Fully Bayesian Inference for Structural MRI: Application to Segmentation and Statistical Analysis of T2-Hypointensities

Paul Schmidt[1,2,3], Volker J. Schmid[3], Christian Gaser[4,5], Dorothea Buck[1,2], Susanne Bührlen[1,2], Annette Förschler[2,6], Mark Mühlau[1,2]*

1 Department of Neurology, Technische Universität München, Munich, Germany, 2 TUM-Neuroimaging Center, Technische Universität München, Munich, Germany, 3 Department of Statistics, Ludwig-Maximilians-Universität München, Munich, Germany, 4 Department of Psychiatry, Jena University Hospital, Jena, Germany, 5 Department of Neurology, Jena University Hospital, Jena, Germany, 6 Department of Neuroradiology, Technische Universität München, Munich, Germany

Abstract

Aiming at iron-related T2-hypointensity, which is related to normal aging and neurodegenerative processes, we here present two practicable approaches, based on Bayesian inference, for preprocessing and statistical analysis of a complex set of structural MRI data. In particular, Markov Chain Monte Carlo methods were used to simulate posterior distributions. First, we rendered a segmentation algorithm that uses outlier detection based on model checking techniques within a Bayesian mixture model. Second, we rendered an analytical tool comprising a Bayesian regression model with smoothness priors (in the form of Gaussian Markov random fields) mitigating the necessity to smooth data prior to statistical analysis. For validation, we used simulated data and MRI data of 27 healthy controls (age: 30 ± 9; range, $20 - 58$). We first observed robust segmentation of both simulated T2-hypointensities and gray-matter regions known to be T2-hypointense. Second, simulated data and images of segmented T2-hypointensity were analyzed. We found not only robust identification of simulated effects but also a biologically plausible age-related increase of T2-hypointensity primarily within the dentate nucleus but also within the globus pallidus, substantia nigra, and red nucleus. Our results indicate that fully Bayesian inference can successfully be applied for preprocessing and statistical analysis of structural MRI data.

Introduction

This work was motivated by the aim to analyze iron-related T2-hypointensity automatically in a complex set of MRI data. Increased iron content within deep gray matter (GM) regions decreases the T2-weighted MRI signal. It has been demonstrated in both normal aging and several neurodegenerative conditions [1]. Therefore, GM T2-hypointensity may be a marker of early neurodegeneration and has even been regarded of potential predictive value in neurological diseases such as Multiple Sclerosis [2]. Intriguingly, increased iron content goes not only along with a remarkable signal loss in T2-weighted sequences (i.e. T2-hypointensity) but also blurs the differences in signal intensities between GM and white matter (WM) in T1-weighted sequences. Our MRI data sets included three sequences: gradient-echo T1-weighted and two different T2-weighted images. We decided to include all three sequences as their information is complementary. T1-weighted sequences are suitable for well established normalization pipelines. Besides the sensitivity to T2-hypointensity, one of the T2-weighted sequences was of high contrast but low image definition (FLAIR), while the other was of low contrast but high

image definition (T2-weighted). Further, accessibility of these data to scientific investigations has been desirable as this MRI protocol has been used in routine clinical practice so that a still growing data base, including several large patient groups, has been available at our institution and cooperating institutions. Besides the fact that initially non-Bayesian approaches had failed with regard to both segmentation [3] and statistical analysis [4], we decided to develop algorithms based on Bayesian inference as some inherent features may be advantageous for preprocessing and statistical analysis of structural neuroimaging data [5]. For example, more realistic modeling of complex data is possible by incorporation of prior knowledge, and results do not have to be corrected for multiple statistical tests *post hoc*.

First, we developed a segmentation algorithm for the localization of T2-hypointensities by using outlier detection based on model checking techniques within a Bayesian mixture model. We used Markov Chain Monte Carlo (MCMC) methods for both model fitting and checking, as they allowed to incorporate not only the uncertainty of the data but also the uncertainty of model parameters, which often leads to results that are more realistic

than those based on point estimates only [6]. The result of this segmentation tool was validated by simulated data and by visual inspection through relating segmented T2-hypointensities to anatomical GM regions known to contain iron above average.

Second, we adapted a Bayesian voxel-wise regression model that included spatial information during the estimation step via smoothness priors to mitigate the necessity to smooth images before statistical analysis. As in the mixture model and suggested earlier [7,8], we also used MCMC methods for model fitting. Here, we had to adapt and extend earlier approaches designed for functional MRI (fMRI) data. Either those models [9,10] were designed for first level analyses requiring an autoregressive component of the data so that they could not be applied directly to structural MRI data, or second level models did not account for the spatial structure of voxels so that images still had to be smoothed prior to statistical analysis. For validation, we used simulated data and data from healthy controls. We compared the results derived from our approach to those derived from standard software using either frequentist or Bayesian inference.

