

# Segmentation of Myelin-like Signals on Clinical MR Images for Age Estimation in Preterm Infants

Maria Deprez<sup>a,\*</sup>, Siying Wang<sup>b,1</sup>, Christian Ledig<sup>c</sup>, Joseph V. Hajnal<sup>a</sup>, Serena J. Counsell<sup>a</sup>, Julia A. Schnabel<sup>a</sup>

<sup>a</sup>*School of Imaging Sciences & Biomedical Engineering, King's College London*

<sup>b</sup>*Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK*

<sup>c</sup>*Department of Computing, Imperial College London, UK*

---

## Abstract

Myelination is considered to be an important developmental process during human brain maturation and closely correlated with gestational age. Assessment of the myelination status requires dedicated imaging, but the conventional T<sub>2</sub>-weighted scans routinely acquired during clinical imaging of neonates carry signatures that are thought to be associated with myelination. In this work, we propose a new segmentation method for myelin-like signals on T<sub>2</sub>-weighted magnetic resonance images that could be used to assess neonatal brain maturation in clinical practice. Firstly we define a segmentation protocol for myelin-like signals, and delineate manual annotations according to this protocol. We then develop an expectation-maximization framework through which we obtain the automatic segmentations of myelin-like signals. We incorporate an explicit class for partial volume voxels whose locations are configured in relation to the composing pure tissues via second-order Markov random fields. We conduct experiments in the thalami and brainstem where the majority of myelination occurs during the perinatal period for 16 test subjects aged between 29 and 44 gestational weeks. The proposed method performs accurately and robustly in both regions with respect to the manual annotations over a range of intensity percentile thresholds that are used to generate the initial segmentation estimates. Finally, we construct spatio-temporal growth models for myelin-like signals in the thalami and brain-

---

\*Corresponding author

Email address: [maria.murgasova@kcl.ac.uk](mailto:maria.murgasova@kcl.ac.uk) (Maria Deprez)

<sup>1</sup>Joint first author

stem to demonstrate the applicability of the proposed method for age estimation in preterm infants.

**Keywords:** Neonatal brain MRI, myelination, brain segmentation, partial volume, Markov random fields

---

## 1. Introduction

Human brain maturation involves a complex series of morphological, structural and functional changes. Among these changes is the process of myelin growth and axon ensheathment known as myelination, which facilitates electrical conduction in the neural system [19]. In clinical practice, birth weights are commonly used for measuring gestational maturity [26]. However, Rorke *et al.* [26] reported that some well-developed infants based on the measurements of body weight show retarded brain maturation, whereas the neurodevelopmental patterns in preterm infants with extensive myelination are similar to those found in full-term babies, regardless of the birth weights. Considering the critical role of myelination in neurodevelopment [12], it is of great interest to utilize myelin as an early marker of abnormal brain development in newborns.

Non-invasive non-ionizing magnetic resonance imaging (MRI) can be used to track myelination in the developing brain, for example  $T_1$ -weighted ( $T_1w$ ) and  $T_2w$  imaging, myelin water imaging (MWI) [1], magnetization transfer imaging (MTI) [16], and diffusion tensor imaging (DTI) [20]. In this work, we use  $T_2w$  images, where myelin appears as low intensities, due to the close correlation between  $T_2$  relaxation time and the final stage of myelination [11]. Moreover,  $T_1w$  and  $T_2w$  acquisitions remain the most routine MRI techniques compared to the advanced MWI, MTI and DTI which require elaborate scans that are constrained in clinical practice when imaging fragile preterm neonates [29, 32]. Qualitative descriptions of myelination on  $T_2w$  images [4, 8] are in great consistency with histological observations [42]. Thus the ability to numerically assess myelination using  $T_2w$  scans would facilitate neurodevelopmental evaluation in preterm infants directly from large-scale clinical MRI data.

In this study, we refer to the tissue that is likely to contain myelin in  $T_2w$  neonatal brain MRI as “myelin-like signals” (MLS). Automatic segmentation of MLS is challenging. Firstly, myelin is not included in any of the existing neonatal brain atlases [18, 30] or manual annotation database [13]. Most neonatal brain segmentation methods adapt the approaches developed for adults [3, 15, 34, 39, 43], and use a probabilistic atlas or manual annotations to obtain prior informa-

tion on the expected tissue locations. Prastawa *et al.* [25] simulated a newborn brain atlas by averaging the semi-automatic segmentations from three subjects, and separated the myelinated and unmyelinated white matter (WM) using minimum spanning tree. Xue *et al.* [41] performed intensity-based  $k$ -means clustering that divided the developing brain into gray matter (GM), WM and cerebrospinal fluid (CSF) without requiring any probabilistic or manual atlases. Anbeek *et al.* [2] trained a  $k$ NN classifier using samples selected from the manual annotations of 12 patients based on the features of voxel intensities and spatial locations. Recently developed methods in the NeoBrainS12 challenge [14, 22, 24, 38, 40] all showed great promise in general neonatal brain segmentation. However, none of them performed well in segmenting myelination due to insufficient spatial prior information [17]. Secondly, partial volume (PV) voxels containing both MLS and the background (BKG) tissue in a region of interest (ROI) need to be modeled explicitly when segmenting MLS. Fig. 1 illustrates that a simple two-class Gaussian mixture model (GMM) without PV modeling fails to separate the intensity distributions of MLS and BKG. The current PV estimation methods [6, 21, 23, 31, 33, 36] depend on prior knowledge of the spatial locations of the composing pure tissues to guide the search for PV voxels, which is, however, very limited in MLS segmentation.

We therefore develop a new expectation-maximization (EM) framework for MLS segmentation on  $T_2w$  neonatal brain images that does not require any probabilistic atlas or manual annotation of myelin. We introduce an explicit PV class whose locations are configured in relation to MLS and BKG in a predefined ROI using a 3D connectivity tensor via second-order Markov random fields (MRFs) [21]. This approach allows us to distinguish the small proportion of MLS in the presence of substantial PV voxels. Our method achieves automatic MLS segmentations of high Dice coefficients (DCs) [10] with respect to the manual annotations for 16 preterm infants at one-week intervals between 29 and 44 weeks gestational age (GA). We further use the segmentations obtained from 114 preterm infants to build a spatio-temporal model of progressing myelination.

This paper is structured as follows. In Section 2, we define an MLS segmentation protocol in two ROIs, namely the thalami and brainstem. Manual annotations are delineated according to this protocol. Section 3 presents the proposed MLS segmentation method, followed by the experiments and results in Section 4. Finally, we demonstrate the applicability of this segmentation approach for assessing neonatal brain development in Section 5.

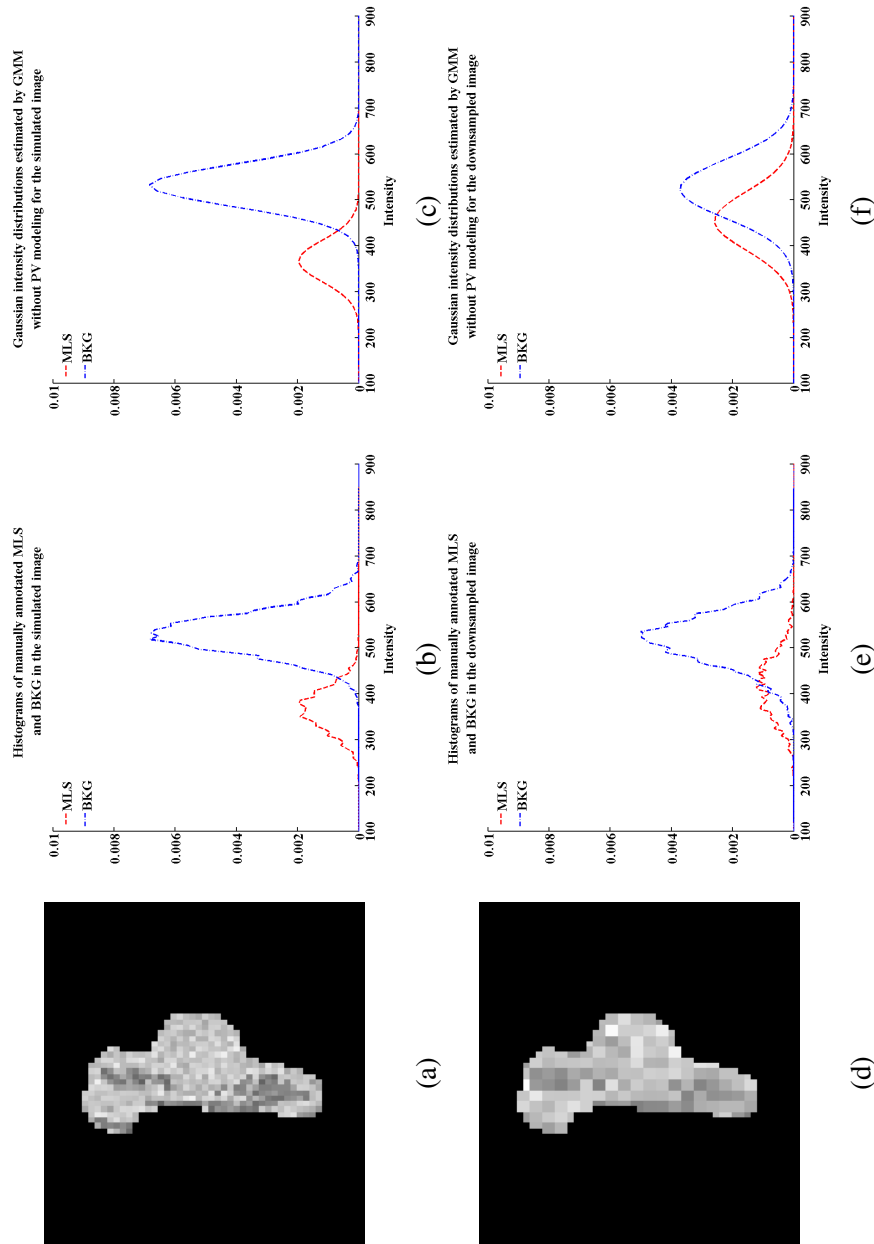


Figure 1: Simulated  $T_2$ -weighted images in the brainstem (sagittal view) for a preterm infant at 38 gestational weeks (a) before and (d) after downsampling. The intensities of myelin-like signals (MLS) and the remaining background (BKG) are generated from two Gaussians, and the MLS voxels are labeled based on the manual annotation. A two-class Gaussian mixture model (GMM) recovers the distributions correctly before downsampling. Partial volume (PV) voxels created after downsampling shift the histogram of manually annotated MLS towards BKG, and the simple GMM without PV estimation can no longer approximate the histograms accurately.

## 2. Materials

### 2.1. Subjects and MR acquisition

114 preterm infants without brain injuries were scanned between 29 and 44 weeks GA at Hammersmith Hospital, London, UK. Ethical permission for this study was granted by the Hammersmith and Queen Charlotte's and Chelsea Research Ethics Committee (07/H0707/101). Written parental consent was obtained prior to scanning. T<sub>2</sub>w fast spin-echo brain images were acquired on a 3T Philips Intera system with repetition time = 8700 ms, echo time = 160 ms and voxel sizes = 0.86 mm × 0.86 mm × 1 mm.

### 2.2. Segmentation protocol and manual annotations

We choose the thalami and brainstem as the ROIs for assessing MLS because these are the regions where the majority of myelination occurs between 29 and 44 weeks GA [8, 28]. The automatic procedure through which we obtain the binary ROI masks of the thalami and brainstem for individual subjects is explained in Section 3.1.

The MLS segmentation protocol is defined within the ROIs based on both image intensities and anatomical knowledge [8]. We delineate three central brain structures in the thalami, including the posterior limbs of the internal capsule (PLIC), ventrolateral nuclei (VLN) and subthalamic nuclei (STN). Particularly the PLIC tracts in the preterm brain, which show MLS at approximately 40 weeks GA on T<sub>2</sub>w images [8], is an important signature that can be used to assess neonatal brain maturation [5]. We define the segmentation protocol in the brainstem mainly based on intensities because the image resolution is insufficient to identify the detailed anatomical structures.

Manual annotations of MLS, delineated in the ROIs according to the defined segmentation protocol, are available for 16 preterm infants, each at a different time point, between 29 and 44 weeks GA at one-week intervals. An example is shown in Fig. 2. Eight of the 16 subjects, from 30 weeks GA onward at two-week intervals, have a repeated manual segmentation delineated by the same rater in order to assess the intra-rater reliability. The quality of these manual annotations was confirmed by a clinical expert.

