In what follows, a need for customizing GAMLSS or GAMLSS-MX method according to such a unique structure of brain imaging data is discussed.

In high-resolution images, the number of voxels in each segmented WM brain atlas well exceeds 10^3 and reaches 10^6 if the whole volume of the brain is of interest. Such a large number of measurements per subject provide opportunities for estimating subject-level density function while it tackles application of statistical methods developed in general settings. Given this order of magnitude in imaging data, we structure our proposed method in two steps: (1) subject-level density estimation and (2) group-level or specific risk-level density function. It is noteworthy that the GAMLSS or GAMLSS-MX method (Stasinopoulos and Rigby, 2007) has no separation of these two steps. We in this study prefer nonparametric density function estimation due to its flexible adaptation, and tailor the existing GAMLSS-MX to fit the brain imaging data. We utilize the Gaussian mixture model (GMM) approach for Step 1 because GMM can flexibly characterize unusual distributions as a mixture of multiple Gaussian distributions (Molas and Lesaffre, 2012) with mixing probabilities. We recently proposed a method (Kim et al., 2014) that differentiates compositions of mixing probabilities to latent Gaussian densities for subject-level density functions. This method *indirectly* estimates subject-specific mixing probabilities *a posteriori* based on a mixture of Gaussian densities estimated from that likelihood function of the pooled data. In the present study, we extend and enhance the *indirect* estimation approach to one that *directly* incorporates subject-level mixing probabilities into the likelihood function. We call this method "*direct* estimation" as opposed to the *indirect* estimation described previously. Once subject-level density functions are estimated at Step 1, we use a multinomial logistic regression model with subject-wise pseudo multinomial responses (Stasinopoulos and Rigby, 2007), generated using the GMM results in Step 1, to estimate density function at each risk level across subjects for Step 2. A diagram showing the flow of the proposed method is shown in Fig. 1. Finally, we characterize four moment statistics, mean, variance, skew, and kurtosis, by using mixing probabilities, each of which was estimated as a function of putative risk factors. These four moment statistics are typically adopted for describing a distribution. Let us call the method proposed here "*distribution-based disease trait detection method*".

We introduce the proposed method in Section 2. We demonstrate the proposed estimation method with a simulated data set in Section 3, where a linear model focused on mean FA fails to detect disease trait. Demonstration with real FA image data is followed in Section 4, where the effects of aging on FA in two WM tracts are examined. Finally, in Section 5, we discuss the proposed methods in terms of their strengths and weaknesses.

2. Methods

2.1. Diffusion tensor image acquisition and preprocessing

Imaging was performed using a 3.0 T MRI scanner (Achieva; Philips Medical Systems, Best, The Netherlands) with an eight-channel head coil (Sense Head Coil; Philips Medical Systems). T1-weighted whole-head structural imaging was performed using sagittal three-dimensional magnetization-prepared rapid acquisition gradient echo (MP-RAGE; TR/TE/TI = 9.9/4.6/1250 ms; field of view, 240 mm^2; matrix, 240 × 240; and section thickness, 1 mm). T2-weighted whole-head imaging was performed using axial two-dimensional turbo spin-echo (TR/TE = 4000/100 ms; field of view, 240 mm^2; matrix, 384 × 512; and section thickness, 2 mm). DTI was performed using single-shot echo-planar imaging (TR/TE = 11,000/51 ms; field of view, 240 mm^2; matrix, 120 × 120; section thickness, 2 mm; independent diffusion sensitizing directions, 32; and $b = 800 s/mm^2$).

FA was derived from DTI at each voxel using the FMRIB Diffusion Toolbox (Smith et al., 2007). Preprocessing procedures implemented for DTI included skull stripping, echo-planar imaging distortion correction, intermediate rigid-body registration, registration to standard space, transformation of DTI to standard space, and white matter segmentation, in sequence. Non-brain voxels were removed from the MP-RAGE and turbo spin-echo images using FMRIB-FSL software (Smith et al., 2004). Each brain volume was inspected section-by-section, and residual non-brain voxels were removed manually. Turbo spin-echo images were acquired with the same section thickness, position and orientation as DTI. Distortion correction was accomplished using a nonlinear deformation algorithm to match each echo-planar image to the corresponding turbo spin-echo volumes (Lim et al., 2006). For intermediate rigid-body registration, each subject's turbo spin-echo images were registered to their three-dimensional MP-RAGE volume using the Automated Registration Toolbox (ART) (Ardekani, 1995) three-dimensional rigid-body approach (Ardekani et al., 2005). For registration to standard space, the nonlinear registration module in ART was used to register each subject's three-dimensional MP-RAGE volume to a standard T1-weighted template, the Johns Hopkins University (JHU) T1 (JHU-MNI-SS-T1) (Holmes et al., 1998; Oishi et al., 2009). For transformation of DTI to the standard space, distortion correction, intermediate rigid-body registration, and standard space registration were applied to the calculated FA maps in a single resectioning operation using ART. Final cubic voxel size was 1 mm^3, masked to exclude non-brain voxels from the analysis. For white matter segmentation, the fast automated segmentation tool in the FMRIB-FSL package (Smith et al., 2004) was used to generate a white matter mask for the three-dimensional MP-RAGE template brain images and restrict subsequent statistical analysis of FA to white matter voxels. A JHU-MNI-SS atlas (JHU-MNI-SS-WMPM-Type-II) with comprehensive WM parcellation was used to extract WM anatomical information (Oishi et al., 2009).

2.2. A limitation in the current stream of FA image data analysis

We here discuss a limitation of current FA brain image data analysis. A typical statistical model for relating subject-wise mean intensity at each voxel with putative risk factors can be written as follows.

$$y_i = \sum_{q=1}^{Q} \beta_q x_{iq} + e_i, \quad (1)$$

where x_{iq} is the q-th risk factor for the i-th subject, β_q is the coefficient, and $e_i \sim N(0, \sigma_e^2)$. In Model (1), the mean change of measurements (y_i's) by risk factor levels, which is $E(y_i) = \sum_{q=1}^{Q} \beta_q x_{iq}$, is of interest. Extension of such arguments can be similarly made with variance, skewness, and kurtosis, which are second, third, and fourth moment statistics from y_i's. Such models, associating higher moments (up to fourth order) statistics with putative risk factors, have been studied and proposed in the GAMLSS-MX, R package. While their methods have been applied to a general data structure which consists of one or a few measurements per subject, those have not been applied to brain imaging data. It is also known that implementing iterative algorithms, e.g. GAMLSS-MX, across individual voxels in the brain is highly exhaustive approach in terms of required memory space and computing time.