Materials and Methods

Subjects

MRI scans were obtained from 27 subjects (female, 17; age in years: range, 20–58; median, 29; mean standard deviation, 30.48 ± 9.1) that had served as healthy controls in an MRI study (Departments of Neurology and Neuroradiology, Technische Universität München, Munich, Germany). Beforehand, written informed consent was obtained after description of the study to the subjects. The study was approved by the ethics committee of the medical faculty of the Technische Universität München, Munich, Germany, and performed in accord with the Declaration of Helsinki.

Magnetic resonance imaging

All brain images were acquired on the same 3 Tesla scanner (Achieva, Philips, Netherlands). A three-dimensional (3D) GRE T1-weighted sequence (orientation, 170 contiguous sagittal 1 mm slices; field of view, 240×240 mm^2; voxel size, $1.0 \times 1.0 \times 1.0$ mm^3; repetition time (TR), 9 ms; echo time (TE), 4 ms), 3D T2-weighted sequence (orientation, 144 contiguous sagittal 1.5 mm slices; field of view, 230×172 mm^2; voxel size, $1.0 \times 1.0 \times 1.5$ mm^3; TR, 4000 ms; TE, 35 ms) and a 3D FLAIR sequence (orientation, 144 contiguous axial 1.5 mm slices; field of view, 230×185 mm^2; voxel size, $1.0 \times 1.0 \times 1.5$ mm^3; TR, 10^4 ms; TE, 140 ms; inversion time, 2750 ms) were used.

Preprocessing

In this section, we describe preprocessing steps of our data with freely available software. SPM8 (http://www.fil.ion.ucl.ac.uk/spm) and its VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm) were used. An overview is given in Figure 1. First, T2-weighted and FLAIR images are coregistered to the T1-weighted images in the original ('native') space. These images are then prepared for the segmentation of T2-hypointensities, which includes correction of T2-weighted and FLAIR images for magnetic field inhomogeneity by VBM8 (function 'estimate and write', default option; output option, 'bias corrected' and 'native space') and segmentation of T1-weighted images into the tissue classes of GM, WM, and cerebrospinal fluid (CSF) (function 'estimate and write', default option) as this information is necessary to segment hypointensities as explained in the next section. The resulting images of segmented T2-hypointensity (T2-hypointensity images) are normalized in two steps: First, T1-weighted images are affine normalized and respective parameters applied to FLAIR and T2-hypointensity images. Second, affine normalized T1-weighted and FLAIR images of all subjects are used to produce individual flow fields by high-dimensional warping as implemented in SPM8 ('DARTEL', [11]). Each sequence was entered as one class to improve normalization by simultaneously accounting for information of both sequences and, hence, also accounting for regional T2-hypointensity. The resulting normalized images of segmented T2-hypointensity were analyzed for age-related effects by our voxel-wise regression approach.

Segmentation of T2-hypointensity

This section describes the first objective of this study, namely the segmentation of T2-hypointensities. It contains four subsections: 1) Introduction of the Bayesian mixture model, 2) estimation of model parameters by MCMC methods, 3) detection of T2-hypointensities by Posterior Predictive Checks (PPC), and 4) validation through simulated and real data.

Bayesian mixture model. Here we present a Bayesian mixture model [12,13] that is used to fit the intensities of the two T2-weighted sequences (T2-weighted and FLAIR). It is first explained in general terms and later adjusted for the data at hand.

In a mixture model, it is assumed that the observed data $\mathbf{y}$ can be divided into K unobserved classes. The components of the data vector are vectors of dimension d, i.e. $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ with $\mathbf{y}_i = (y_{i1}, \ldots, y_{id})$. In order to derive the likelihood for the data conditioned onto the respective class, a class indicator $\mathbf{x}_i = (x_{i1}, \ldots, x_{iK})$ for each observation $i = 1, \ldots, n$ is introduced, where the kth element of $\mathbf{x}_i$ is set to 1 if voxel i belongs to class k. In most applications, the indicators $\mathbf{x}' = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ are missing and it is the purpose of a mixture model to estimate these missing observations. Conditioned on the class indicator $\mathbf{x}_i$, the distribution of observation $\mathbf{y}_i$ is given by

$$p(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}) = \prod_{k=1}^{K} f(\mathbf{y}_i | \boldsymbol{\theta}_k)^{x_{ik}}.$$

Commonly, it is assumed that the mixture components f are all from the same parametric family and differ only by their parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$, which is not a real restriction but simplifies notation. For a given mixture distribution $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$ with $\lambda_k > 0, k = 1, \ldots, K$ and $\sum_{k=1}^{K} \lambda_k = 1$, the distribution of the unknown class indicators is a multinomial distribution:

$$\mathbf{x}_i | \boldsymbol{\lambda} \sim \text{Multinomial}(1; \boldsymbol{\lambda}).$$