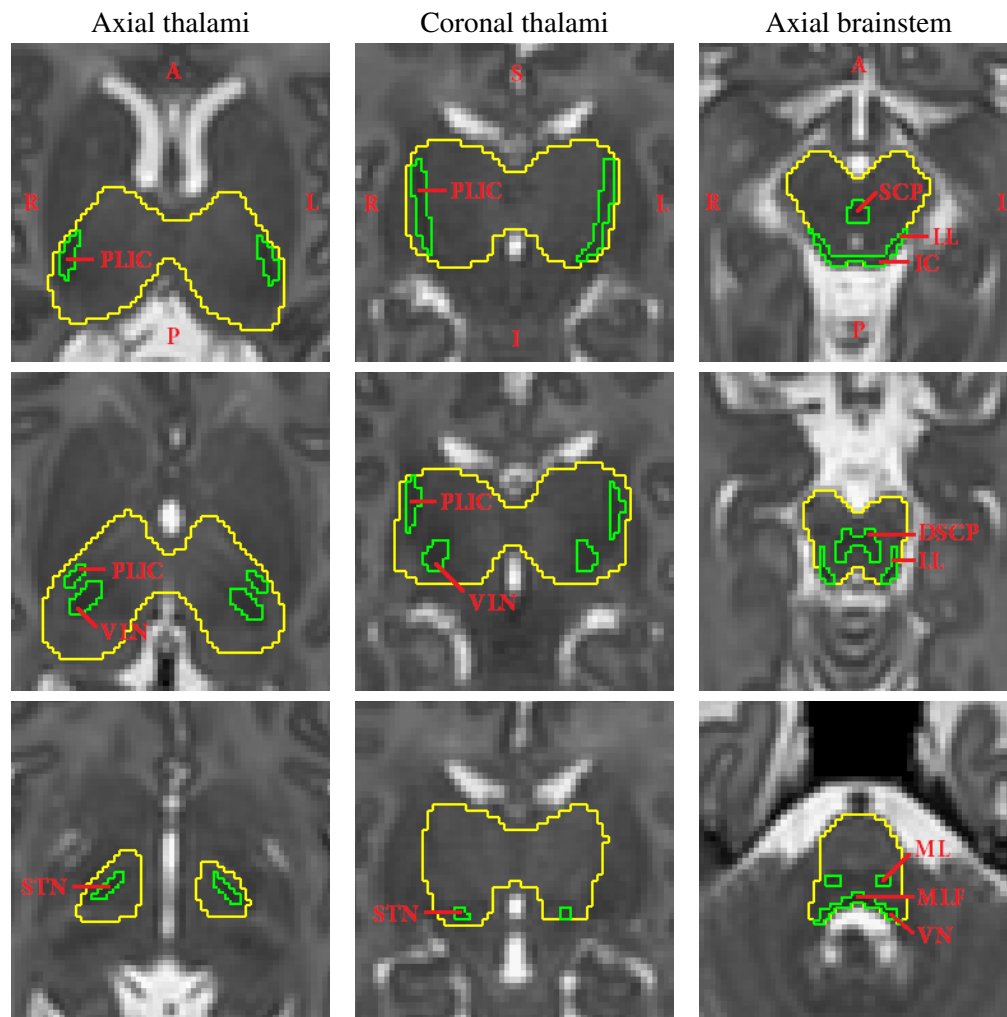


Figure 2: Manual annotations of myelin-like signals (green label) in the thalami and brainstem of a subject at 42 gestational weeks, delineated according to the defined segmentation protocol. The regions of interest are labeled in yellow. The columns from left to right show the thalami in the axial and coronal views, and the brainstem in the axial view. Abbreviations: PLIC–posterior limb of the internal capsule, VLN–ventrolateral nucleus, STN–subthalamic nucleus, SCP–superior cerebellar peduncle, DSCP–decussation of the superior cerebellar peduncle, IC–inferior colliculus, LL–lateral lemniscus, ML–medial lemniscus, MLF–medial longitudinal fasciculus, VN–vestibular nucleus.

### 3. Method

The overall pipeline proposed for MLS segmentation on T<sub>2</sub>w neonatal brain MR images is shown in Fig. 3. In this section, we first explain the procedure of image preprocessing. We then describe a purely intensity-based GMM without PV modeling, followed by the incorporation of an explicit PV class and spatial regularization in the EM framework. Lastly, we explain the implementation details of the proposed method for obtaining the final binary MLS segmentations.

#### 3.1. Image preprocessing

The aim of image preprocessing is to achieve brain extraction and bias field correction, and to obtain the binary ROI masks of the thalami and brainstem for individual subjects.

We remove the non-brain tissues using label propagation [15] of manually annotated brain masks, and segment the skull-stripped images using the Statistical Parametric Mapping (SPM) software<sup>2</sup> (version SPM8) [3]. The spatial priors are provided by a 4D probabilistic neonatal brain atlas<sup>3</sup> [18]. Subsequently, we obtain the bias-corrected images as well as the segmentations of the cortical GM, WM, deep gray matter (DGM), brainstem, cerebellum and CSF.

We use the automatic SPM segmentations of the brainstem directly as the ROI masks. The DGM segmentations include the basal ganglia, thalami, hippocampi and amygdalae. Since the myelinated thalamic nuclei can have similar intensities to the densely organized GM in the basal ganglia which does not contain myelin between 29 and 44 weeks GA [8], we extract the ROI masks for the thalami from the DGM segmentations in order to prevent misclassifications by the intensity-based segmentation model. We first manually delineate the thalami on the T<sub>2</sub>w image of a single reference subject at 36 weeks GA. We register the reference image with the dilated DGM region of each subject using free-form deformation (FFD) non-rigid registration [27], and then transform the manual delineation of the thalami from the reference space to each individual subject's space. The ROI masks of the thalami and brainstem are illustrated in Fig. 4.

---

<sup>2</sup><http://www.fil.ion.ucl.ac.uk/spm>

<sup>3</sup><http://brain-development.org/brain-atlases>

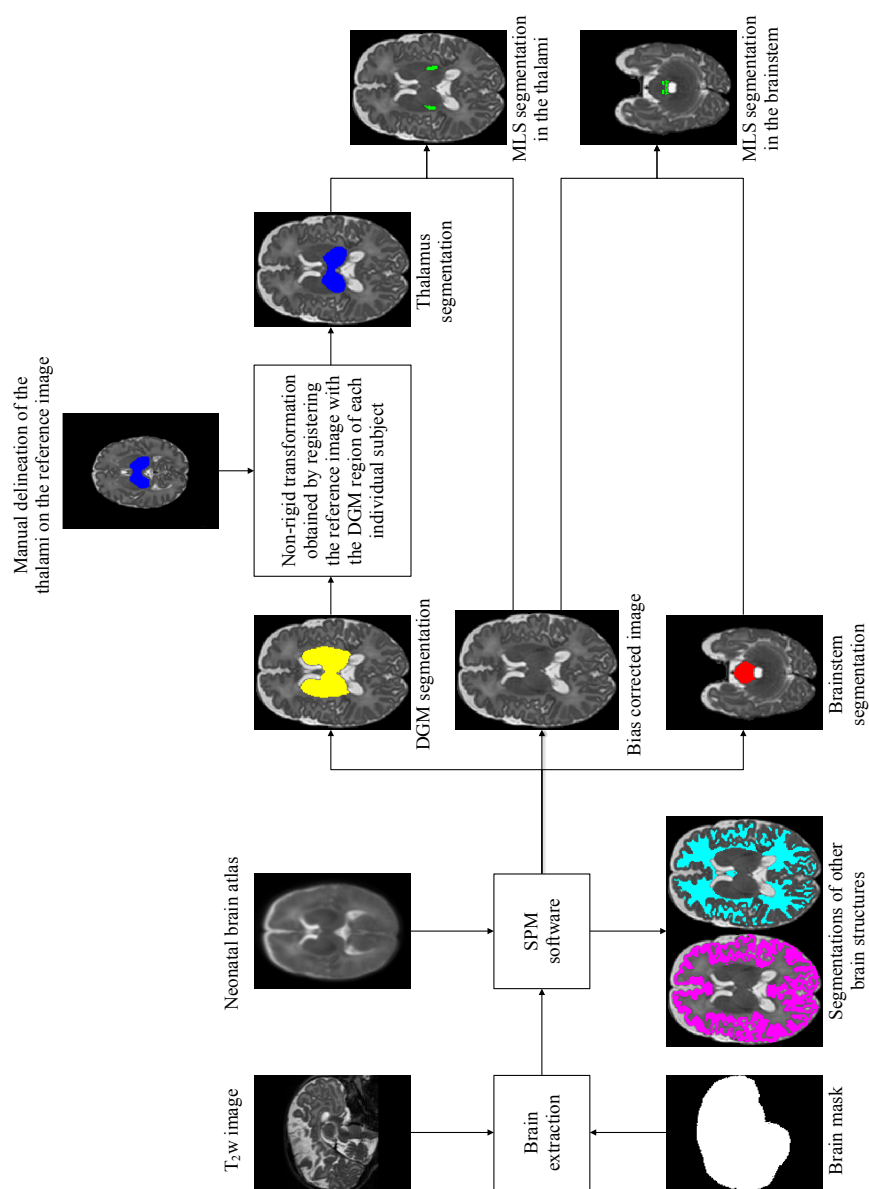


Figure 3: Schematic flow of the proposed segmentation approach for myelin-like signals (MLS) on T<sub>2</sub>w neonatal brain images. Through image preprocessing, we achieve brain extraction and bias field correction, and obtain the binary masks of the thalami and brainstem for individual subjects. The automatic segmentations of the deep gray matter (DGM), obtained using the Statistical Parametric Mapping (SPM) software, contain the basal ganglia, thalami, hippocampi and amygdalae. We extract the masks for the thalami from the DGM segmentations via label propagation.



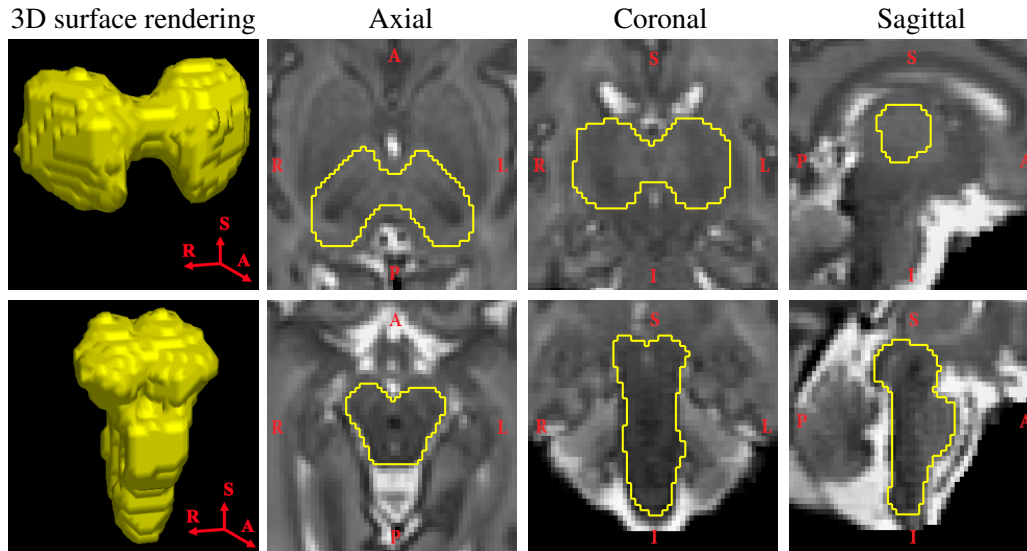


Figure 4: Two separate regions of interest (ROIs) labeled in yellow. The first and second rows show the thalami and brainstem, respectively, for a subject at 34 gestational weeks. The columns from left to right show the ROIs in 3D surface rendering, axial, coronal and sagittal views.

### 3.2. Purely intensity-based GMM without PV modeling

First we segment MLS using a standard GMM without PV modeling. We define two classes in the ROI: myelin-like signals (MLS) and background (BKG). The conditional probability density function (PDF) of class  $k$  is approximated as a Gaussian with mean  $\mu_k$  and standard deviation (SD)  $\sigma_k$ . The prior probability that voxel  $i$  belongs to class  $k$  is modeled by the spatially constant mixing proportion  $c_k$ . We apply the EM algorithm [9] as follows:

- E-step: Calculate the probability  $p_{ik}$  that voxel  $i$  belongs to class  $k$ :

$$p_{ik}^{(m+1)} = \frac{G(y_i, \mu_k^{(m)}, \sigma_k^{(m)}) c_k^{(m)}}{\sum_{k'=1}^K G(y_i, \mu_{k'}^{(m)}, \sigma_{k'}^{(m)}) c_{k'}^{(m)}} \quad (1)$$

where  $y_i$  is the observed intensity,  $K$  the number of classes, and  $m$  the current EM iteration number.

- M-step: We assume that the SDs of the MLS and BKG classes are equal. Ideally the MLS and BKG tissues in the ROI have uniform intensities. Variations arise mainly due to noise and can be considered constant across the MR image. Therefore, the two classes may be regarded as Gaussians with equal SDs

provided that their mean intensities are distant from zero. Based on this assumption, the maximum likelihood (ML) estimations for the GMM parameters of class  $k$  yield the following equations:

$$\mu_k^{(m+1)} = \frac{\sum_{i=1}^N p_{ik}^{(m+1)} y_i}{\sum_{i=1}^N p_{ik}^{(m+1)}} \quad (2)$$

$$\left(\sigma_k^{(m+1)}\right)^2 = \left(\sigma^{(m+1)}\right)^2 = \frac{1}{N} \sum_{l=1}^K \sum_{i=1}^N p_{il}^{(m+1)} \left(y_i - \mu_l^{(m+1)}\right)^2 \quad (3)$$

$$c_k^{(m+1)} = \frac{1}{N} \sum_{i=1}^N p_{ik}^{(m+1)} \quad (4)$$

where  $N$  is the number of voxels in the ROI.