In neuroimaging studies, characterization of a region of interest (ROI) as a function of risk factors is more frequently applied in lieu of voxel-wise analysis to avoid repeating countless univariate analyses. We thus propose a method by modifying the GAMLSS-MX that enables estimating association between risk factors and higher order moment statistics from ROI-based brain imaging data. It is noteworthy that a typical size of ROI is sometimes exceedingly large (>10000) and approaching to the size of the whole brain. As such, the proposed method is structured in two steps: (1) subject-level density estimation and (2) group-level or specific risk-level density function. Of note, the proposed method examines how FA distribution from each ROI changes by each risk factor level and enables researchers to answer such questions as "*Does variance of the FA measurements in a ROI become smaller or larger in older ages?*" and "*Does mean of the FA measurements in a ROI become smaller in older ages*?"

2.3. Proposed method

2.3.1. Voxel-wise normalization—We note that a distribution from pooling of all the voxels in a large scaled ROI can be highly irregular and complicated by presence of multi-modes and/or highly skewed tails. A part of such complexity is due to heterogeneous mean and variance at each voxel across subjects. Such a complexity can cause longer iteration or failure of convergence, and necessitates a much larger number of Gaussian densities required for approximating to the true underlying density function with the Expectation-Maximization (EM) algorithm which will be introduced in the following Section. To reduce complexity and enhance computation efficiency, we introduce a voxel-wise normalization step as below.

$$z_{ij}=\frac{y_{ij}-\overline{y}_{\cdot j}}{s_{\cdot j}}, \quad (2)$$

where $\overline{y}_{\cdot j}$ is the mean intensity and $s_{\cdot j}$ is the SD of subjects at j-th voxel. The efficacy of the normalization in Eq. (2) is demonstrated with simulated data in Section 3. We further discuss in Supplementary material 1 that the proposed normalization in Eq. (2) maintains two higher order (3rd and 4th) moment statistics after pooling which could be altered otherwise.

2.3.2. Step 1: estimation of subject-specific density functions—Our proposed *direct estimation* method incorporates specification of subject-level densities directly to the likelihood function through parameterization of subject-specific mixing probabilities. Unlike the previous *indirect method* (Kim et al., 2014), subject-specific mixing probabilities assigned to underlying Gaussian densities are estimated specifically for individual subjects while mean and variance for each latent Gaussian component are estimated from pooled data from all subjects. Differences in FA distribution between subjects are therefore characterized by differences in mixing probabilities assigned to underlying Gaussian densities. When compared to the direct method, the indirect method fails to fully specify likelihood function of FA data because it does not incorporate parameters necessary for specification of individual subject characteristics. It is also noteworthy that, although we can estimate subject-specific density functions by Kernel Density Estimation (KDE) (Gu, 2002), these approaches do not provide parameters for comparison between subjects as discussed in (Kim et al., 2014). Details of the results are shown in Supplementary material 2.

In contrast, for direct method, the subject-level mixing probability for the i-th subject is denoted by $\boldsymbol{\lambda}_i = (\lambda_{i,1},\ldots, \lambda_{i,\mathrm{m}})$ for m Gaussian densities φ_k ($k = 1,\ldots,m$) with mean μ_k and a common variance σ^2. We assume this equal variances σ^2 of all m Gaussian densities since this assumption allows relatively easy interpretation of differences between density functions as detailed in (Kim et al., 2014). The entire configuration of the mixing probabilities of all n subjects is denoted by $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1,\ldots, \boldsymbol{\lambda}_n)$. The sum of the mixing probabilities for each subject constrained to be 1, i.e. $\sum_{k=1}^{m}\lambda_{i,k}=1$ for all i. This parameterization yields a Gaussian mixture for the i-th subject,

$$f_i(z_{ij};\boldsymbol{\lambda}_i,\boldsymbol{\theta})=\sum_{\mathbf{k=1}}^{\mathbf{m}}\lambda_{\mathbf{i,k}}\varphi_{\mathbf{k}}(\mathbf{z_{ij}};\mu_{\mathbf{k}},\sigma^{\mathbf{2}}), \quad (3)$$

where $\boldsymbol{\theta} = (\mu_k, \sigma/k = 1,\ldots, m)$ is common for all subjects. The likelihood function for all FA data from all subjects, based on (3), is expressed as follows:

$$L_{\min}(\mathbf{z};\boldsymbol{\Lambda},\boldsymbol{\theta})=\prod_{i=1}^{n}\prod_{j=1}^{N}f_i(z_{ij};\boldsymbol{\lambda}_i,\boldsymbol{\theta})=\prod_{i=1}^{n}\prod_{j=1}^{N}\left[\sum_{k=1}^{m}\lambda_{i,k}\varphi_k(z_{ij})\right]. \quad (4)$$

The number of parameters to be estimated in the Gaussian mixture (3) is $dim(\boldsymbol{\theta}) + dim(\Lambda) = (n + 1) \times (m - 1) + 2$. To estimate every element of Λ and $\boldsymbol{\theta}$, we used the EM algorithm (Table 1) for the likelihood function in (4). Determination of optimal m is based on Akaike Information Criterion (AIC).

2.3.3. Step 2: estimation of density functions across subjects at risk factors—

Estimated subject-level mixing probabilities for m-Gaussian densities obtained from Step 1 are then utilized for Step 2 between-subjects analysis which estimates a specific risk-level density function identified with mixing probabilities at the specific risk-level. In this step, each Gaussian density is treated as an individual class. A multinomial logistic regression model with pseudo multinomial responses adopted in GAMLSS-MX (Stasinopoulos and Rigby, 2007) was utilized, where each pseudo multinomial response (y_{ik}) for each subject is a class indicator to the k-th Gaussian density (i=1,…,n; k = 1,…,m). Each pseudo multinomial response generated for each subject is then weighted by subject-level mixing probabilities ($\lambda_{i,k}$) estimated from Model (1), where

$$\sum_{i=1}^{n}\sum_{k=1}^{m}\lambda_{i,k}=n.$$

We order the estimated Gaussian densities, denoted by G_k, by their means in ascending manner so that G_1 and G_m represent the Gaussian densities with the lowest and the greatest means, respectively. Specifically, the multinomial log odds of the k-th density (k = 1,…,m) over the lowest Gaussian density (the reference) is given as below.