Thus, the joint distribution of the observed data $\mathbf{y}$ and the missing label indicators $\mathbf{x}$ can be written as

$$p(\mathbf{y}, \mathbf{x} | \boldsymbol{\lambda}, \boldsymbol{\theta}) \propto \prod_{i=1}^{n} \prod_{k=1}^{K} (\lambda_k f(\mathbf{y}_i | \boldsymbol{\theta}_k))^{x_{ik}}.$$

In order to perform inference, prior distributions have to be specified for $\boldsymbol{\lambda}$ and the parameters in $\boldsymbol{\theta}$. One possible choice for $\boldsymbol{\lambda}$ is its natural conjugate, that is a Dirichlet distribution with hyperparameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ [14]. For $\boldsymbol{\alpha} = (1, \ldots, 1)'$ this prior can be seen as non-informative. The choice of priors for the parameters in $\boldsymbol{\theta}$ depends on the choice of the mixture components f.

Figure 1. Segmentation and normalization of T2-hypointensity. T2-weighted and FLAIR images are first coregistered to the T1-weighted images and then prepared for the segmentation of T2-hypointensities, which includes correction of T2-weighted and FLAIR images for magnetic field inhomogeneity by VBM8 and segmentation of T1-weighted images into the tissue classes of GM, WM, and CSF. These images are then used to segment hypointensities. The resulting T2-hypointensity images are normalized in two steps: First, T1-weighted images are affine normalized and respective parameters applied to FLAIR and T2-hypointensity images. Second, affine normalized T1-weighted and FLAIR images of all subjects are used to produce individual flow fields by DARTEL; these flow fiields are then applied to T2-hypointensity images.
doi:10.1371/journal.pone.0068196.g001

In summary, the joint posterior distribution of all unknown parameters is given by

$$p(\mathbf{x},\lambda,\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y},\mathbf{x}|\lambda,\boldsymbol{\theta})p(\lambda|\mathbf{a})p(\boldsymbol{\theta}). \qquad (1)$$

This distribution is analytically not feasible but MCMC methods can be used to simulate this posterior.

For the segmentation of our data, we adjusted the mixture model as two different T2-weighted sequences were available, which were either of high image definition but low contrast (T2-weighted), or of low image definition but high contrast (FLAIR). Aiming at the best possible segmentation of T2-hypointensity, we

utilized both pieces of information in order to reduce the effect of sequence-specific artifacts. Yet, as the tissue class of CSF is also hypointense in FLAIR sequences, we excluded voxels representing CSF according to the segmented images of the T1-weighted sequence by VBM8, i.e. the number of classes K in our study is two. Intriguingly, the distinction between GM and WM based on T1-weighted images is problematic particularly in the areas of interest in this study, namely T2-hypointense GM regions, since they have an increased iron content, which increases the T1-weighted signal and hence shifts its intensity from GM towards WM [15]. Against this backdrop, we decided to model T2-hypointensity for the two tissue classes of GM and WM separately but to generate a single image of T2-hypointensity across both GM and WM.

With the restrictions outlined above, the probabilities regarding the brain tissue classes (GM and WM) are already known from the segmentation of the T1-weighted images and can therefore used as additional prior information. Let $\hat{\boldsymbol{\lambda}} = (\hat{\boldsymbol{\lambda}}_1, \ldots, \hat{\boldsymbol{\lambda}}_n)$ with $\hat{\boldsymbol{\lambda}}_i = (\hat{\lambda}_{i,\mathrm{GM}}, \hat{\lambda}_{i,\mathrm{WM}})$ denote these probabilities, then it is assumed that the class indicators are multinomial distributed with parameter $\hat{\boldsymbol{\lambda}}$. Note that VBM8 incorporates spatial prior information of adjacent voxels into the segmentation estimation by a Markov Random Field [16]. Therefore, we did not further account for spatial correlation at this step. However, we will later consider neighbouring information during outlier detection.

Finally, because the marginal histograms of the two remaining tissue classes are considerably skewed, we use two bivariate mixture models for the mixture components and therefore introduce the subclass indicators $\boldsymbol{\zeta}'_{ik} = (\zeta_{ik1}, \zeta_{ik2})$ with

$$\zeta_{ikj} = \begin{cases} 1 & \text{if voxel } i \text{ belongs to subclass } j \text{ of class } k, \\ 0 & \text{otherwise.} \end{cases}$$

In similarity to the class indicators, the subclass indicators $\boldsymbol{\zeta}_{ik}$ follow a multinomial distribution with mixture distribution $\boldsymbol{\pi}_k = (\pi_{k1}, \pi_{k2})$, which leads to the following joint distribution for the observed intensities and the missing class and subclass indicators

$$p(\mathbf{y}, \mathbf{x}, \boldsymbol{\zeta} | \hat{\boldsymbol{\lambda}}, \boldsymbol{\pi}, \boldsymbol{\theta}) \propto \prod_{i=1}^{n} \prod_{k \in \{\mathrm{GM,WM}\}} (\hat{\lambda}_{ik} f(\mathbf{y}_i, \boldsymbol{\zeta}_{ik} | \boldsymbol{\pi}_k, \boldsymbol{\theta}_k))^{x_{ik}} \quad (2)$$

with mixture components defined as

$$f(\mathbf{y}_i, \boldsymbol{\zeta}_{ik} | \boldsymbol{\pi}_k, \boldsymbol{\theta}_k) = \prod_{j=1}^{2} (\pi_{kj} \phi(\mathbf{y}_i | \boldsymbol{\mu}_{kj}, \sum_{kj}))^{\zeta_{ikj}}.$$