### 3.3. Explicit PV modeling

As explained in Fig. 1, MLS becomes much less distinguishable in the histogram of the downsampled image due to the intermediate PV intensities. The simple two-class GMM described in Section 3.2 over-estimates the MLS class and cannot recover the Gaussians before downsampling. Therefore, we introduce an explicit PV class in addition to the previous MLS and BKG classes to model the intensities of mixed tissues.

The E-step (Eq. 1) remains unchanged at this stage. In the M-step, we approximate the PV class mean as the arithmetic mean of the composing pure tissue means, and assume that the SDs of the MLS, PV and BKG classes are identical for the reason explained in Section 3.2:

$$\mu_{PV} = \frac{1}{2}(\mu_{MLS} + \mu_{BKG}) \quad (5)$$

$$\sigma_{MLS} = \sigma_{PV} = \sigma_{BKG} \quad (6)$$

By substituting Eqs. 5 and 6 into the objective function of log likelihood [35], the SD  $\sigma_k$  of class  $k$  can be updated using Eq. 3 and the ML estimations for the means of the MLS and BKG classes result in the following system of linear equations (see more details in the appendix):

$$\begin{bmatrix} \mu_{MLS}^{(m+1)} \\ \mu_{BKG}^{(m+1)} \end{bmatrix} = \begin{bmatrix} \frac{1}{4}B_{PV} & B_{BKG} + \frac{1}{4}B_{PV} \\ B_{MLS} + \frac{1}{4}B_{PV} & \frac{1}{4}B_{PV} \end{bmatrix}^{-1} \begin{bmatrix} A_{BKG} + \frac{1}{2}A_{PV} \\ A_{MLS} + \frac{1}{2}A_{PV} \end{bmatrix} \quad (7)$$

where

$$A_k = \sum_{i=1}^N y_i p_{ik}^{(m+1)} \quad \text{and} \quad B_k = \sum_{i=1}^N p_{ik}^{(m+1)}$$

### 3.4. Spatial regularization via second-order MRFs

Van Leemput *et al.* [36] pointed out that introducing an additional PV class without any spatial prior can disrupt the process of EM convergence. Therefore, robust PV estimation methods generally require prior knowledge of the spatial locations of the composing pure tissues to guide the search for PV voxels [6, 21, 23, 31, 33, 36]. However, MLS is not included in any of the existing neonatal brain atlases [18, 30] or manual annotation database [13]. To overcome this difficulty, we incorporate spatial regularization in the EM framework via second-order MRFs [21] where we configure the dependencies among a triplet of classes in the neighborhood using a 3D connectivity tensor (Fig. 5).

First we modify the E-step (Eq. 1) by replacing the spatially constant mixing proportion  $c_k$  with the spatial prior  $p(\mathbf{z}_i = \mathbf{e}_k | \mathbf{p}_{\mathcal{N}_i}^{(m)}, \Phi_z)$  acquired from the neighborhood as described in [35]:

$$p_{ik}^{(m+1)} = \frac{G(y_i, \mu_k^{(m)}, \sigma_k^{(m)}) p(\mathbf{z}_i = \mathbf{e}_k | \mathbf{p}_{\mathcal{N}_i}^{(m)}, \Phi_z)}{\sum_{k'=1}^K G(y_i, \mu_{k'}^{(m)}, \sigma_{k'}^{(m)}) p(\mathbf{z}_i = \mathbf{e}_{k'} | \mathbf{p}_{\mathcal{N}_i}^{(m)}, \Phi_z)} \quad (8)$$

where

$$p(\mathbf{z}_i = \mathbf{e}_k | \mathbf{p}_{\mathcal{N}_i}^{(m)}, \Phi_z) = \frac{e^{-U_{\text{MRF}}(\mathbf{e}_k | \mathbf{p}_{\mathcal{N}_i}^{(m)}, \Phi_z)}}{\sum_{k''=1}^K e^{-U_{\text{MRF}}(\mathbf{e}_{k''} | \mathbf{p}_{\mathcal{N}_i}^{(m)}, \Phi_z)}} \quad (9)$$

The symbol  $\mathbf{z}_i$  indicates to which class voxel  $i$  belongs, and  $\mathbf{e}_k$  is a unit vector of length  $K$  whose components are equal to zero except the  $k$ th component. The probabilities in the 26-neighborhood  $\mathcal{N}_i$  of voxel  $i$  are denoted as  $\mathbf{p}_{\mathcal{N}_i}$ , and the MRF parameters as  $\Phi_z$ . We compute the energy function  $U_{\text{MRF}}(\mathbf{e}_k | \mathbf{p}_{\mathcal{N}_i}, \Phi_z)$  for second-order MRFs as follows:

$$\begin{aligned} U_{\text{MRF}}(\mathbf{e}_k | \mathbf{p}_{\mathcal{N}_i}^{(m)}, \Phi_z) &= \sum_{k_1=1}^K v_{k_1}^{(m)} \sum_{k_2=1}^K \mathbf{T}_k(k_1, k_2) v_{k_2}^{(m)} \\ &= [v_1 \quad v_2 \quad \cdots \quad v_K] \begin{bmatrix} T_k(1,1) & T_k(1,2) & \cdots & T_k(1,K) \\ T_k(2,1) & T_k(2,2) & \cdots & T_k(2,K) \\ \vdots & \vdots & \ddots & \vdots \\ T_k(K,1) & T_k(K,2) & \cdots & T_k(K,K) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_K \end{bmatrix} \end{aligned} \quad (10)$$

	BKG	PV	MLS		BKG	PV	MLS		BKG	PV	MLS
BKG	$t_1$	0	$t_2$	BKG	$t_3$	0	0	BKG	0	0	$t_2$
PV	0	0	0	PV	0	0	0	PV	0	0	0
MLS	$t_2$	0	0	MLS	0	0	$t_3$	MLS	$t_2$	0	$t_1$
$k = \text{MLS}$				$k = \text{PV}$				$k = \text{BKG}$			

Figure 5: Matrix  $\mathbf{T}_k$  in the 3D connectivity tensor for  $k = \text{MLS}$ ,  $\text{PV}$  and  $\text{BKG}$ . Each element  $\mathbf{T}_k(k_1, k_2)$  indicates the penalty when classes  $k_1$  and  $k_2$  are both present in the neighborhood of class  $k$ . We will determine the values of penalties  $t_1$ ,  $t_2$  and  $t_3$  empirically in Section 4.1. Abbreviations: MLS—myelin-like signals, PV—partial volume, BKG—background.

where

$$v_k^{(m)} = \sum_{j \in \mathcal{N}_i} \frac{p_{jk}^{(m)}}{d_{ij}} \quad (11)$$

Here  $p_{jk}$  is the probability that neighbor  $j$  belongs to class  $k$ , and  $d_{ij}$  the Euclidean distance between voxel  $i$  and neighbor  $j$ . The element of the connectivity tensor, denoted as  $\mathbf{T}_k(k_1, k_2)$ , indicates the penalty when classes  $k_1$  and  $k_2$  are both present in the neighborhood of class  $k$ . By assigning a large value to  $\mathbf{T}_k(k_1, k_2)$  in Eq. 10, we are able to penalize a particular configuration when the probabilities of classes  $k_1$  and  $k_2$  are both high in the neighborhood. This would result in a large value of the energy function  $U_{\text{MRF}}(\mathbf{e}_k | \mathbf{p}_{\mathcal{N}_i}, \Phi_z)$ , making voxel  $i$  less probable to belong to class  $k$ .

We assign the penalties  $\mathbf{T}_k(k_1, k_2)$  according to the following rules based on the triplet that contains class  $k$  of the current voxel and classes  $k_1$  and  $k_2$  of a pair of neighboring voxels:

- $\mathbf{T}_k(\text{MLS}, \text{MLS})$  and  $\mathbf{T}_k(\text{BKG}, \text{BKG})$ : We set the penalty to 0 if class  $k$  is the same as the pair. We assign a penalty  $t_1$  if  $k$  is a pure tissue class different from the pair to encourage the PV class as MLS and BKG are both in the triplet. We assign a penalty  $t_3$  if  $k$  is PV as there is no evidence for PV in the triplet. This effectively prevents over-estimation of the PV class.
- $\mathbf{T}_k(\text{MLS}, \text{BKG})$ : We assign a penalty  $t_2$  if  $k$  is MLS or BKG, and set the penalty to 0 if  $k$  is PV. We aim to encourage the PV class by deliberately

penalizing the MLS and BKG classes if they are both present in the neighborhood.

- $\mathbf{T}_k(\text{MLS}, \text{PV})$ ,  $\mathbf{T}_k(\text{PV}, \text{PV})$  and  $\mathbf{T}_k(\text{BKG}, \text{PV})$ : We set the penalty to 0 regardless of the class of  $k$ .

The values of penalties  $t_1$ ,  $t_2$  and  $t_3$  will be determined empirically in Section 4.1. We again assume that the PV class mean equals the arithmetic mean of the composing pure tissue means and that the SDs of the MLS, PV and BKG classes are identical. Provided that the penalties remain constant, the SD  $\sigma_k$  and mean  $\mu_k$  of class  $k$  in the EM framework with MRFs can be estimated using Eqs. 3 and 7, respectively, in the same way as the purely intensity-based GMM with an explicit PV class (see the appendix for details).

### 3.5. Implementation

The EM algorithm is initialized as follows. We first sort the  $T_2w$  intensities in the thalami of individual subjects in ascending order and choose the 6th percentile as the threshold. This is because the average volume fraction of manually annotated MLS in the thalami is 6% over the 16 test subjects at one-week intervals between 29 and 44 weeks GA. Voxels with intensities below the threshold constitute the initial MLS segmentation, and the remaining voxels in the ROI form the initial BKG segmentation. We will verify in Section 4.3 that the percentile estimated this way optimizes the MLS segmentations that could possibly be achieved for all the 16 subjects using a single threshold. We assign the initial MLS and BKG segmentations to the corresponding posterior probability maps (PPMs), and set the PPM of the PV class equal to zero. Similarly, we threshold the  $T_2w$  intensities in the brainstem of individual subjects at the 25th percentile to obtain the initial segmentation estimates in this region.

Tissue classification is achieved by interleaving the E-step and M-step as described in Section 3.4 until the relative change in the objective function of log likelihood [35] is less than 0.01%. The resultant PPMs are converted into hard segmentations using the maximum-vote rule. In addition, we reclassify the PV voxels as one of the composing pure tissues by calculating the fraction of the MLS class, denoted as  $f_i$ , at PV voxel  $i$ :

$$f_i = \frac{\mu_{\text{BKG}} - y_i}{\mu_{\text{BKG}} - \mu_{\text{MLS}}} \quad (12)$$

where  $\mu_{\text{MLS}}$  and  $\mu_{\text{BKG}}$  are the means of the MLS and BKG classes, and  $y_i$  the observed intensity. PV voxels with fractions above 0.5 are reclassified as MLS

and combined with the voted hard segmentation of the MLS class to form the final segmentation.

#### 4. Experiments and results

In this section, we first determined the optimal values of the penalties in the connectivity tensor for the proposed method with explicit PV modeling via second-order MRFs (GMM-PV-MRF). We then evaluated GMM-PV-MRF in terms of Dice overlap with manually annotated MLS, as compared to two other segmentation models in Table 1. The initial MLS segmentations generated by optimal thresholding exploited manual annotations of MLS that are usually unavailable in practice. We thus further investigated the dependence of each method on the initial threshold. Lastly, we assessed the intra-rater reliability compared to the accuracies of the automated model-based methods.

Segmentation model	Description	Partial volume (PV) modeling	Spatial regularization
GMM	Purely intensity-based Gaussian mixture model without PV modeling	×	×
GMM-PV	GMM incorporating a spatially unconstrained PV class	✓	×
GMM-PV-MRF	GMM with explicit PV modeling via second-order MRFs	✓	✓

Table 1: Segmentation models for myelin-like signals.

#### 4.1. Optimization of the penalties in the connectivity tensor

As explained in Section 3.4, we acquired the MLS, PV and BKG priors by configuring their spatial dependencies in the connectivity tensor (Fig. 5) via second-order MRFs. For each individual ROI, we optimized the penalties  $t_1$ ,  $t_2$  and  $t_3$  using a coarse-to-fine grid search method.

We first determined the coarse value of each penalty from the following set  $\{0.003, 0.007, 0.03, 0.07, 0.3, 0.7\}$ . Two values were included for the lower and upper halves of each tenfold scale. We computed MLS segmentations using all 216 combinations of the  $t_1$ ,  $t_2$  and  $t_3$  values. We chose the combination that yielded the highest average DC over the 16 test subjects with respect to the manual annotations. A DC is defined for two overlapping volumes as the ratio of the number of voxels in the intersection to the average number of voxels in the two volumes:

$$DC = \frac{2N(A \cap B)}{N(A) + N(B)} \quad (13)$$

where  $N(A)$  denotes the number of voxels in volume  $A$ .

We then performed a fine grid search through ten different values for each penalty with the selected coarse value at the center. For instance, we examined the set  $\{0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07\}$  for  $t_1$  in the thalami, whose best value at the coarse level was found to be 0.03. The test sets and optimal values of all three penalties at the coarse and fine levels are summarized in Table 2 for both ROIs. The final  $(t_1, t_2, t_3)$  combination that yielded the highest average DC was  $(0.05, 0.03, 0.01)$  in the thalami, and  $(0.05, 0.03, 0.009)$  in the brainstem, which were highly consistent even though optimized separately.