$$\log\left(\frac{Pr(y=k|\mathbf{x})}{Pr(y=1|\mathbf{x})}\right)=\boldsymbol{\beta}_{\boldsymbol{k}}'\mathbf{x}=\beta_{0,k}+\beta_{1,k}x_1+\cdots+\beta_{Q,k}x_Q, \quad (5)$$

where $\boldsymbol{x}$ is a covariate vector of Q-risk factors and $\boldsymbol{\beta}_1 = \mathbf{0}$. The probability of a pseudo response (y) belonging to the k-th Gaussian density with a risk factor level ($\mathbf{x}$) is denoted as $p_k(\mathbf{x})$[= $Pr(y = k|\mathbf{x})$] and calculated as

$$p_k(\mathbf{x})=\frac{\exp\left(\boldsymbol{\beta}_{\boldsymbol{k}}'\mathbf{x}\right)}{\sum_{l=1}^{m}\exp\left(\boldsymbol{\beta}_{\boldsymbol{l}}'\mathbf{x}\right)}. \quad (6)$$

Accordingly, the likelihood function from the multinomial logistic regression model with weights is given as below.

$$L_{\text{multi}}(\mathbf{x};\boldsymbol{\Lambda},\boldsymbol{\beta})=\prod_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{p_1}(\mathbf{x_i})^{\lambda_{\mathbf{i,1}}}\times\cdots\times\mathbf{p_m}(\mathbf{x_i})^{\lambda_{\mathbf{i,m}}}. \quad (7)$$

Plugging in $\hat{\beta}$ estimated from (7) to (6) estimates $p_k(\mathbf{x})$ which is called mixing probability for the k-th Gaussian density at risk level $\boldsymbol{x}$. Density function at risk level x can be written as

$$g(z{:}\theta, p_1(\mathbf{x}), \cdots, p_m(\mathbf{x})) = \sum_{k=1}^{m} p_k(\mathbf{x}) \varphi(z; \mu_k, \sigma^2). \tag{8}$$

2.4. Four moment statistics as function of risk factors

We provide an additional procedure relating the results from Step 1 & 2, which characterizes density function at a specific risk-level **x**, to moment-based summary statistics, mean, variance, skew, kurtosis, which are conventionally adopted for characterizing density function. Specific calculation for mean (α), variance (ν^2), skew (sk), and kurtosis (kr) at a risk level **x** with the density functions ($\varphi_1(z; \mu_1, \sigma), \ldots, \varphi_m(z; \mu_m, \sigma)$, $p_1(\mathbf{x}), \ldots, p_m(\mathbf{x})$) estimated from Step 1 & 2, is shown as below:

$$\alpha(\boldsymbol{x}) = \int_{-\infty}^{\infty} z \sum_{k=1}^{m} p_k(\boldsymbol{x}) \varphi_k(z) \mathrm{dz} = \sum_{k=1}^{m} p_k(\boldsymbol{x}) \mu_{k'} \tag{9}$$

$$\nu^2(\boldsymbol{x}) = \int_{-\infty}^{\infty} (z - \alpha(\boldsymbol{x}))^2 \sum_{k=1}^{m} p_k(\boldsymbol{x}) \varphi_k(z) \mathrm{dz} = \sigma^2 + \sum_{k=1}^{m} p_k(\boldsymbol{x}) (\mu_k - \alpha(x))^2, \tag{10}$$

$$sk(x) = \int_{-\infty}^{\infty} \left(\frac{(z - \alpha(\boldsymbol{x})}{\nu(\boldsymbol{x})} \right)^3 \sum_{k=1}^{m} p_k(x) \varphi_k(z) \mathrm{dz} = \frac{1}{\nu^2(x)} \left(\sum_{k=1}^{m} p_k(x) \left(\mu_k^3 - \alpha^3(x) \right) - 3\alpha(x) \left(\nu^2(x) - \sigma^2 \right) \right), \tag{11}$$

$$kr(x) = \frac{1}{\nu^4(x)} \int_{-\infty}^{\infty} ((z - \alpha(\boldsymbol{x}))^4 \sum_{k=1}^{m} p_k(x) \varphi_k(z) \mathrm{dz} - 3 = \frac{1}{\nu^4(x)} \left(\sum_{k=1}^{m} p_k(x) (\mu_k - \alpha(\boldsymbol{x}))^4 + 3\sigma^2 (2\nu^2(x) - \sigma^2) \right) - 3, \tag{12}$$

where **x** is a vector of risk factors, and z is realization from the Gaussian mixture model. Mean and variance of the Gaussian mixture densities ($\varphi_1, \ldots, \varphi_m$) are denoted as (($\mu_1, \ldots, \mu_m$), σ^2). Mixing probabilities at a risk level **x** estimated from (6) are denoted as ($p_1(\mathbf{x})$, $\ldots, p_m(\mathbf{x})$).

3. Simulation

We simulate a dataset to demonstrate the proposed method, where disease traits are characterized in terms of variance, and thus current analytic practices focusing on seeking relationship of mean intensity to risk factors cannot detect such disease traits. In this simulation, the primary risk factor (x_i) is binary group indicator, exposed or control group. True underlying distribution for each group is parameterized by differing combination of mixing probabilities of a GMM shown as below.

$$f_g(z{:}\mu_1'\mu_2'\mu_3'\sigma^2, \lambda_{g,1}, \lambda_{g,2}, \lambda_{g,3}) = \lambda_{g,1}\varphi_1(z{:}\mu_1'\sigma^2) + \lambda_{g,2}\varphi_2(z{:}\mu_2'\sigma^2) + \lambda_{g,3}\varphi_3(z{:}\mu_3'\sigma^2),$$

where common parameters ($\boldsymbol{\theta}$) in the GMM are set as $\mu_1 = -1$, $\mu_2 = 0$, $\mu_3 = 1$, and $\sigma = 1$. Mixing probabilities for the exposed group (g=1) are set as $\lambda_{1,1} = \lambda_{1,3} = 0.4$, and $\lambda_{1,2} = 0.2$ while those for the control group (g = 0) are as $\lambda_{0,1} = \lambda_{0,3} = 0.2$, and $\lambda_{0,2} = 0.6$. In this example, the distribution of the exposed group has larger variance (1.8 vs. 1.4) and smaller kurtosis (−0.34 vs. −0.04) while mean and skewness stay the same across two groups (mean = 0; skewness = 0). The number of subjects for each group was 20 ($i = 1,\ldots,n$; $n = 40$) and the number of measurements per subjects, which represents individual voxels for this study, was 100 ($j = 1,\ldots,N$; $N = 100$). Measurements from a subject-wise density function were simulated through two steps, which is specification of random subject and voxel effect. In the first step for specification of random subject effects, we add random jitters to the mixing probabilities specified for each group, where $s_{i,k} \sim N(0, \sigma^2_{s,k})$ with $\sigma^2_{s,k} = 0.05$, and $\sum_{k=1}^{3} s_{i,k} = 0$ ($k = 1$–3) for all i. Individual density function with random jitters is determined as below.