Here, $\phi(\cdot | \boldsymbol{\mu}, \sum)$ is the density function of the multivariate (bivariate, in this case) normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\sum$ and all subclass indicators are collected in $\boldsymbol{\zeta}' = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n)$ with $K \times J = 2 \times 2$ elements $\boldsymbol{\zeta}_i = (\boldsymbol{\zeta}_{i1}, \boldsymbol{\zeta}_{i2})$.

Since the size of the data is quite large (over 1.1 million relevant voxels for each brain), the influence of prior distributions on the parameters of the mixture components as well as the mixture distribution will be limited and the inference will be dominated by the likelihood. We therefore choose non-informative flat prior distributions for these parameters. In detail, we use independent Dirichlet priors for the mixture distributions $\boldsymbol{\pi}_k, k \in \{1,2\}$ with

hyperparameters set to $(1,1)'$ and an independent Jeffrey's prior for $\boldsymbol{\mu}_k$ and $\sum_k$, that is $p(\boldsymbol{\mu}_k, \sum_k) \propto |\sum_k|^{-(d+1)/2}$. This does not only reflect our lack of knowledge about these parameters but also simplifies the MCMC algorithm. In summary, the joint posterior distribution of all unknown parameters is given by

$$p(\mathbf{x}, \boldsymbol{\zeta}, \boldsymbol{\pi}, \boldsymbol{\mu}, \sum | \mathbf{y}) \propto p(\mathbf{y}, \mathbf{x}, \boldsymbol{\zeta} | \hat{\boldsymbol{\lambda}}, \boldsymbol{\pi}, \boldsymbol{\mu}, \sum) p(\boldsymbol{\pi} | \boldsymbol{\alpha}) p(\boldsymbol{\mu}, \sum). \quad (3)$$

Parameter estimation. For the proposed model, all full conditional distributions can be derived in closed form. As this part is not crucial for understanding the general segmentation approach, the reader may skip to the next subsection.

The full conditional distributions for the subclass indicators of voxel i that belongs to class k can be derived from the joint distribution of $\mathbf{y}_i$ and $\boldsymbol{\zeta}_{ik}$:

$$\begin{aligned} &p(\boldsymbol{\zeta}_{ik} | \mathbf{y}_i, \boldsymbol{\pi}_k, \boldsymbol{\theta}_k, \mathbf{x}_i) \\ &\propto \prod_{j=1}^{J} (\pi_{kj} \phi(\mathbf{y}_i | \boldsymbol{\mu}_{kj}, \sum_{kj}))^{\zeta_{ikj}} \\ &\propto \prod_{j=1}^{J} \left(\frac{\pi_{kj} \phi(\mathbf{y}_i | \boldsymbol{\mu}_{kj}, \sum_{kj})}{\sum_{m=1}^{J} \pi_{kj} \phi(\mathbf{y}_i | \boldsymbol{\mu}_{kj}, \sum_{kj})} \right)^{\zeta_{ikj}} \end{aligned}$$

The last term is the core of a multinomial distribution with parameters

$$\hat{\pi}_{ikj} = \frac{\pi_{kj} \phi(\mathbf{y}_i | \boldsymbol{\mu}_{kj}, \sum_{kj})}{\sum_{m=1}^{J} \pi_{kj} \phi(\mathbf{y}_i | \boldsymbol{\mu}_{kj}, \sum_{kj})}, j = 1, \ldots, J. \quad (4)$$

Since the Dirichlet prior is the natural conjugate for the parameters of a multinomial distribution, the full conditional for the mixture distribution $\boldsymbol{\pi}_k$ is a Dirichlet distribution with updated parameters $n_{k1} + \alpha_1, \ldots, n_{kJ} + \alpha_J$, where n_{kj} is the number of observations in subclass j of class k.