Search level	Penalty	Set of test values	Optimal value
Coarse	$t_1$	$\{0.003, 0.007, 0.03, 0.07, 0.3, 0.7\}$	0.03
	$t_2$	Same as $t_1$ at the coarse level	0.03
	$t_3$	Same as $t_1$ at the coarse level	0.03
Fine	$t_1$	$\{0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07\}$	0.05
	$t_2$	Same as $t_1$ at the fine level	0.03
	$t_3$	Same as $t_1$ at the fine level	0.01

(a) Grid search results in the thalami

Search level	Penalty	Set of test values	Optimal value
Coarse	$t_1$	$\{0.003, 0.007, 0.03, 0.07, 0.3, 0.7\}$	0.03
	$t_2$	Same as $t_1$ at the coarse level	0.03
	$t_3$	Same as $t_1$ at the coarse level	0.007
Fine	$t_1$	$\{0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07\}$	0.05
	$t_2$	Same as $t_1$ at the fine level	0.03
	$t_3$	$\{0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.02, 0.03\}$	0.009

(b) Grid search results in the brainstem

Table 2: Optimal values of the penalties  $t_1$ ,  $t_2$  and  $t_3$  in different regions of interest (ROIs) determined using a coarse-to-fine grid search method. These penalties are assigned to the connectivity tensor (Fig. 5) for configuring spatial dependencies of myelin-like signals, partial volume voxels and the background in each ROI.



#### 4.2. Evaluation of the MLS segmentation methods

We segmented MLS using GMM, GMM-PV and GMM-PV-MRF (see Table 1) in the thalami and brainstem for the 16 test subjects at one-week intervals between 29 and 44 weeks GA. From here on, we used the optimal values of  $t_1$ ,  $t_2$  and  $t_3$  determined in Section 4.1 for GMM-PV-MRF. We implemented GMM and GMM-PV as described in Sections 3.2 and 3.3 respectively.

The DCs of the resultant segmentations with respect to the manual annotations are summarized for both ROIs in Table 3. We also evaluated the initial MLS segmentations obtained by optimal thresholding as a baseline method. We compared GMM-PV-MRF with all the other methods using two-tailed Student's  $t$ -tests at the 5% significance level, and the  $p$ -values are shown in Table 3.

Method	Average DC	$p$ -value
Thresholding at the 6th percentile	$0.789 \pm 0.091$	0.043
GMM	$0.762 \pm 0.092$	0.001
GMM-PV	$0.662 \pm 0.192$	0.002
GMM-PV-MRF	$0.837 \pm 0.057$	—

(a) Evaluation results in the thalami

Method	Average DC	$p$ -value
Thresholding at the 25th percentile	$0.827 \pm 0.035$	0.020
GMM	$0.675 \pm 0.206$	0.010
GMM-PV	$0.762 \pm 0.057$	$5.449 \times 10^{-4}$
GMM-PV-MRF	$0.831 \pm 0.038$	—

(b) Evaluation results in the brainstem

Table 3: Average Dice coefficients (DCs) ( $\pm$  standard deviations) of myelin-like signals segmented by optimal thresholding and the model-based methods (see Table 1 for method descriptions) for 16 test subjects aged between 29 and 44 gestational weeks. GMM-PV-MRF is compared with all the other methods using two-tailed Student's  $t$ -tests at the 5% significance level. GMM-PV-MRF achieves the highest average DC and significantly outperforms all the other methods in both regions of interest.

Thresholding at the optimal level resulted in a lower average DC and a larger SD in the thalami than in the brainstem. The 6th percentile threshold was derived from the average volume fraction of manually annotated MLS in the thalami over the 16 subjects, and hence decreased the DCs at the early and late GAs due to the appearance of new myelination in this region [8]. In contrast, the myelinated brainstem grew with the overall increase in the brain size with relatively little new myelination appearing between 29 and 44 weeks GA [7, 28], which implied an approximately constant MLS volume fraction.

The purely intensity-based GMM without PV modeling displayed large segmentation variations in the brainstem compared to the thalami. In Figs. 6 and 7, we studied the Gaussian distributions of the PPMs estimated using different segmentation models for a subject at 37 weeks GA. With respect to the manual annotation histograms, GMM over-estimated the MLS class in the thalami with a shift of the mean towards BKG, and failed to separate the two classes in the brainstem. This is caused by PV voxels represented as the overlapping area between the histograms of manually annotated MLS and BKG.

The extra Gaussian in GMM-PV caused large segmentation variations in the thalami due to the absence of spatial regularization. In the brainstem, GMM-PV appeared to reduce the variations previously shown by GMM, but nonetheless yielded significantly lower DCs ( $p < 0.05$ ) than the baseline method of optimal thresholding. Fig. 9(c) revealed that, without any spatial constraints, the PPM of the PV class in GMM-PV for the same subject at 37 weeks GA spread across the brainstem with little variation in space.

GMM-PV-MRF achieved high DCs consistently across all the GAs in both ROIs. The extra Gaussian modeled the intermediate PV intensities, and the spatial constraints imposed via second-order MRFs confined the PV locations at the boundary between MLS and BKG. This approach prevented the PV Gaussian distribution from shifting towards BKG and hence the spread of the PV class in space, as demonstrated in Figs. 6(c), 7(c), 8(f) and 9(f). The histograms of the binary segmentations obtained using GMM-PV-MRF in Figs. 6(d) and 7(d) highly resembled those of the manual annotations. Student's  $t$ -tests in Table 3 confirmed that GMM-PV-MRF outperformed all the other methods in both ROIs.

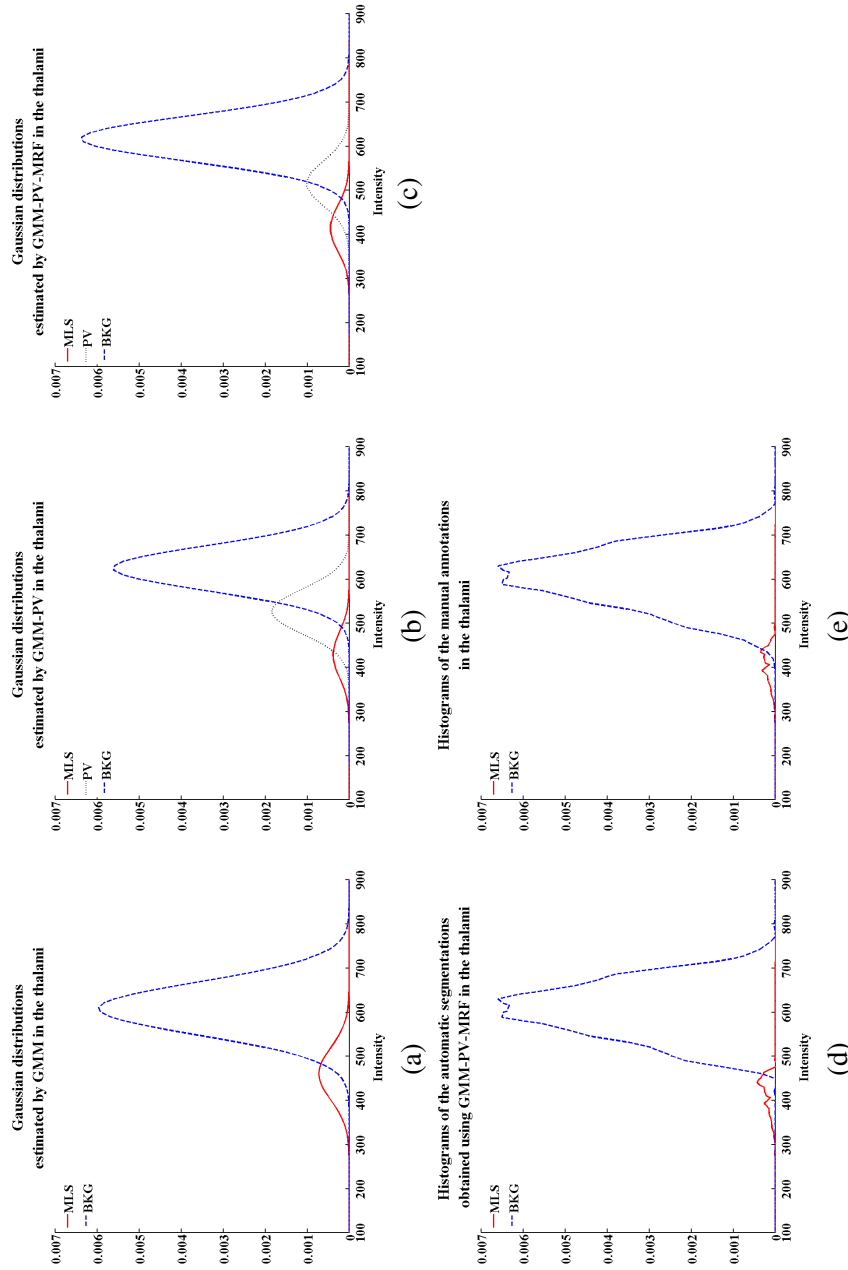


Figure 6: Gaussian distributions of the posterior probability maps estimated using the segmentation models in Table 1 for myelin-like signals (MLS), background (BKG) and partial volume voxels (PV) in the thalami of a subject at 37 gestational weeks. The sum of the graph areas for all the classes equals one. GMM over-estimates the MLS class with a shift of the mean towards BKG compared to the manual annotation histograms. GMM-PV over-estimates the PV class compared to GMM-PV-MRF due to the absence of spatial regularization, as confirmed in Fig. 8(c). The histograms of the binary segmentations obtained using GMM-PV-MRF are almost the same as those of the manual annotations.

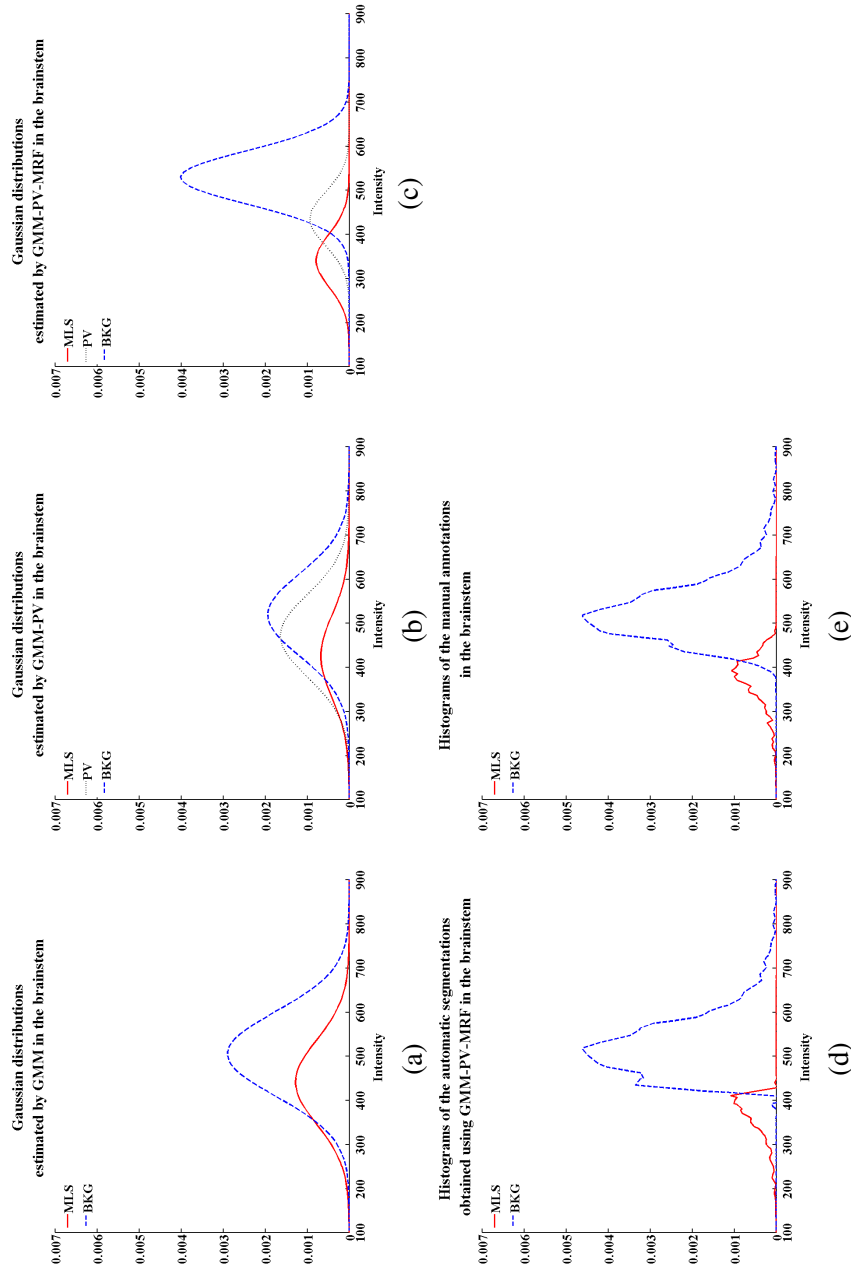


Figure 7: Gaussian distributions of the posterior probability maps estimated using the segmentation models in Table 1 for myelin-like signals (MLS), background (BKG) and partial volume voxels (PV) in the brainstem of a subject at 37 gestational weeks. The distributions and histograms are shown in the same scale as Fig. 6. The brainstem contains more PV voxels than the thalami, which can be seen from the larger overlapping area between the histograms of manually annotated MLS and BKG. EM classification only succeeds in GMM-PV-MRF, and the histograms of the resultant binary segmentations highly resemble those of the manual annotations.