$$f_i(z{:}\mu_1'\mu_2'\mu_3'\sigma^2, \lambda_{i,1}, \lambda_{i,2}, \lambda_{i,3}) = \lambda_{i,1}\varphi_1(z{:}\mu_1'\sigma^2) + \lambda_{i,2}\varphi_2(z{:}\mu_2'\sigma^2) + \lambda_{i,3}\varphi_3(z{:}\mu_3'\sigma^2), \quad (13)$$

where $\lambda_{i,k} = \lambda_{g,k} + s_{i,k}$. In the second step, individual voxel effects, parameterized with two parameters of heteroscedastic mean and SD at each voxel (j) denoted as $(\nu_j^m, \sqrt{\nu_j^s})$, were added to each measurement (z_{ij}) from the individual density functions in (13) as below.

$$y_{ij} = z_{ij}\sqrt{\nu_j^s} + \nu_j^m. \quad (14)$$

The mean and variance of measurements at each voxel (j) is then $E(y_{ij}) = \sqrt{\nu_j^s}\kappa_1 + \nu_j^m$ and $\mathrm{Var}(y_{ij}) = \nu_j^s\kappa_2$ from the equation (14), where K_1 and K_2 are mean and variance of z_{ij} across subjects (i) and denoted as $K_1 = \mathrm{E}(z_{ij})$ and $K_2 = \mathrm{Var}(z_{ij})$, respectively. The proposed normalization procedure was applied and followed by the two subsequent procedures of estimating density functions for individuals and groups. In Fig. 2, two histograms with raw and normalized data from all the subjects of each group are illustrated, where normalized data provides better contrast between two groups. Three Gaussian components were selected based on AIC. We order the estimated Gaussian densities, denoted by G_k, by their means in

ascending manner so that G_1 and G_3 represent the Gaussian densities with the lowest and the greatest means, respectively. The subsequent multinomial logistic regression model takes the G_2 as reference and models multinomial log-odds for each of the other two Gaussians over the referent. While group effect was not significant with the mean of measurements from 2-group *t*-test ($p = 0.61$), significance of group effect was found from the proposed approach for G_1 over G_2 (beta= 1.72; $p = 4.0 \times 10^{-4}$) and G_3 over G_2 (beta= 1.96; $p = 7.0 \times 10^{-5}$).

The estimated mixing probability $(\hat{\lambda}_{g,1}, \hat{\lambda}_{g,2}, \hat{\lambda}_{g,3})$ was (0.39, 0.18, 0.42) for the exposed group and (0.22, 0.59, 0.19) for the control group while estimated common parameters in the GMM were $\hat{\theta}' = (\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\sigma}) = (-1.045,\ -0.106,\ 0.872,\ 0.706)$.

4. Real FA data analysis

4.1. Ethics Statement

After Albert Einstein College of Medicine Institutional Review Board (IRB) approval, Health Insurance Portability and Accountability Act (HIPAA) compliance and written informed consent, subjects were prospectively enrolled between August 2006 and May 2010 through advertisements to examine effect of aging on brain and to demonstrate the proposed approach.

4.2. Demonstration of the aging effects

Forty-nine normal subjects were included: 25 males and 24 females, with ages ranging from 20 to 60 years, years of education ranging from 7 to 25 years, and no history of medical, neurological or psychiatric disease.

We use the proposed method to demonstrate heterogeneous aging mechanism by difference in the brain development phase with two WM tracts, the posterior limb of internal capsule and the posterior thalamic radiation, where maturity of the former is followed by the latter. Two tracts were delineated by a JHU-MNI-SS atlas (JHU-MNI-SS-WMPM-Type-II) with comprehensive WM parcellation and each of which consist of 6592 and 12952 voxels, respectively. Three Gaussian densities ($k = 3$) were determined as optimal for the proposed direct method based on AIC criterion for the two WM tracts. Three putative risk factors that we examined were: age, gender, and years of education.

Multinomial logistic regression was applied for estimating mixing probabilities across subjects with three risk-factors, age, gender, and years of education, and the interaction terms. The multinomial logistic regression model takes the G_1 as reference and models multinomial log-odds for each of the other two Gaussians over the referent. As a result, we found a significant age effect (beta = −0.085; *p*-value = 0.047) in the multinomial log-odds of G_3 over G_1 for the posterior thalamic radiation while a borderline significance (beta= −0.057; *p*-value = 0.061) was found in the multinomial log-odds of G_2 over G_1 for the posterior limb of internal capsule. We note that no significant effects of gender, education or interaction effects were found: *p*-values = 0.98, 0.70, 0.67, 0.75 respectively for gender, education, gender × age, gender × education with the posterior limb of internal capsule; *p*-values = 0.85, 0.63, 0.75, 0.69 with the posterior thalamic radiation, where the reported *p*-value was the minimum taken from two logit analyses, G_2 over G_1 and G_3 over G_1 for each

covariate. Significantly increased mixing probability of G_1 provides much further information on density shape by aging beyond decreased mean FA as detailed below.

Estimated mixing probabilities (p_1, p_2, p_3) against age from Eq. (6) are shown in Fig. 3 for each WM tract, the posterior limb of internal capsule (2a) and the posterior thalamic radiation (2b), where p_1, p_2, and p_3 are represented by solid gray, dashed black, and dash-dot black, respectively. It is observed that the mixing probability to the lowest Gaussian density, p1, increases rapidly while either p_2 or p_3 declines from both WM tracts. Specifically, two mixing probabilities dominating age-related change are ($p_1\uparrow$, $p_2\downarrow$) for the posterior limb of internal capsule and ($p_1\uparrow$, $p_3\downarrow$) for the posterior thalamic radiation, respectively while the remainder shows very subtle change over age. These results suggest that distributions of FA are characterized in an aging-related pattern; mean becomes smaller and positive skewness becomes larger as age increases. We also found in the posterior limb of internal capsule that two dominating Gaussian densities attain almost evenly divided weights ($p_1 \approx p_2$) progressively as age approaches to sixty while p_2 is close to one at age of twenty ($p_1 \ll p_2$). This pattern implies higher variance in the FA distribution for older ages while such a change in variance is much smaller in the posterior thalamic radiation.