For given class and subclass indicators, the parameters of the mixture components are updated for each class and subclass separately. For $\sum_{kj}$ the marginal posterior under the proposed flat prior is an inverse Wishart distribution

$$\sum_{kj} \sim \mathrm{InvWish}(n_{kj} - 1, \mathbf{S}_{kj}).$$

Here, $\mathrm{Inv-Wish}(v, \mathbf{S})$ stands for the inverse Wishart distribution with v degrees of freedom and scale matrix $\mathbf{S}$. The matrix $\mathbf{S}_{kj}$ is the sample covariance matrix of the intensities in class k and subclass j

$$\mathbf{S}_{kj} = \sum_{i=1}^{n_{kj}} (\mathbf{y}_{kji} - \bar{\mathbf{y}}_{kj})(\mathbf{y}_{kji} - \bar{\mathbf{y}}_{kj})'. \quad (5)$$

The posterior for $\boldsymbol{\mu}_k$ conditioned on $\sum_{kj}$ is then a multivariate normal distribution with mean equal to the mean of the intensities

in class k and subclass j, $\bar{\mathbf{y}}_{kj}$, and covariance matrix $\sum_{kj}/n_{kj}$.

In summary, the following Gibbs sampler can be used to simulate distribution (3):

1. Initialize the class and subclass indicators $\mathbf{x}^{(0)}$ and $\boldsymbol{\zeta}^{(0)}$.
2. For $t=1,\ldots,n_{\text{sim}}$ repeat the following steps:

   - For current $\mathbf{x}^{(t-1)}$ and $\boldsymbol{\zeta}^{(t-1)}$ calculate the sample covariance matrix $\mathbf{S}_{kj}^{(t)}$ for $k\in\{\text{GM},\text{WM}\}$ and $j\in\{1,2\}$ according to (5) and draw $\sum_{kj}^{(t)}$ from an inverse Wishart distribution with scale matrix $\mathbf{S}_{kj}^{(t)}$ and $n_{kj}^{(t)}-1$ degrees of freedom.
   - For $k\in\{\text{GM},\text{WM}\}$ and $j\in\{1,2\}$ draw $\boldsymbol{\mu}_{kj}^{(t)}$ from a normal distribution with mean $\bar{\mathbf{y}}_{kj}^{(t)}$ and covariance matrix $\sum_{kj}^{(t)}/n_{kj}^{(t)}$.
   - For $k\in\{\text{GM},\text{WM}\}$ draw $\boldsymbol{\pi}_k^{(t)}$ from a Dirichlet distribution with parameters $(n_{k1}^{(t)}+\alpha_1, n_{k2}^{(t)}+\alpha_2)$.
   - For $i=1,\ldots,n$ and $k\in\{\text{GM},\text{WM}\}$ draw $\zeta_{ik}^{(t)}$ from a multinomial distribution with parameters $\hat{\boldsymbol{\pi}}_{ik}^{(t)}=(\hat{\pi}_{ik1}^{(t)},\hat{\pi}_{ik2}^{(t)})$ according to (4).
   - For $i=1,\ldots,n$ draw $\mathbf{x}_i$ from a multinomial distribution with parameters $\hat{\boldsymbol{\lambda}}_i=(\hat{\lambda}_{i,\text{GM}},\hat{\lambda}_{i,\text{WM}})$.

3. After discarding the realizations of an initial burn-in phase, the remaining samples can be considered as dependent samples of the joint posterior (3).

For each subject, three parallel chains of length 1500 were calculated by the Gibbs Sampler described above. Class and subclass indicators were initialized at random. Figure 2 shows trace plots of such a chain for one randomly chosen subject. The left panel displays the components of $\boldsymbol{\mu}_{\text{GM},1}$ and the right panel those of $\sum_{\text{GM},1}$. As it can be seen, mixing of chains is quite good and label switching [13] does not occur. For the calculation of T2-hypointensities, we discarded the initial 500 draws and kept every second sample. For the remaining draws we calculated Gelman and Rubin's potential scale reduction factor [17] for the mean of the parameters of the mixture components and the mixture
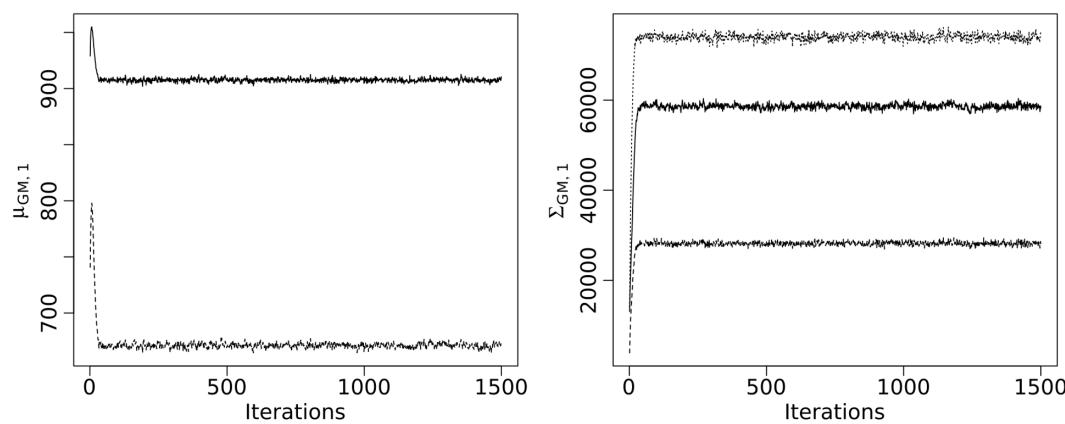
distributions. In all cases, the value was nearly indistinguishable from 1 indicating that the simulation converged to the target distribution [17].