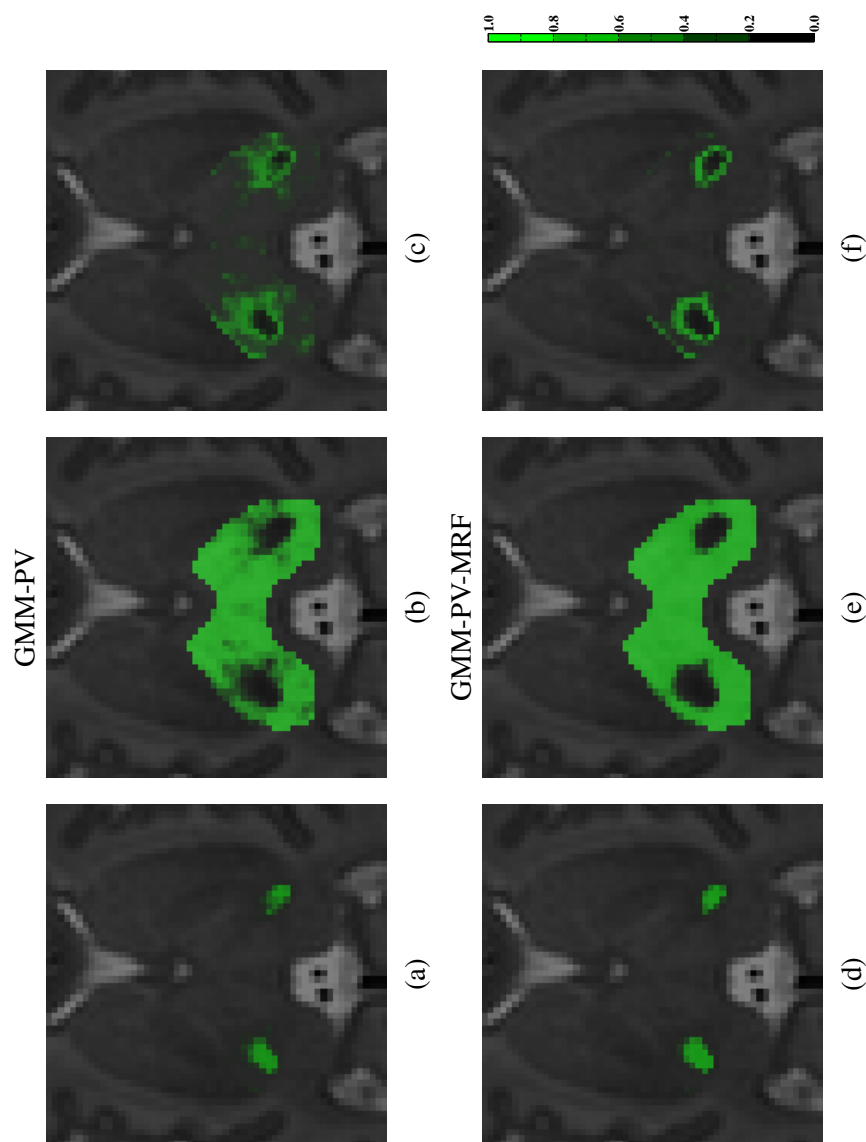


Figure 8: Posterior probability maps (PPMs) estimated using GMM-PV and GMM-PV-MRF (see Table 1 for method descriptions) in the thalami of a subject at 37 gestational weeks. The columns from left to right show the PPMs in the axial view for myelin-like signals (MLS), background (BKG) and partial volume voxels (PV). The PV class in GMM-PV spreads across the thalami due to the absence of spatial regularization. In contrast, GMM-PV-MRF configures spatial dependencies of all the classes via second-order MRFs and indicates the PV locations accurately at the boundary between MLS and BKG.

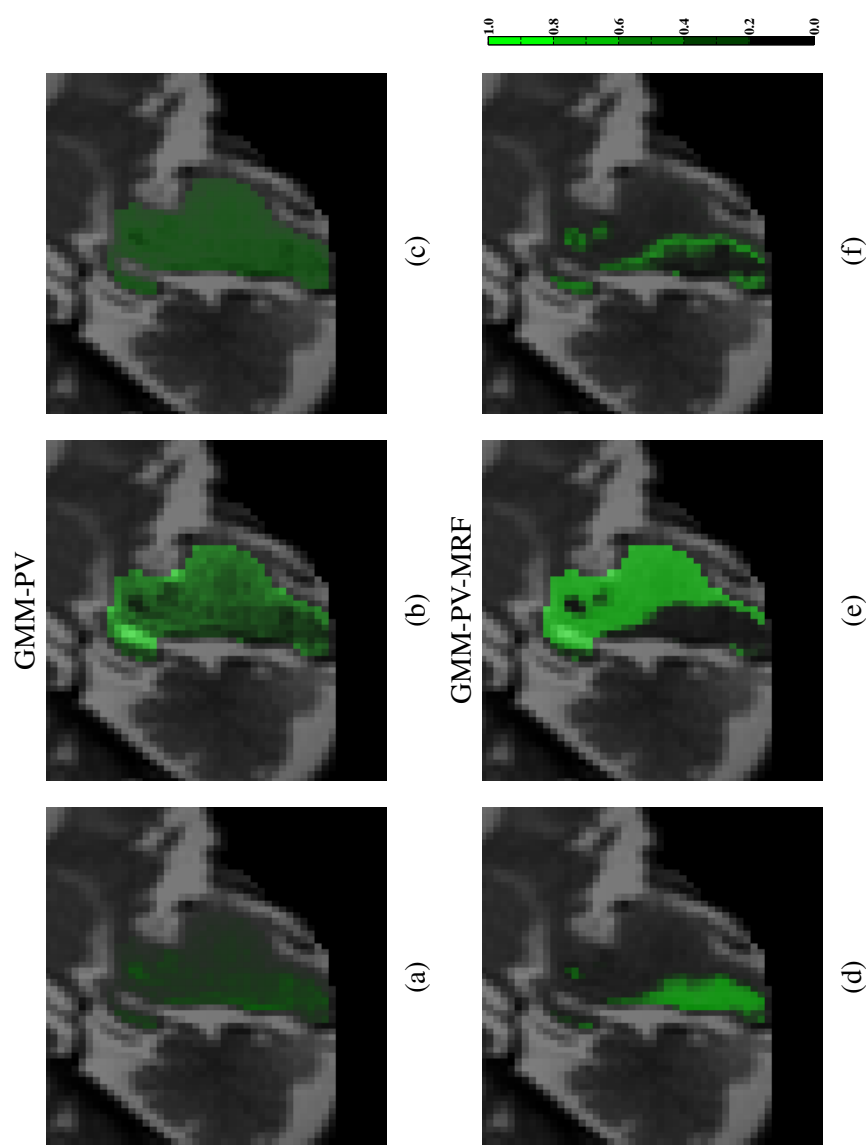


Figure 9: Posterior probability maps (PPMs) estimated using GMM-PV and GMM-PV-MRF (see Table 1 for method descriptions) in the brainstem of a subject at 37 gestational weeks. The columns from left to right show the PPMs in the sagittal view for myelin-like signals (MLS), background (BKG) and partial volume voxels (PV). GMM-PV-MRF performs accurately and consistently as in the thalami, whereas PV estimation using GMM-PV becomes much worse than in the thalami and shows little variation in space. This is consistent with the nearly overlaid Gaussians of the BKG and PV classes in Fig. 7(b).

### 4.3. *Dependence on the initial threshold*

We initialized MLS segmentations for all the methods in Table 1 using an intensity percentile threshold estimated from the volume fraction of manually annotated MLS in the ROI. However, manual reference is usually unavailable in practice, without which it is difficult to estimate the MLS content as it can vary with many factors, such as brain regions and GAs. Therefore, we continued to investigate the dependence of each method on the initial threshold.

We segmented MLS in the thalami and brainstem for the 16 test subjects by thresholding, GMM, GMM-PV and GMM-PV-MRF based on a range of initial thresholds. At each threshold, we calculated the average DC of the resultant segmentations over the subjects for each method, which was then plotted against the threshold in Figs. 10 and 11.

GMM-PV-MRF performed most stably and most accurately as the initial threshold varied from the 1st to 20th percentile in the thalami, and from the 21st to 40th percentile in the brainstem. The segmentation robustness towards the threshold is a significant advantage as the choice of the initial threshold is difficult without manual annotations, and moreover, the need for choosing the initial threshold is undesirable for practical applications. Especially when comparing the automatic MLS segmentation of an individual subject against the average growth trajectory for assessing brain maturation, it is of great importance that the volume fraction of the initial MLS estimate does not influence the final segmentation.

Thresholding, GMM and GMM-PV were particularly sensitive to the initial threshold, and performed inconsistently in different ROIs. Compared to the brainstem, the DCs obtained by thresholding underwent considerable changes in the thalami. The additional Gaussian in GMM-PV lowered the DCs of GMM in the thalami due to the absence of spatial regularization. In the brainstem with substantial PV voxels, GMM-PV helped to stabilize the variations shown by GMM, but was nonetheless unable to outperform the baseline method of thresholding.

Lastly, we verified that the 6th and 25th percentiles of the  $T_2w$  intensities in the thalami and brainstem, respectively, optimized the DCs that could possibly be achieved using a single threshold for all the 16 subjects. Nevertheless, even the optimal threshold only performed equally well as GMM-PV-MRF in the brainstem, and could not reach the DC level of GMM-PV-MRF in the thalami.

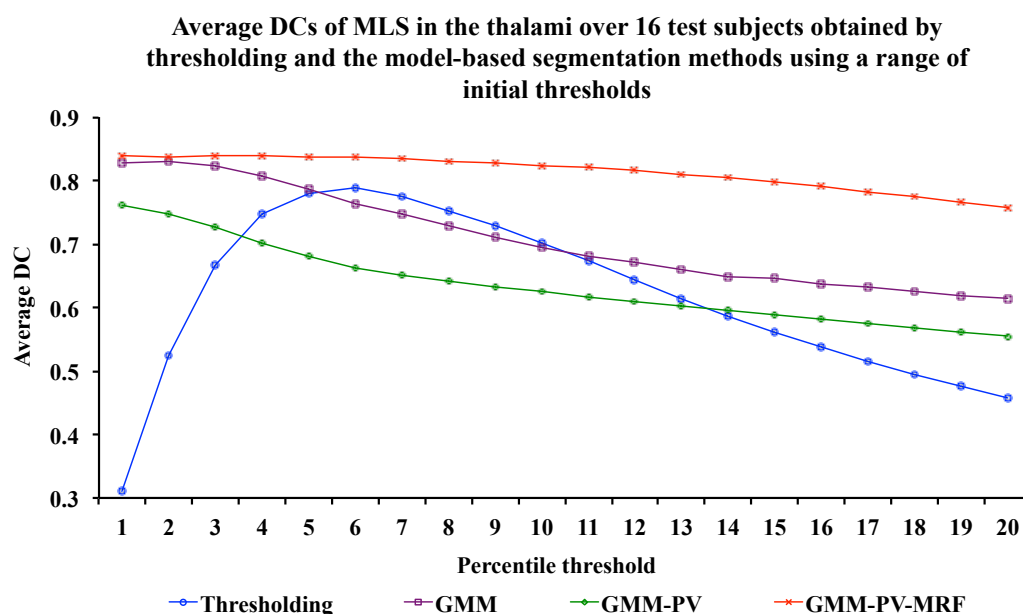


Figure 10: Average Dice coefficients (DCs) of myelin-like signals (MLS) in the thalami over 16 test subjects obtained by thresholding and the model-based methods (see Table 1 for method descriptions) using a range of initial thresholds. GMM-PV-MRF performs most robustly with the highest DCs as the threshold varies from the 1st to 20th percentile of the  $T_2w$  intensities in this region. Thresholding shows considerable changes of the DC values, and the additional Gaussian in GMM-PV lowers the DCs compared to GMM due to the absence of spatial regularization.