A centile curve at a γ is a plot of $z_\gamma(x)$ against a covariate x, where $z_\gamma(x)$ is $\gamma \times 100\%$ percentile from the density function estimated at a risk-level x. Centile curve was adopted to visualize distributional shape change by varying x values. $z_\gamma(x)$ from a Gaussian mixture density can be calculated as below by using `nor1Mix` (R package).

$$\gamma = \int_{-\infty}^{z\gamma} \sum_{k=1}^{m} p_k(x)\varphi_k(z)dz \qquad (15)$$

Fig. 4 shows seven centile curves for the mixture Gaussian density estimated from Eq. (15), with percentiles $\gamma = 0.005, 0.025, 0.1, 0.5, 0.9, 0.975, 0.995$ for each WM tract, (3a) the posterior limb of internal capsule and (3b) the posterior thalamic radiation. We find smaller mean and larger variance in FA distribution in older subjects with the posterior limb of internal capsule, where such increase in variance with age is greater compared to the posterior thalamic radiation.

Each of the four typical summary statistics describing shapes of the density functions, mean, variance, skew and kurtosis, expressed as a function of age ((9)–(12)) were plotted in Fig. 5 (a)–(d). We found similar decrease of mean and kurtosis with age in both WM tracts, while variance and skewness increased with age. The rate of change of each shape parameter differed in each tract reflecting heterogeneous aging mechanism.

5. Discussion

A limitation in the current FA imaging data analysis is that it focuses on only first order moment (or simply mean) in identifying risk factors. Consequently, identification of effects of risk factors on higher order moments, which enables fuller characterization of distribution, have been ignored. While GAMLSS-MX is an available R routine package for

assessing such associations in a general data structure which has one or a few measurements per subject, its application for brain imaging data analysis may not be possible due to a massive number of measurements per subject. We demonstrated that our proposed method was able to identify putative risk factors by examining additional three higher order moment statistics: variance, skewness, and kurtosis.

The proposed algorithm was applied to characterize distributional change of FA by age within two WM regions: the posterior limb of internal capsule and the posteriorthalamic radiation, where maturity of the former is followed by the latter. We found that the distributions in those two tracts tended to exhibit lower mean and lower kurtosis, higher variance and higher skewness as age increases. Although the two WM tracts showed similarity in aging-related trend of FA distribution, we observed that the posterior thalamic radiation showed a greater rate of decline in mean and a more shrunken distribution toward lower FA suggested by smaller mean and variance as age increases. This finding supports a neurodevelopmental theory, in that the posterior thalamic radiation matured later showed greater aging-related change. We recognize that larger and more diverse samples will be required to fully characterize age related change at the population level.

While we demonstrated the proposed method with two WM tracts, the sizes of which are between 7000 and 13,000 voxels, the proposed method can be applied to the whole brain-wide data, where a typical size is about a half million voxels as shown in Supplementary material 3. We also tested significance of aging effect on the mean FA from each of the two white matter ROIs by applying general linear model with two other covariates, gender and education. Significant age effects were found in both ROIs with $p = 0.0021$ (posterior thalamic radiation) and $p = 0.0201$ (posterior limb of internal capsule) while no significant gender and years of education effects were found. It is noteworthy that aging also triggers distribution-wide change in the white matter tracts as revealed in this study.

A few limitations of the proposed method are discussed as follows. The number of parameters for the direct estimation method increases linearly with increasing number of subjects (n), since the ratio of the number of parameters in the direct method compared to the indirect method is $1/2(n + 1)(1 - 1/m) + 1/m$ where the number of m-Gaussian components is fixed. This steep increase may create many local maxima in the search space; a stochastic search based on the EM algorithm could be problematic with a large number of subjects. In addition, due to the larger number of parameters to be estimated, more intensive computation is required for the direct method compared to the indirect method. To reduce the number of parameters estimated for subject-wise density function, a GMM that facilitates penalization on excessive number of parameters could be considered.

The optimal number of Gaussian components required for GMM was determined based on AIC in this study by treating each voxel measurement as independent observation. It is known that voxel values in a neighborhood are somewhat correlated. A further study on comparison of other criteria, e.g. Bayesian Information Criterion (BIC), and cross-validation, with regard to optimality in correlated brain imaging data is warranted. We believe that this examination is an important topic for future study.

In conclusion, our proposed method is powerful for analysis of brain imaging data with very large volume sizes of voxels when detection of risk factors associated with any type of microstructural neurodegenerations is of interest.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by The Dana Foundation, David Mahoney Neuroimaging Program and by NIH/NINDS (R01 NS082432).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/jjneumeth.2016.04.024.

References

Ardekani B. A fully automatic multimodality image registration algorithm. J Comput Assist Tomogr. 1995; 19(4):615–623. [PubMed: 7622696]

Ardekani B, Guckemus S, Bachman A, Hoptman MJ, Wojtaszek M, Nierenberg J. Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans. J Neurosci Methods. 2005; 142(1):67–76. [PubMed: 15652618]

Benson RR, Meda SA, Vasudevan S, Kou Z, Govindarajan KA, Hanks RA, Millis SR, Makki M, Latif Z, Coplin W, Meythaler J, Haacke EM. Global white matter analysis of diffusion tensor images is predictive of injury severity in traumatic brain injury. J Neurotrauma. 2007; 24(3):446–459. http://dx.doi.org/10.1089/neu.2006.0153, Epub 2007/04/04. [PubMed: 17402851]

Gu, C. Smoothing Spline ANOVA Models. Springer; New York: 2002.

Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. Enhancement of MR images using registration for signal averaging. J Comput Assist Tomogr. 1998; 22(2):324–333. [PubMed: 9530404]

Kim N, Branch CA, Kim M, Lipton ML. Whole brain approaches for identification of microstructural abnormalities in individual patients: comparison of techniques applied to mild traumatic brain injury. PLoS One. 2013; 8(3):e59382. http://dx.doi.org/10.1371/journal.pone.0059382, Epub 2013/04/05. [PubMed: 23555665]

Kim N, Heo M, Fleysher R, Branch CA, Lipton ML. A Gaussian mixture model approach for estimating and comparing the shapes of distributions of neuroimaging data: diffusion-measured aging effects in brain white matter. Front Public Health. 2014; 2 http://dx.doi.org/10.3389/fpubh.2014.00032.