Posterior predictive checks. Fitting the model will yield n_{sim} realizations of the posterior distribution. Denote these samples by $\boldsymbol{\theta}^{(t)}, t=1,\ldots,n_{\text{sim}}$. Those samples can be used to perform PPC in order to check the fit of the model [18,19] or to identify outliers, as explained next. The basis for PPC are replicated samples ('fake' data) $\mathbf{y}^{\text{rep}}$ of the observed data $\mathbf{y}$ according to the posterior predictive distribution (PPD):

$$p(\mathbf{y}^{\text{rep}}|\mathbf{y}) = \int p(\mathbf{y}^{\text{rep}}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}. \tag{6}$$

To generate samples $\mathbf{y}^{\text{rep},(t)}, t=1,\ldots,n_{\text{sim}}$ out of this distribution, one proceeds as follows: For each realization of the unknown parameters generate n samples according to the likelihood. In the case of the mixture model explained above, we generate $n_{kj}^{(t)}$ samples (according to the actual label configuration $\mathbf{x}^{(t)},\boldsymbol{\zeta}^{(t)}$) of a bivariate normal distribution with parameters $\boldsymbol{\mu}_{kj}^{(t)}$ and $\sum_{kj}^{(t)}$. Figure 3 illustrates this procedure for a randomly chosen subject. The first row shows a slice of the observed T2-intensities followed by three simulated slices. The last panel in the first row displays the mean and standard deviation of the simulated intensities. The second row displays the same information for the same slices of the FLAIR-image. In both cases it can be seen that hypointense structures visible in the observed images are not present in the calculated mean images. This illustrates that we can detect T2-hypointensities by comparing the replicated images to the observed image.

In general, once the replicated data sets are available, they can serve to measure the discrepancy between the model and observed data by analyzing test quantities, or general discrepancy measures $T(\mathbf{y},\boldsymbol{\theta})$. This discrepancy measure is calculated for the observed and replicated data. It can be any kind of scalar summary of the data. The calculated discrepancies may be displayed graphically to perform visual model checks or by using Bayesian posterior predictive p-values [6]. For segmentation, we record if the replicated intensity of voxel i is greater or equal than its observed intensity, hence, we choose $T(\mathbf{y}_i,\boldsymbol{\theta})$ to be the intensity value itself.

Figure 2. Trace plots for the Gibbs sampler of the mixture model for T2-hypointensity segmentation of one randomly chosen subject. Components of $\boldsymbol{\mu}_{\text{GM},1}$ (left) and of $\sum_{\text{GM},1}$ (right). See text for details.
doi:10.1371/journal.pone.0068196.g002
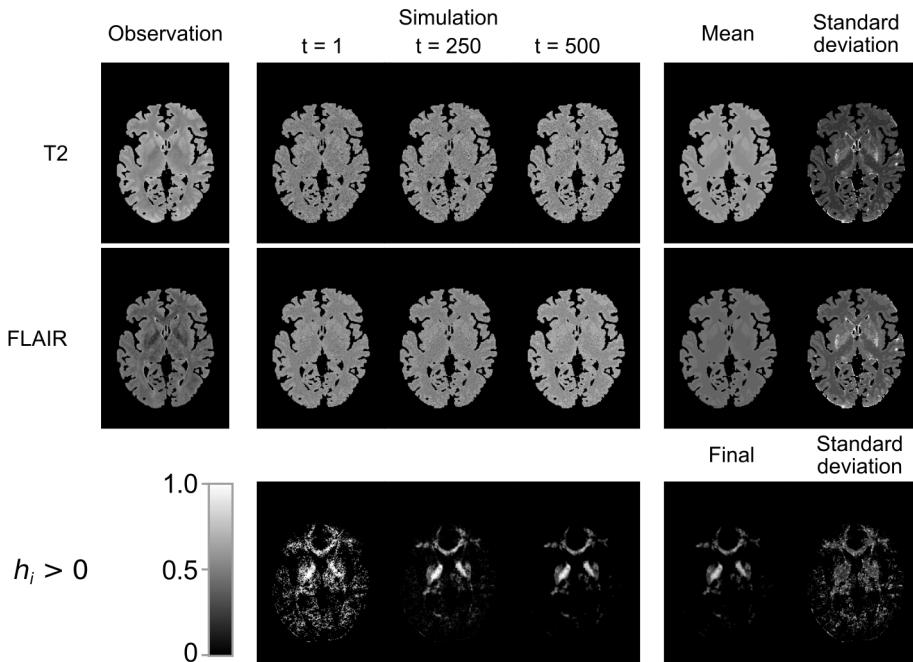
Figure 3. Simulation and outlier detection of T2-hypointensity. Images were derived from a randomly chosen subject. On the left, a normalized T2-weighted (top) and a normalized FLAIR (bottom) image is shown (only gray and white matter). Three examples of respective simulated images and their means and standard deviations are shown in the middle and on the right, respectively. In the lower row, respective positive values of the hypointensity score, derived from both T2-weighted and FLAIR images, as well as their final image and standard deviation are shown and gray-scaled according to the bar in the lower left. See text for details.
doi:10.1371/journal.pone.0068196.g003