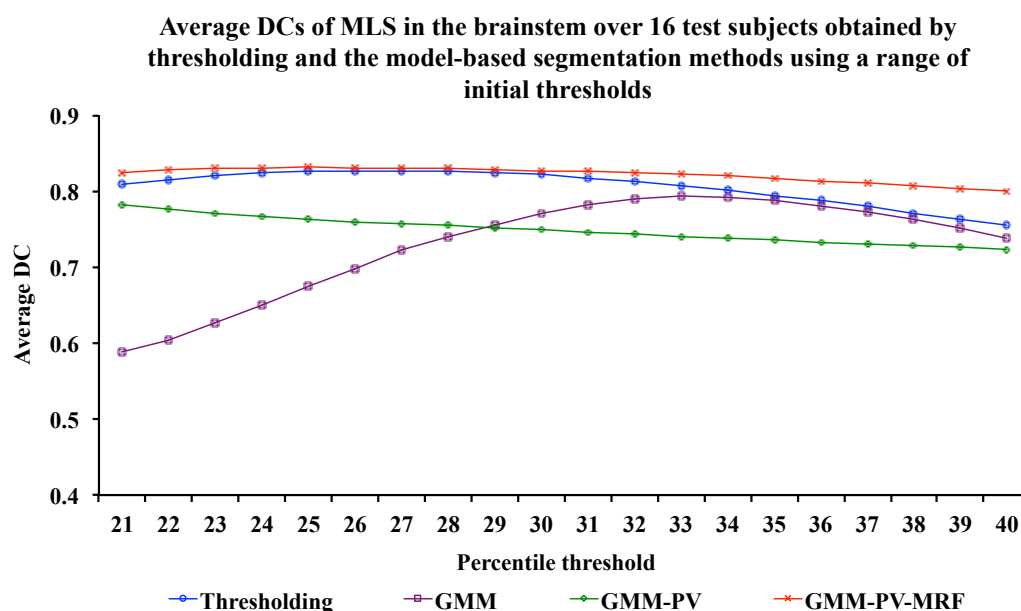


Figure 11: Average Dice coefficients (DCs) of myelin-like signals (MLS) in the brainstem over 16 test subjects obtained by thresholding and the model-based methods (see Table 1 for method descriptions) using a range of initial thresholds. GMM-PV-MRF performs most stably and most accurately with nearly constant DCs throughout the threshold range from the 21st to 40th percentile. The DCs of thresholding are much less affected by the varying threshold because the brainstem contains more myelinated tissue than the thalami. GMM-PV helps to stabilize the variations shown by GMM in the presence of substantial partial volume voxels, but nonetheless lowers the DCs of the baseline initial segmentations obtained from thresholding.

#### 4.4. Comparison to the intra-rater reliability

We assessed the intra-rater reliability as DCs between two sets of manual annotations that were available for eight of the 16 test subjects at two-week intervals between 30 and 44 weeks GA. We evaluated the automated model-based methods with respect to each manual set. The average DCs over the eight subjects are shown in Table 4 along with the  $p$ -values of two-tailed Student's  $t$ -tests with the intra-rater DCs at the 5% significance level. It can be seen that only GMM-PV-MRF performed equally well as a human rater in both ROIs, and demonstrated no significant differences with respect to the repeated manual annotations.

## 5. Applications

### 5.1. Volume growths of MLS in the ROIs

Based on the results in Section 4, we applied GMM-PV-MRF as proposed to segment MLS in the ROIs for 114 preterm infants aged between 29 and 44 weeks GA. The volumes of the resultant segmentations in the thalami and brainstem were plotted against GA in Figs. 12 and 13 respectively.

We found that the MLS volume appeared to grow exponentially with GA in the thalami, and this trend persisted in the MLS volume fraction after we corrected for the different thalamus sizes in individual subjects. In contrast, the MLS volume appeared to grow linearly with GA in the brainstem, which resulted in an approximately constant MLS volume fraction after correcting for the different brainstem sizes in individual subjects. This is because most of the brainstem appears to have completed myelination before 29 weeks GA, whereas new brain structures become myelinated in the thalami between 29 and 44 weeks GA [7, 8, 28]. Consequently, the MLS volume increased in the brainstem mainly due to the overall brain growth, and the increase in the thalami was attributed to both brain growth and appearance of new myelination. In the next section, we demonstrate the different myelination progressions in the ROIs by constructing spatio-temporal growth models for MLS.

Method	Manual set 1		Manual set 2		Intra-rater reliability
	Average DC	<i>p</i> -value	Average DC	<i>p</i> -value	
GMM	0.785 ± 0.085	0.257	0.754 ± 0.054	0.023	0.834 ± 0.043
GMM-PV	0.662 ± 0.210	0.060	0.626 ± 0.217	0.028	
GMM-PV-MRF	0.843 ± 0.051	0.624	0.816 ± 0.046	0.413	

(a) Comparison results in the thalami

Method	Manual set 1		Manual set 2		Intra-rater reliability
	Average DC	<i>p</i> -value	Average DC	<i>p</i> -value	
GMM	0.702 ± 0.171	0.046	0.695 ± 0.156	0.026	0.842 ± 0.034
GMM-PV	0.763 ± 0.047	0.002	0.766 ± 0.077	0.031	
GMM-PV-MRF	0.832 ± 0.034	0.608	0.815 ± 0.024	0.087	

(b) Comparison results in the brainstem

Table 4: Average Dice coefficients (DCs) ( $\pm$  standard deviations) over eight test subjects obtained by thresholding and the model-based segmentation methods (see Table 1 for method descriptions) with respect to two sets of manual annotations compared to the intra-rater reliability. The *p*-values indicate the outcomes of two-tailed Student's *t*-tests with the intra-rater DCs at the 5% significance level. The performance of GMM-PV-MRF is close to the intra-rater reliability in both regions of interest.

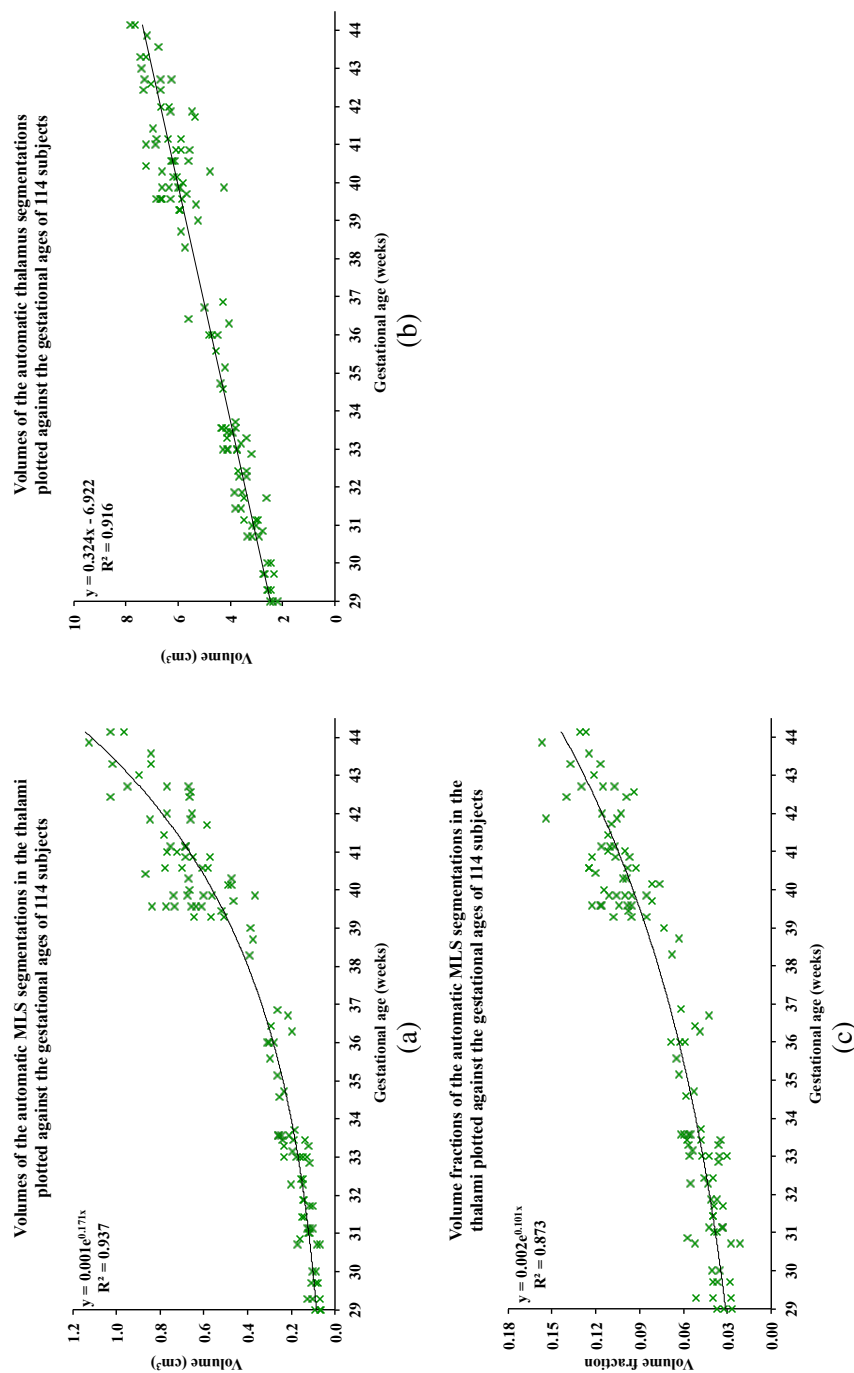


Figure 12: (a) Volumes of the automatic segmentations for myelin-like signals (MLS) in the thalami obtained using the proposed method GMM-PV-MRF, (b) volumes of the automatic thalamus segmentations obtained during image preprocessing, and (c) volume fractions of MLS plotted against the gestational ages (GAs) of 114 subjects between 29 and 44 weeks GA. The MLS volume appears to grow exponentially with GA, and this trend persists in the MLS volume fraction after correcting for the different sizes of the thalami in individual subjects.

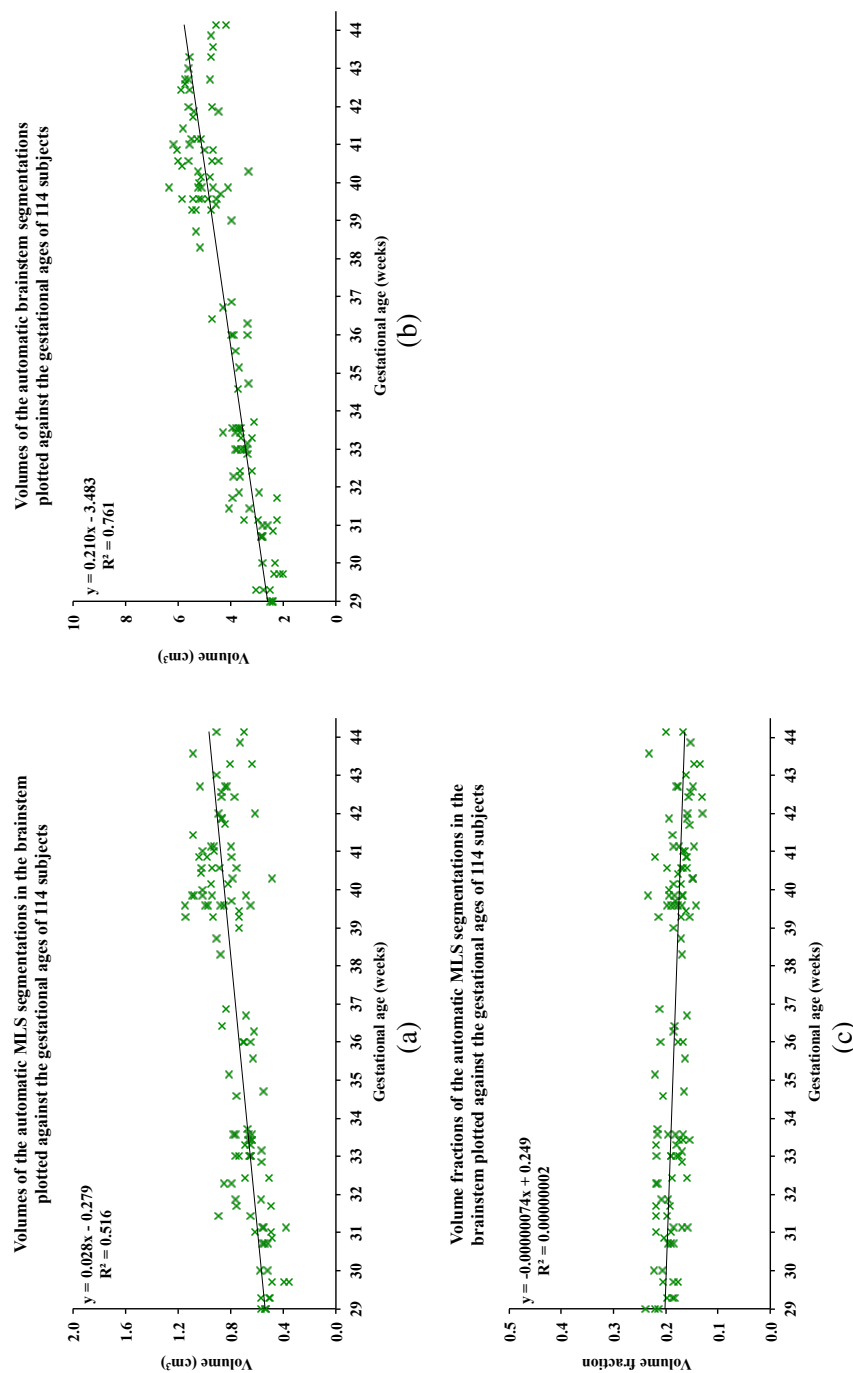


Figure 13: (a) Volumes of the automatic segmentations for myelin-like signals (MLS) in the brainstem obtained using the proposed method GMM-PV-MRF, (b) volumes of the automatic brainstem segmentations obtained during image preprocessing, and (c) volume fractions of MLS plotted against the gestational ages (GAs) of 114 subjects between 29 and 44 weeks GA. The MLS volume appears to grow linearly with GA, which results in an approximately constant MLS volume fraction after correcting for the different sizes of the brainstem in individual subjects.