Kochunov P, Thompson PM, Lancaster JL, Bartzokis G, Smith S, Coyle T, Royall DR, Laird A, Fox PT. Relationship between white matter fractional anisotropy and other indices of cerebral health in normal aging: tract-based spatial statistics study of aging. NeuroImage. 2007; 35(2):478–487. http://dx.doi.org/10.1016/j.neuroimage.2006.12.021, Epub 2007/02/13. [PubMed: 17292629]

Kou Z, Wu Z, Tong KA, Holshouser B, Benson RR, Hu J, Haacke EM. The role of advanced MR imaging findings as biomarkers of traumatic brain injury. J Head Trauma Rehab. 2010; 25(4):267–282. http://dx.doi.org/10.1097/HTR.0b013e3181e54793, Epub 2010/07/09.

Lim KO, Ardekani BA, Nierenberg J, Butler PD, Javitt DC, Hoptman MJ. Voxelwise correlational analyses of white matter integrity in multiple cognitive domains in schizophrenia. Am J Psychiatry. 2006; 163(11):2008–2010. http://dx.doi.org/10.1176/appi.ajp.163.11.2008, Epub 2006/11/01. [PubMed: 17074956]

Lipton ML, Gellella E, Lo C, Gold T, Ardekani BA, Shifteh K, Bello JA, Branch CA. Multifocal white matter ultrastructural abnormalities in mild traumatic brain injury with cognitive disability: a voxel-wise analysis of diffusion tensor imaging. J Neurotrauma. 2008; 25(11):1335–1342. http://dx.doi.org/10.1089/neu.2008.0547, Epub 2008/12/09. [PubMed: 19061376]

Lipton ML, Kim N, Park YK, Hulkower MB, Gardin TM, Shifteh K, Kim M, Zimmerman ME, Lipton RB, Branch CA. Robust detection of traumatic axonal injury in individual mild traumatic brain injury patients: intersubject variation, change overtime and bidirectional changes in anisotropy. Brain Imaging Behav. 2012; 6(2):329–342. http://dx.doi.org/10.1007/s11682-012-9175-2, Epub 2012/06/12. [PubMed: 22684769]

Molas M, Lesaffre E. Finite mixture models with fixed weights applied to growth data. Biom Lett. 2012; 49(2) http://dx.doi.org/10.2478/bile-2013-0008.

Muller HP, Unrath A, Riecker A, Pinkhardt EH, Ludolph AC, Kassubek J. Intersubject variability in the analysis of diffusion tensor images at the group level: fractional anisotropy mapping and fiber tracking techniques. Magn Reson Imaging. 2009; 27(3):324–334. http://dx.doi.org/10.1016/j.mri.2008.07.003, Epub 2008/08/15. [PubMed: 18701228]

Della Nave R, Foresti S, Pratesi A, Ginestroni A, Inzitari A, Salvadori E, Giannelli M, Diciotti S, Inzitari D, Mascalchi M. Whole-brain histogram and voxel-based analyses of diffusion tensor imaging in patients with leukoaraiosis: correlation with motor and cognitive impairment. Am J Neuroradiol. 2007; 28(7):1313–1319. http://dx.doi.org/10.3174/ajnr.A0555, Epub 2007/08/19. [PubMed: 17698534]

Oishi K, Faria A, Jiang H, Li X, Akhter K, Zhang J, Hsu JT, Miller MI, van Zijl PC, Albert M, Lyketsos CG, Woods R, Toga AW, Pike GB, Rosa-Neto P, Evans A, Mazziotta J, Mori S. Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and Alzheimer's disease participants. NeuroImage. 2009; 46(2):486–489. Epub 2009/04/23. [PubMed: 19385016]

Rosenbaum SB, Lipton ML. Embracing chaos: the scope and importance of clinical and pathological heterogeneity in mTBI. Brain Imaging Behav. 2012; 6(2):255–282. http://dx.doi.org/10.1007/s11682-012-9162-7, Epub 2012/05/03. [PubMed: 22549452]

Silverman, D. PET in the Evaluation of Alzheimer's Disease and Related Disorders, vol xiv. Springer; New York, NY: 2009. p. 215

Stasinopoulos DM, Rigby RA. Generalized additive models for location scale and shape (GAMLSS). R J Stat Softw. 2007; 23(7)

Smith SM, Johansen-Berg H, Jenkinson M, Rueckert D, Nichols TE, Miller KL, Robson MD, Jones DK, Klein JC, Bartsch AJ, Behrens TE. Acquisition and voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics. Nat Protoc. 2007; 2(3):499–503. http://dx.doi.org/10.1038/nprot.2007.45, Epub 2007/04/05. [PubMed: 17406613]

Smith S, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage. 2004; 23(Suppl. 1):S208–S219. [PubMed: 15501092]

Thiebaut de Schotten M, Ffytche DH, Bizzi A, Dell'Acqua F, Allin M, Walshe M, Murray R, Williams SC, Murphy DGM, Catani M. Atlasing location, asymmetry and intersubject variability of white matter tracts in the human brain with MR diffusion tractography. NeuroImage. 2011; 54(1):49–59. http://dx.doi.org/10.1016/j.neuroimage.2010.07.055. [PubMed: 20682348]

Walhovd KB, Fjell AM, Reinvang I, Lundervold A, Dale AM, Eilertsen DE, Quinn BT, Salat D, Makris N, Fischl B. Effects of age on volumes of cortex, white matter and subcortical structures. Neurobiol Aging. 2005; 26(9):1261–1270. http://dx.doi.org/10.1016/j.neurobiolaging.2005.05.020. [PubMed: 16005549]

Yan H, Wang H, Wang YH, Zhang YM. Volumetric magnetic resonance imaging classification for Alzheimer's disease based on kernel density estimation of local features. Chin Med J Peking. 2013; 126(9):1654–1660. http://dx.doi.org/10.3760/cma.j.issn.0366-6999.20122683.

HIGHLIGHTS

- Analyses focused on mean intensity fail to detect deviation in higher order moments by risk factors.
- A two-step Gaussian mixture model approach was proposed to meet such a limitation in current imaging data analysis.
- Aging-related FA change was found in mean, variance, skewness, and kurtosis.