We denote this by

$$T(\mathbf{y}_i, \boldsymbol{\theta}^{(t)}) = I(y_{i,FLAIR} < y_{i,FLAIR}^{\mathrm{rep},(t)} \wedge y_{i,T2} < y_{i,T2}^{\mathrm{rep},(t)}).$$

Based on this information, we iteratively build the hypointensity score h_i for each voxel i by applying the following formula

$$h_i^{(t)} = \begin{cases} \dfrac{(t-1)h_i^{(t-1)}+1}{t} & , \text{if } T(\mathbf{y}_i, \boldsymbol{\theta}^{(t)}) = 1 \\[2mm] \dfrac{(t-1)h_i^{(t-1)}-1}{t} & , \text{otherwise.} \end{cases} \quad (7)$$

In this form, the final hypointensity score is simply the mean of minus and plus ones, where plus one results if the replicated intensity is greater than the observed one. Thus the intensity score takes values between -1 and 1 with positive voxels indicating more hypointensity and negative values less hypointensity. To account for spatial dependencies between adjacent voxels, we expand equation (7) to

$$h_i^{(t)} = \begin{cases} \dfrac{(t-1)h_i^{(t-1)}+(h_{N_i,\mathrm{pos}}^{(t)})^{1/\delta}}{t} & , \text{if } T(\mathbf{y}_i, \boldsymbol{\theta}^{(t)}) = 1 \\[2mm] \dfrac{(t-1)h_i^{(t-1)}-|h_{N_i,\mathrm{neg}}^{(t)}|^{1/\delta}}{t} & , \text{otherwise.} \end{cases}$$

Here, $h_{N_i,\mathrm{pos}}$ is the mean of all neighboring intensity scores that are positive. Likewise, $h_{N_i,\mathrm{neg}}$ is the mean of all neighboring intensity scores that are negative. The parameter δ controls the influence of neighboring intensity scores on $h_i^{(t)}$.

The last row in Figure 3 displays positive values of $h_i^{(t)}$ for three different iterations and the final segmentation along with its standard deviation. According to the validation, the parameter δ was set to 1.4. See next paragraph for details.

Validation. First, we validated our segmentation procedure by a simulation study. Accounting for the lack of a commonly accepted gold standard, we manually labeled hypointense regions that are visible in the mean FLAIR images of our healthy controls. Before averaging, images were normalized by the use of the deformation field derived from standard normalization of T1-weighted images as implemented in VBM8. We then added this binary label as an extra class to BrainWeb's (http://brainweb.bic.mni.mcgill.ca/brainweb/) discrete anatomical model [20] and simulated T1-weighted, T2-weighted and FLAIR images by BrainWeb's MRI simulator [21]. Selected slices of the simulated T2-weighted and FLAIR images without and with T2-hypointensities as well as of the binary label are shown in Figure 4. We applied our algorithm to the simulated images with values of δ ranging from 1 to 4 with an increment of 0.05 and determined the optimal value by calculating the Dice coefficient (DC, [22,23]). We also considered the influence of different values for the binary threshold ranging from 0 to 1 with an increment of 0.05 for each value of δ. Beyond the simulation study, we biologically validated the algorithm by visually comparing the segmented T2-hypointensity images with both the T2-weighted images and the FLAIR images. To evaluate the effect of the incorporation of both T2-weighted sequences (T2-weighted and FLAIR) into the segmentation by the mixture model, we repeated segmentation of T2-hypointensity with an adapted version of the model twice after having subjected either only the T2-weighted or only the FLAIR images.
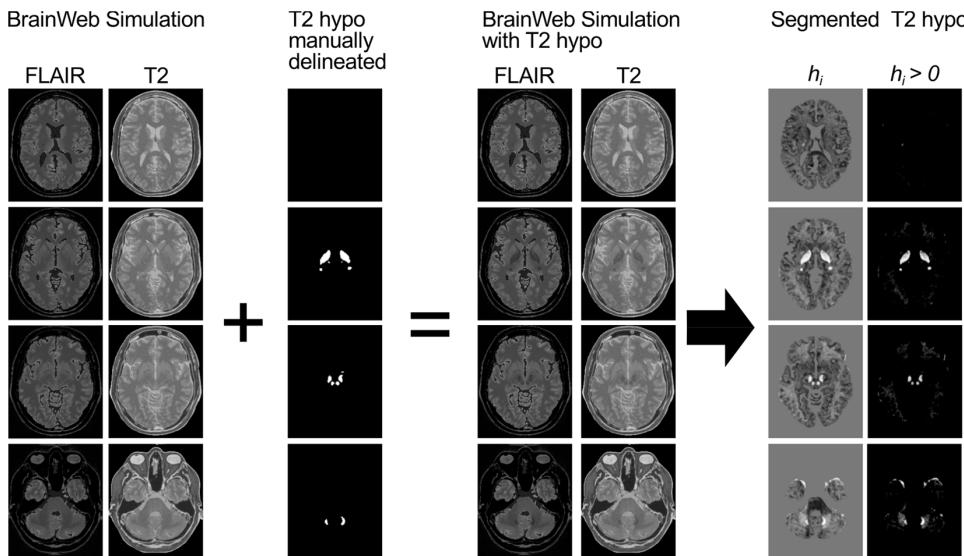
Figure 4. Segmentation of simulated T2-hypointensities. Manually delineated T2-hypointensities were added as an extra class to BrainWebs discrete anatomical model. This way, T1-weighted, T2-weighted and FLAIR images were simulated. Hypointensities were then segmented from the simulated images.
doi:10.1371/journal.pone.0068196.g004