## 5.2. Spatio-temporal growth models for MLS in the ROIs

In order to assess brain development in preterm neonates, we constructed MLS growth models in the thalami and brainstem using voxelwise logistic regression based on the automatic segmentations computed using GMM-PV-MRF for the 114 subjects. This part of our work has been published in the *Proceedings of the 13th International Symposium on Biomedical Imaging (ISBI)* [37].

First we registered the T<sub>2</sub>w image of each individual subject with the dilated ROI of the single reference subject at 36 weeks GA using FFD non-rigid registration [27]. This was the same reference subject used to create the ROI masks of the thalami in Section 3.1. We used normalized mutual information (NMI) as the similarity measure and 10 mm control point spacing. The automatic MLS segmentations were then transformed accordingly from each subject's space to the common reference space. Lastly, we constructed the growth model in each ROI by fitting a voxelwise logistic function to the transformed segmentations.

We found that there was an evident difference in the onset and rate of myelination in the two ROIs. Fig. 14 showed that the VLN and STN in the thalami were likely to be myelinated before 29 weeks GA. Moreover, the model captured the arrival of MLS in both of the PLIC tracts at approximately 40 weeks GA, which confirmed the clinically observed time point [8]. Myelination of the PLIC tracts is an important landmark for evaluating neonatal brain development [5]. The brainstem, however, appeared to be well developed by the time of 29 weeks GA, and there was relatively little new myelination in the age range of our investigation. Note that because we corrected for the different sizes of each ROI in individual subjects through non-rigid registration [27], the simulated MLS progression was entirely due to the appearance of new myelinated brain structures without contribution from the overall brain growth.

We further used the MLS growth models in the thalami and brainstem to predict GAs of the 114 preterm infants. The age estimates were determined by minimizing the sum of squared differences (SSD) between each individual transformed segmentation and the average growth model constructed in a leave-one-out procedure. The estimated GAs were plotted against the nominal values in Fig. 15. We obtained root mean squared errors (RMSEs) of 1.41 weeks and 2.56 weeks in the thalami and brainstem respectively. Therefore, each ROI has a different predictive power that best assesses a particular period of brain development. The thalami with the most prominent MLS growth between 29 and 44 weeks GA produced the most accurate age estimates for preterm infants in this age range. Details of our work in spatio-temporal modeling and age estimation can be found in [37].

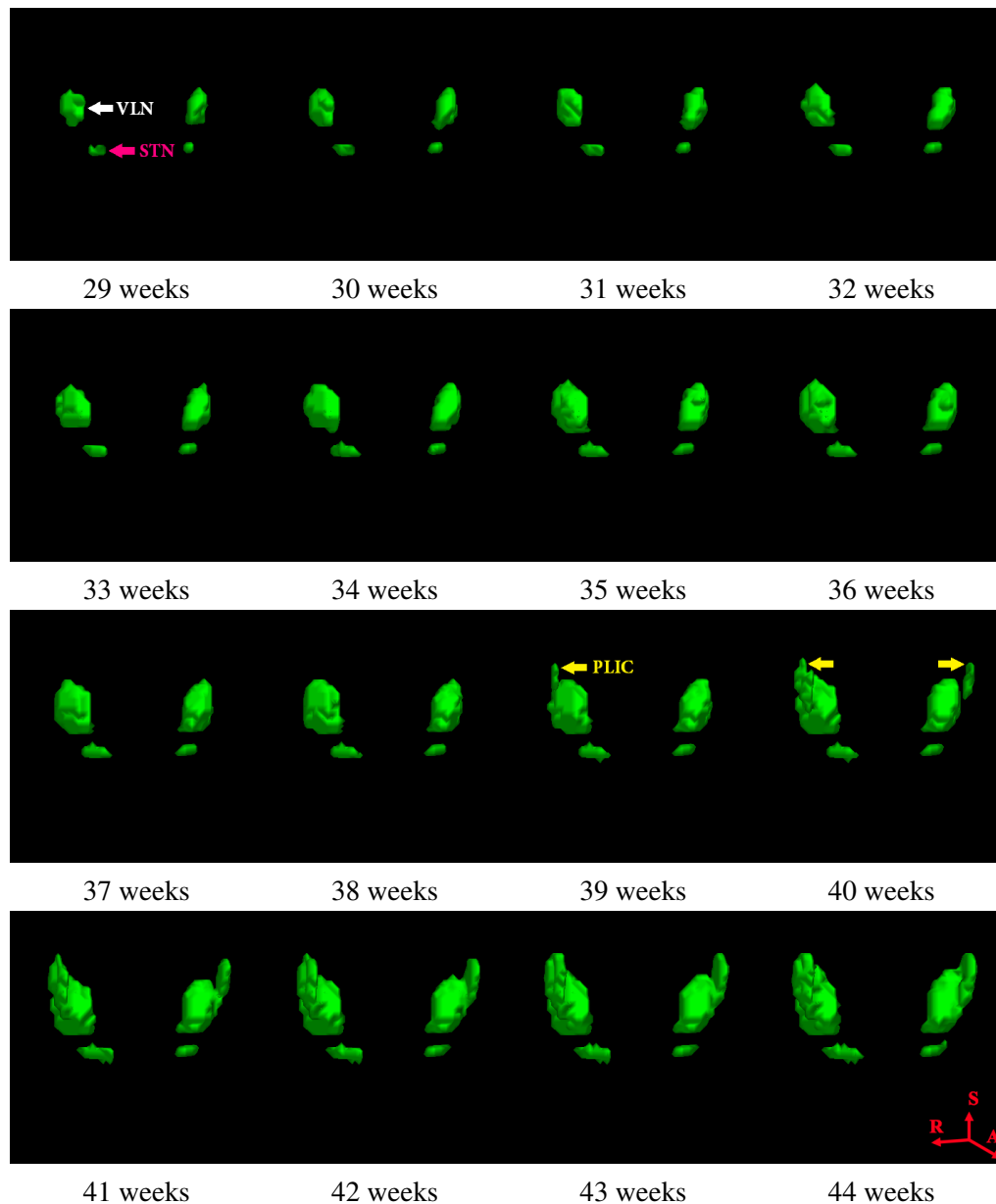


Figure 14: Spatio-temporal growth model for myelin-like signals (MLS) in the thalami between 29 and 44 weeks gestational age (GA). The ventrolateral nuclei (VLN) and subthalamic nuclei (STN) appear to be myelinated before 29 weeks GA. MLS becomes evident in both tracts of the posterior limbs of the internal capsule (PLIC) at approximately 40 weeks GA. This is an important landmark for evaluating neonatal brain development. As the different thalamus sizes in individual subjects have been corrected through non-rigid registration, the progression of MLS is completely due to the appearance of new myelination without contribution from the overall brain growth.

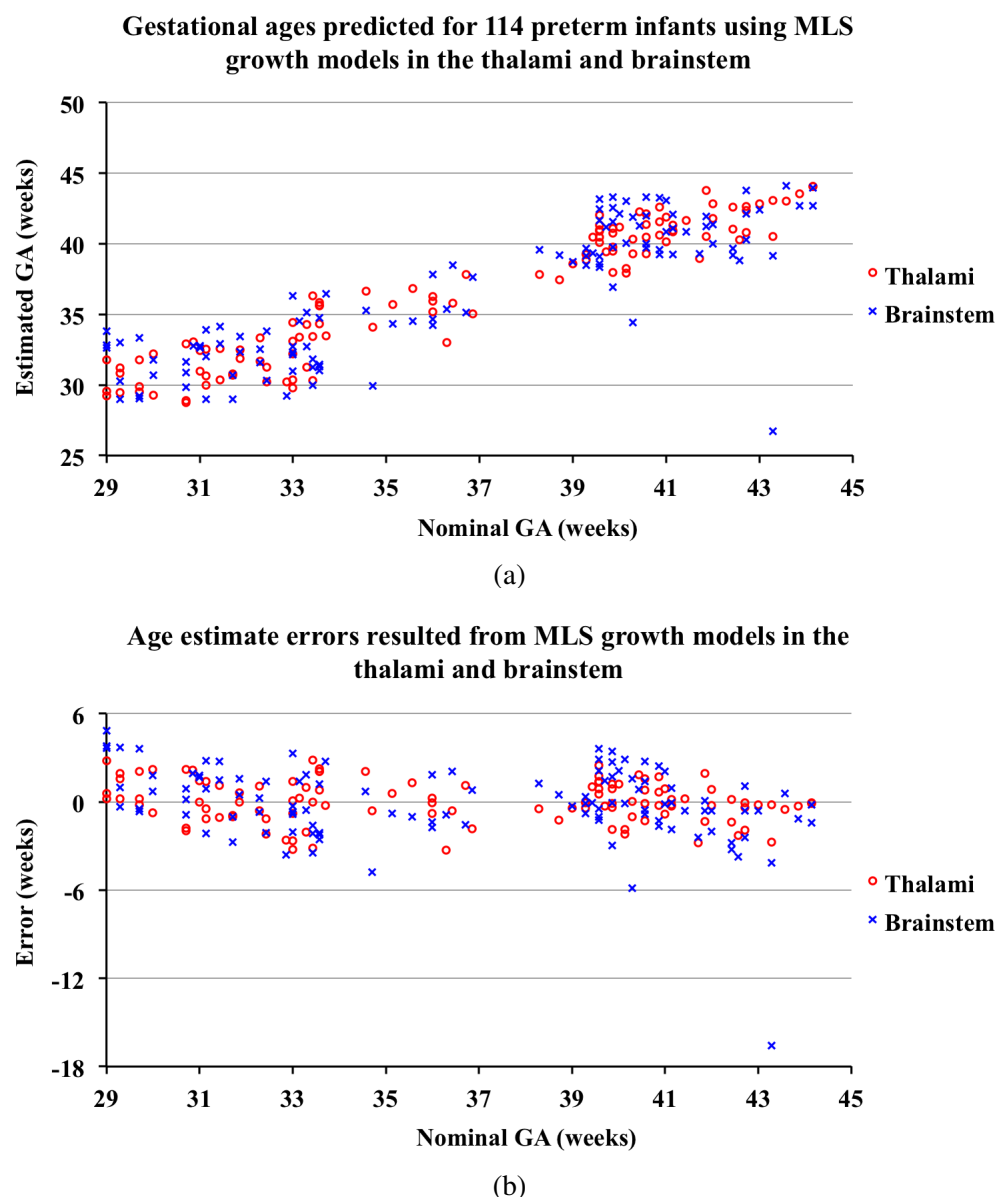


Figure 15: (a) Gestational ages (GAs) predicted for 114 preterm infants and (b) age estimate errors compared to the nominal GAs using the logistic growth models for myelin-like signals (MLS) in the thalami and brainstem. The predictions based on MLS in the brainstem display larger deviations from the nominal GAs than in the thalami. This is because most of the MLS spatio-temporal changes between 29 and 44 weeks GA occur in the thalami, whereas the brainstem provides much less information on progressing myelination during this particular age period.



## 6. Discussion and Conclusion

In this work, we presented an EM segmentation framework for MLS on  $T_2w$  brain MR images of preterm neonates. The two major challenges of this task are the inadequate spatial prior information and the large number of PV voxels typical for the developing brain. The publicly available neonatal brain atlases [18, 30] do not include a probability map for myelin, and manual annotations are impractical in clinical use or for large-scale researches. The intermediate intensities of PV voxels containing both MLS and BKG make the EM optimization process unstable, and cause over-estimation of the MLS class.

To overcome these challenges, we introduced an additional Gaussian to explicitly model the PV voxels. Van Leemput *et al.* [36] observed that, without any spatial prior information, the extra Gaussian hampers segmentation robustness. The previous PV estimation methods thus used prior knowledge of the spatial locations of the composing pure tissues to guide the search for PV voxels [6, 21, 23, 31, 33, 36]. Because there are no probabilistic or manual atlases available for MLS, we identified the PV locations entirely based on tissue connectivities among the triplet of classes in the local neighborhood via second-order MRFs. This approach performed robustly and accurately with respect to manually annotated MLS in different ROIs for subjects between 29 and 44 weeks GA, regardless of the volume fraction used to obtain the initial MLS segmentations. This is crucial in practice as the choice of the initial threshold is difficult without manual annotations. The independence of the automatic MLS segmentations against the initial threshold is also important when assessing deviation from the average MLS growth model in order to estimate the brain maturation status.

In the future, we will explore multi-spectral imaging data for MLS segmentation. Myelinated GM and WM structures are shown differently using  $T_1w$  and  $T_2w$  sequences [4, 8]. In general, myelinated GM structures, such as the thalamic and brainstem nuclei, are more conspicuous on  $T_2w$  images whereas myelinated WM structures, particularly the axonal tracts, are better visualized on  $T_1w$  images [4]. Future work will include developing a multivariate EM framework that incorporates both  $T_1w$  and  $T_2w$  images.