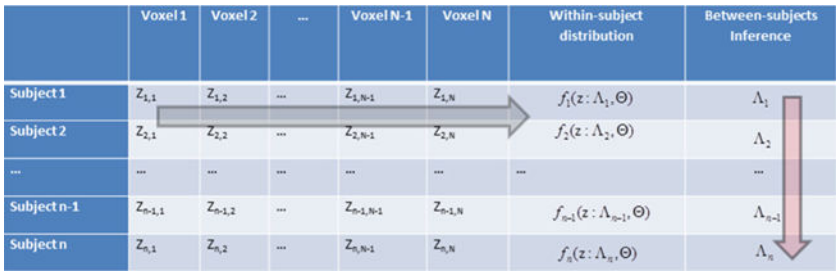

Fig. 1.

Diagram for the proposed method. Data structure for the proposed procedure is shown in a matrix format, where columns are for voxels, and rows for subjects. The proposed procedure consists of two steps, where Step 1 is shown in blue for within-subject analysis and Step 2 in red for between-subjects analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

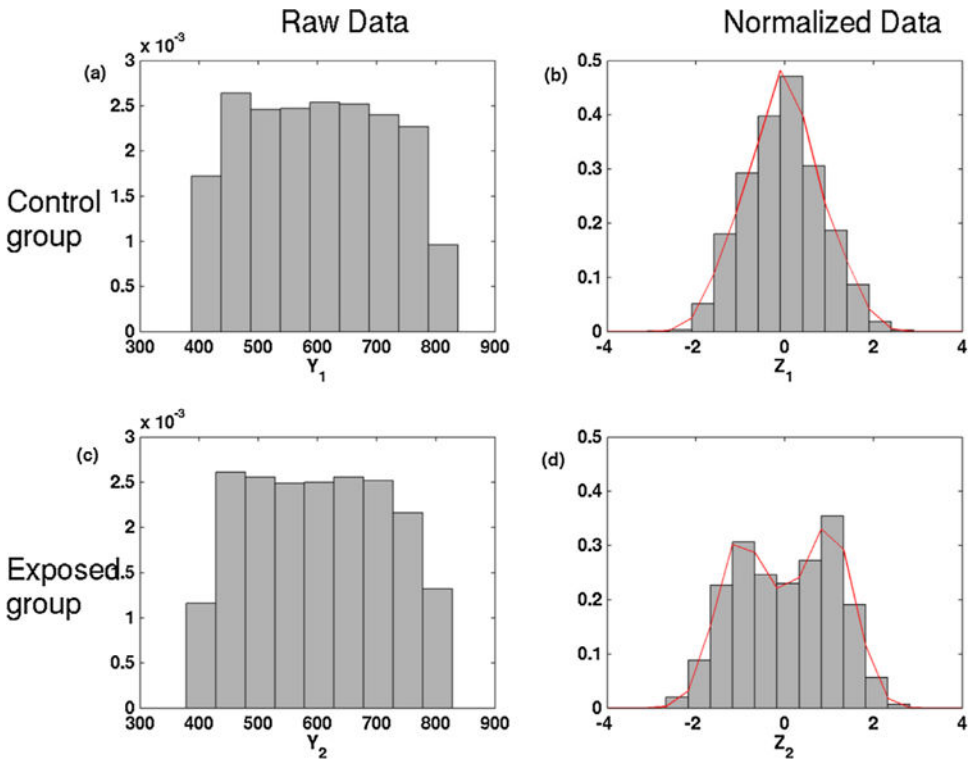

Fig. 2.

Demonstration of normalization effect on simulated data. Two histograms with raw data ((a), (c)) and normalized data ((b) and (d)) from the pooled data set with all the subjects of each group are illustrated. Estimated density function was overlaid in red solid line to the histogram with the normalized data from each group. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

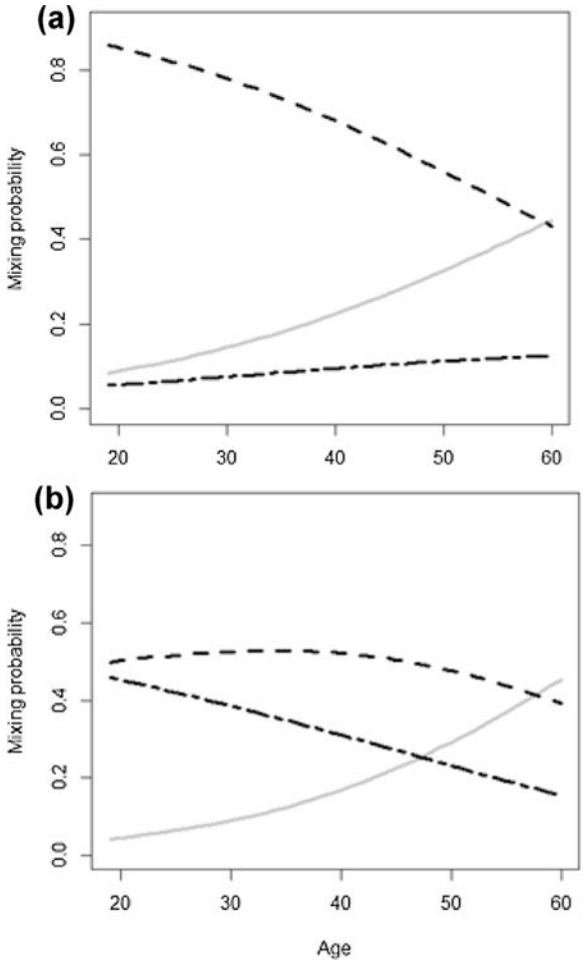

Fig. 3.

Mixing probability against age. Three mixing probabilities (p_1, p_2, and p_3) estimated in Step 2 are plotted against age for each WM region (the posterior limb of internal capsule (a) and the posterior thalamic radiation (b)), where each mixing probability is represented by solid gray, dashed black, and dash-dot black, respectively.

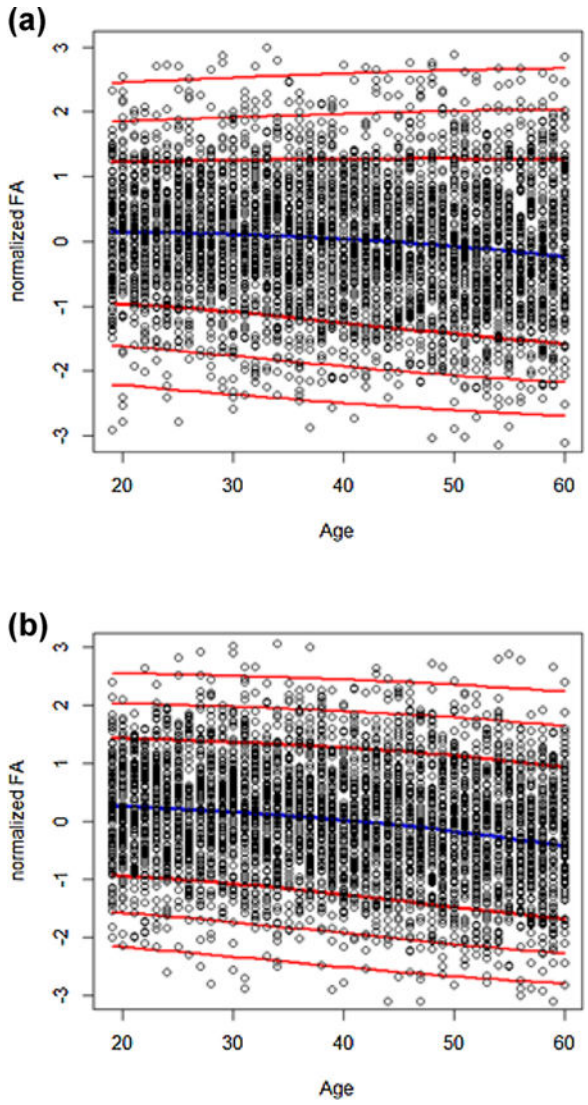

Fig. 4.

Centile curves. Seven centile curves with $\gamma = 0.005, 0.025, 0.1, 0.5, 0.9, 0.975, 0.995$ were estimated with simulated random samples from GMM density estimated at each age for each WM region (the posterior limb of internal capsule (a) and the posteriorthalamic radiation (b)). The middle blue curve is the 50th percentile curve. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

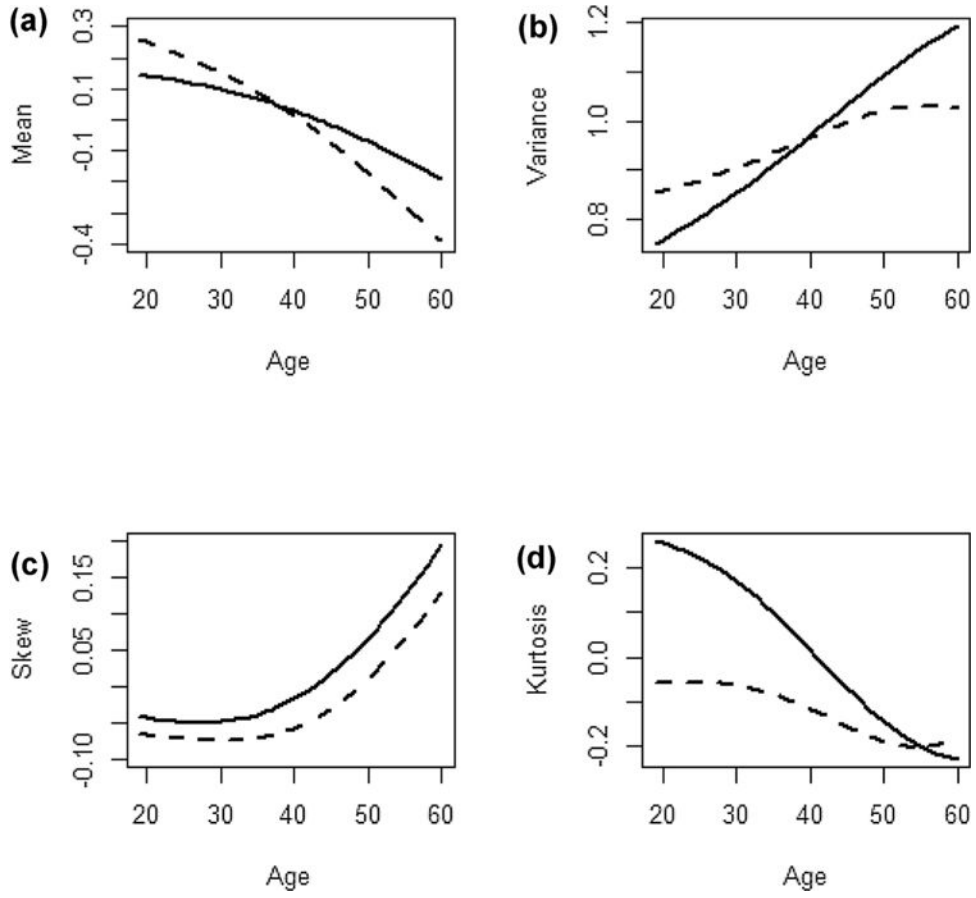

Fig. 5.

Four moment-based summary statistics (mean, variance, skew, kurtosis) against age. Estimated trajectory of mean (a), variance (b), skew(c), and kurtosis (d) against age are demonstrated for two WM tracts: the posterior limb of internal capsule (solid line) and the posterior thalamic radiation (dotted line).

Table 1

EM algorithm for the direct method.

Steps	Equations
E-step	$$\lambda_{ij}^{(t)}(k)=\frac{\lambda_i^{(t)}(k)\phi_k^{(t)}(z_{ij})}{\sum_{l=1}^{m}\lambda_i^{(t)}(l)\phi_l^{(t)}(z_{ij})},\ k=1,\ldots,m$$
M-step	$$\mu_k^{(t+1)}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{N}z_{ij}\lambda_{ij}^{(t)}(k)}{\sum_{i=1}^{n}\sum_{j=1}^{N}\lambda_{ij}^{(t)}(k)}$$
	$$\sigma^{2(t+1)}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{k=1}^{m}(z_{ij}-\mu_k^{(t+1)})^2\lambda_{ij}^{(t)}(k)}{\sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{k=1}^{m}\lambda_{ij}^{(t)}(k)}$$
	$$\lambda_i^{(t+1)}(k)=\frac{1}{N}\sum_{j=1}^{N}\lambda_{ij}^{(t)}(k)$$

Note: 1. A variable noted with (t) is the estimated value for the variable at the t-th iteration.

2. $\varphi_k^{(t)}=\varphi\left(\frac{z_{ij}-\mu_k^{(t)}}{\sigma^{(t)}}\right)$ and φ is the standard Gaussian density function.