In terms of the applications, we demonstrated that the proposed method for MLS segmentation provides the basis for a variety of automatic analyses that can track brain development in fragile preterm neonates using routinely acquired MR images. In particular, age estimation using the average MLS growth model can be useful in clinical practice for infants with subtle brain maturational abnormalities, who would benefit from early medical interventions soon after birth.

## Appendix

Here we derive the M-step for the proposed segmentation method (GMM-PV-MRF) where we model myelin-like signals (MLS), partial volume (PV) voxels and background (BKG) in the ROI via second-order MRFs.

We assume that the observed intensities  $\mathbf{y}$  are generated by the hidden labels  $\mathbf{z}$  with Gaussian probability density functions. The Gaussian parameters  $\Phi_y$  consist of mean  $\mu_k$  and standard deviation (SD)  $\sigma_k$  of class  $k$  that are updated in the M-step. The hidden labels are assumed to be a realization of the random process characterized by the MRF parameters  $\Phi_z$ .

Denoting  $\Phi = \{\Phi_y, \Phi_z\}$ , Van Leemput *et al.* [35] showed that the conditional expectation of the log likelihood for the complete data  $\{\mathbf{y}, \mathbf{z}\}$  can be written as:

$$\begin{aligned} Q(\Phi | \Phi^{(m)}) &= E \left[ \ln p(\mathbf{y}, \mathbf{z} | \Phi) | \mathbf{y}, \Phi^{(m)} \right] \\ &= \sum_{l=1}^K \sum_{i=1}^N p_{il}^{(m+1)} \ln G(y_i, \mu_l, \sigma_l) \\ &\quad - E \left[ U_{\text{MRF}}(\mathbf{z} | \Phi_z) + \ln f(\Phi_z) | \mathbf{y}, \Phi^{(m)} \right] \end{aligned} \quad (14)$$

where

$$\ln G(y_i, \mu_l, \sigma_l) = -\frac{1}{2} \ln(2\pi\sigma_l^2) - \frac{(y_i - \mu_l)^2}{2\sigma_l^2}$$

The energy function  $U_{\text{MRF}}(\mathbf{z} | \Phi_z)$  is computed using Eq. 10 and  $f(\Phi_z)$  denotes a normalization constant summed over all possible configurations of the hidden labels. Explanations of the other symbols can be found in Section 3.2. Throughout the EM iterations, the MRF parameters  $\Phi_z$  remain constant as assigned in the connectivity tensor (Fig. 5), and hence do not contribute to the optimization of the Gaussian parameters  $\Phi_y$  which can be estimated in the same way as a GMM with independent voxels.

We approximate the PV class mean as the arithmetic mean of the composing tissue means (Eq. 5) and assume that the SDs of the MLS, PV and BKG classes are identical (Eq. 6). Maximization of  $Q(\Phi | \Phi^{(m)})$  with respect to  $\sigma_k^2$  by substituting Eq. 6 into Eq. 14 yields Eq. 3 for updating the identical SD. We then differentiate  $Q(\Phi | \Phi^{(m)})$  with respect to each of the composing tissue means by substituting

Eq. 5 into Eq. 14, and obtain the following system of linear equations:

$$\begin{aligned} \frac{1}{4}\mu_{\text{MLS}} \sum_{i=1}^N p_{i\text{PV}} + \mu_{\text{BKG}} \sum_{i=1}^N \left( p_{i\text{BKG}} + \frac{1}{4}p_{i\text{PV}} \right) &= \sum_{i=1}^N y_i \left( p_{i\text{BKG}} + \frac{1}{2}p_{i\text{PV}} \right) \\ \mu_{\text{MLS}} \sum_{i=1}^N \left( p_{i\text{MLS}} + \frac{1}{4}p_{i\text{PV}} \right) + \frac{1}{4}\mu_{\text{BKG}} \sum_{i=1}^N p_{i\text{PV}} &= \sum_{i=1}^N y_i \left( p_{i\text{MLS}} + \frac{1}{2}p_{i\text{PV}} \right) \end{aligned}$$

We omit the iteration number in the posterior probabilities for clarity. Lastly, we solve the linear equations via matrix inversion, as shown in Eq. 7.

## Acknowledgement

S. Wang is a recipient of the Oxford Clarendon Fund Scholarship which is provided by the Oxford University Press through the Life Sciences Interface Doctoral Training Center. The authors would like to thank the Center for the Developing Brain, King's College London, UK, for sharing the MRI data used in this study.

## References

- [1] Alonso-Ortiz, E., Levesque, I., Pike, G., 2015. MRI-based myelin water imaging: A technical review. *Magn. Reson. Med.* 73, 70–81.
- [2] Anbeek, P., Vincken, L., Groenendaal, F., Koeman, A., Van Osch, M., Van der Grond, J., 2008. Probabilistic brain tissue segmentation in neonatal magnetic resonance imaging. *Pediatr. Res.* 63 (2), 158–163.
- [3] Ashburner, J., Friston, K., 2005. Unified segmentation. *NeuroImage* 26, 839–851.
- [4] Barkovich, A., 1998. MR of the normal neonatal brain: Assessment of deep structures. *Am. J. Neuroradiol.* 19, 1397–1403.
- [5] Bird, C., Hedberg, M., Drayer, B., Keller, P., Flom, R., Hodak, J., 1989. MR assessment of myelination in infants and children: Usefulness of marker sites. *Neuroradiology* 10, 731–740.
- [6] Cardoso, M., Clarkson, M., Ridgway, G., Modat, M., Fox, N., Ourselin, S., the Alzheimer's Disease Neuroimaging Initiative, 2011. LoAd: A locally adaptive cortical segmentation algorithm. *NeuroImage* 56, 1386–1397.

- [7] Chi, J., Dooling, E., Gilles, F., 1977. Gyral development of the human brain. *Ann. Neurol.* 1, 86–93.
- [8] Counsell, S., Maalouf, E., Fletcher, A., Duggan, P., Battin, M., Lewis, H., Herlihy, A., Edwards, A., Bydder, G., Rutherford, M., 2002. MR imaging assessment of myelination in the very preterm brain. *Am. J. Neuroradiol.* 23, 872–881.
- [9] Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 39 (1), 1–38.
- [10] Dice, L., 1945. Measures of the amount of ecologic association between species. *Ecology* 23 (3), 297–302.
- [11] Dubois, J., Dehaene-Lambertz, G., Kulikova, S., Poupon, C., Huppi, P., Hertz-Pannier, L., 2014. The early development of brain white matter: A review of imaging studies in fetuses, newborns and infants. *Neuroscience* 276, 48–71.
- [12] Fornari, E., Knyazeva, M., Meuli, R., Maeder, P., 2007. Myelination shapes functional activity in the developing brain. *NeuroImage* 38, 511–518.
- [13] Gousias, I., Edwards, A., Rutherford, M., Counsell, S., Hajnal, J., Rueckert, D., Hammers, A., 2012. Magnetic resonance imaging of the newborn brain: Manual segmentation of labelled atlases in term-born and preterm infants. *NeuroImage* 62, 1499–1509.
- [14] Gui, L., Lisowski, R., Faundez, T., Huppi, P., Lazeyras, F., Kocher, M., 2012. Morphology-driven automatic segmentation of MR images of the neonatal brain. *Med. Image Anal.* 16 (8), 1565–1579.
- [15] Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* 33, 115–126.
- [16] Henkelman, R., Stanisz, G., Graham, S., 2001. Magnetization transfer in MRI: A review. *NMR Biomed.* 14, 57–64.
- [17] Isgum, I., Benders, M., Avants, B., Cardoso, M., Counsell, S., Gomez, E., Gui, L., Huppi, P., Kersbergen, K., Makropoulos, A., Melbourne, A.,

- Moeskops, P., Mol, C., Kuklisova-Murgasova, M., Rueckert, D., Schnabel, J., Srhoj-Egekher, V., Wu, J., Wang, S., De Vries, L., Miergever, M., 2015. Evaluation of automatic neonatal brain segmentation algorithms: The Neo-BrainS12 challenge. *Med. Image Anal.* 20 (1), 135–151.
- [18] Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S., Doria, V., Serag, A., Gousias, I., Boardman, J., Rutherford, M., Edwards, A., Hajnal, J., Rueckert, D., 2011. A dynamic 4D probabilistic atlas of the developing brain. *NeuroImage* 54, 2750–2763.
- [19] Laule, C., Vavasour, I., Kolind, S., Li, D., Traboulsee, T., Moore, G., MacKay, A., 2007. Magnetic resonance imaging of myelin. *Neurotherapeutics* 4, 460–484.
- [20] Le Bihan, D., Mangin, J., Poupon, C., Clark, C., Pappata, S., Molko, N., Chabriet, H., 2001. Diffusion tensor imaging: Concepts and applications. *J. Magn. Reson. Imaging* 13, 534–546.
- [21] Ledig, C., Wright, R., Serag, A., Aljabar, P., Rueckert, D., 2012. Neonatal brain segmentation using second order neighborhood information. In: *Proceedings of the MICCAI Workshop: Perinatal and Paediatric Imaging*. pp. 33–40.
- [22] Makropoulos, A., Gousias, I., Ledig, C., Aljabar, P., Serag, A., Hajnal, J., Edwards, A., Counsell, S., Rueckert, D., 2014. Automatic whole brain MRI segmentation of the developing neonatal brain. *IEEE Trans. Med. Imag.* 33 (9), 1818–1831.
- [23] Manjon, J., Tohka, J., Robles, M., 2010. Improved estimates of partial volume coefficients from noisy brain MRI using spatial context. *NeuroImage* 53, 480–490.
- [24] Melbourne, A., Cardoso, M., Kendall, G., Robertson, N., Marlow, N., Ourselin, S., 2012. NeobrainS12 challenge: Adaptive neonatal MRI brain segmentation with myelinated white matter class and automated extraction of ventricles I-IV. In: *Proceedings of the MICCAI Grand Challenge: Neonatal Brain Segmentation*. pp. 16–21.
- [25] Prastawa, M., Gilmore, J., Lin, W., Gerig, G., 2005. Automatic segmentation of MR images of the developing newborn brain. *Med. Image Anal.* 9, 457–466.

- [26] Rorke, L., Riggs, H., 1969. Myelination of the brain in the newborn. Lippincott, Philadelphia.
- [27] Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D., 1999. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. Med. Imag.* 18 (8), 712–721.
- [28] Rutherford, M., 2002. MRI of the Neonatal Brain. Saunders, Edinburgh.
- [29] Rutherford, M., Srinivasan, L., Dyet, L., Ward, P., Allsop, J., Counsell, S., Cowan, F., 2006. Magnetic resonance imaging in perinatal brain injury: Clinical presentation, lesions and outcome. *Pediatr. Radiol.* 36, 582–592.
- [30] Serag, A., Aljabar, P., Ball, G., Counsell, S., Boardman, J., Rutherford, M., Edwards, A., Hajnal, J., Rueckert, D., 2012. Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. *NeuroImage* 59, 2255–2265.
- [31] Shattuck, D., Sandor-Leahy, S., K.A., S., Rottenberg, D., Leahy, R., 2001. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage* 13, 856–876.
- [32] Srinivasan, L., Rutherford, M., 2008. MRI of the newborn brain. *Paediatr. Child Health* 18 (4), 183–195.
- [33] Tohka, J., Zijdenbos, A., Evans, A., 2004. Fast and robust parameter estimation for statistical partial volume models in brain MRI. *NeuroImage* 23, 84–97.
- [34] Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based bias field correction of MR images of the brain. *IEEE Trans. Med. Imag.* 18 (10), 885–896.
- [35] Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imag.* 18 (10), 897–908.
- [36] Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 2003. A unifying framework for partial volume segmentation of brain MR images. *IEEE Trans. Med. Imag.* 22 (1), 105–119.

- [37] Wang, S., Kuklisova-Murgasova, M., Hajnal, J., Ledig, C., Schnabel, J., 2016. Regression analysis for assessment of myelination status in preterm brains with magnetic resonance imaging. In: Proceedings of the 13th International Symposium on Biomedical Imaging (ISBI). pp. 278–281.
- [38] Wang, S., Kuklisova-Murgasova, M., Schnabel, J., 2012. An atlas-based method for neonatal MR brain tissue segmentation. In: Proceedings of the MICCAI Grand Challenge: Neonatal Brain Segmentation. pp. 28–35.
- [39] Wells, W., Grimson, W., Kikinis, R., Jolesz, F., 1996. Adaptive segmentation of MRI data. *IEEE Trans. Med. Imag.* 15 (4), 429–442.
- [40] Wu, J., Avants, B., 2012. Automatic registration-based segmentation for neonatal brains using ANTs and Atropos. In: Proceedings of the MICCAI Grand Challenge: Neonatal Brain Segmentation. pp. 36–43.
- [41] Xue, H., Srinivasan, L., Jiang, S., Rutherford, M., Edwards, A., Rueckert, D., Hajnal, J., 2007. Automatic segmentation and reconstruction of the cortex from neonatal MRI. *NeuroImage* 38, 461–477.
- [42] Yakolev, P., Lecours, A., 1967. The myelogenetic cycles of regional maturation of the brain. In: Mankowski, A. (Ed.), *Regional Development of the Brain in Early Life*. Blackwell, Oxford, pp. 3–70.
- [43] Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field and the expectation-maximization algorithm. *IEEE Trans. Med. Imag.* 20 (1), 45–57.