Bayesian voxel-wise regression with smoothness priors

This section describes the second objective of this study, namely the adaption of a voxel-wise linear regression model. It contains four subsections: 1) introduction of the model, 2) parameter estimation, 3) calculation of posterior probability maps, and 4) validation.

Description of the model. It is still a challenging question in neuroimaging, how to handle spatial correlations in the data, alongside the ideas that we expect effects of interest to occur in clusters of voxels and that models accounting for these dependencies are likely to be more robust. Existing frequentist methods use a combination of pre-smoothing and spatial statistics based on Random Field Theory. However, these frequentist approaches rely on the subjective selection of a number of parameters, such as the amount of spatial smoothing to impose, and the choice of cluster forming thresholds. Of note, Bayesian inference offers the possibility of a solution to these problems via the integration of the dependency among adjacent voxels into the regression model itself [5]. Those approaches were formulated previously in the context of first level fMRI analyses [7,8] but not for second level analyses as necessary for structural MRI data so that we had to adopt previously proposed approaches.

Let $\mathbf{y}_i$ denote the $m \times 1$ vector of responses of voxel $i, i = 1, \ldots, n$, for the m subjects, the regression model for the ith voxel can be written as

$$\mathbf{y}_i \sim \mathrm{N}(\boldsymbol{\eta}_i, \kappa_i^{-1}\mathbf{I}_m)$$

where $\boldsymbol{\eta}_i$ denotes the linear predictor, $\mathbf{I}_m$ the $m \times m$ identity matrix and κ_i the unknown precision parameter, i.e. the inverse variance. The linear predictor has the form

$$\boldsymbol{\eta}_i = \mathbf{x}'_1 \beta_{i1} + \ldots + \mathbf{x}'_p \beta_{ip}.$$

In this notation, the $m \times 1$ vector $\mathbf{x}_k, k = 1, \ldots, p$ collects the values of the kth covariate for all subjects and β_{ik} represents the
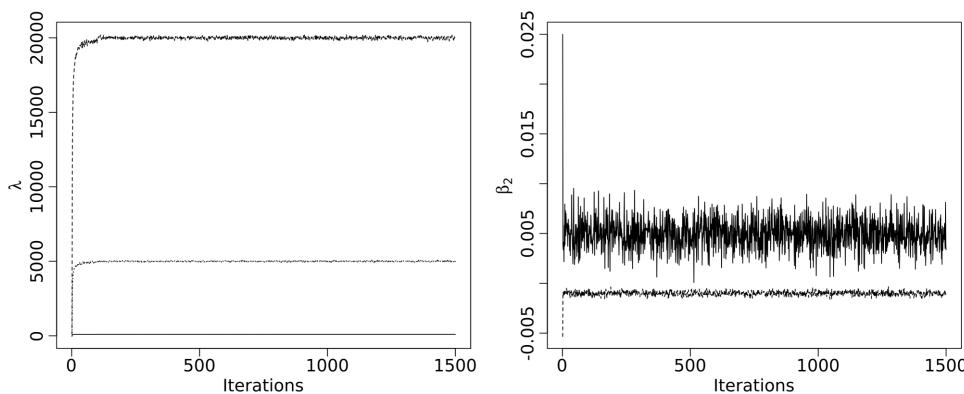
Figure 5. Trace plots of the voxel-wise regression model. Precision parameters (left) and main effect of age for two selected voxels (right). Corresponding MNI coordinates are $-16, -63, -36$ (top) and $-18, 38.5, 7$ (bottom).
doi:10.1371/journal.pone.0068196.g005
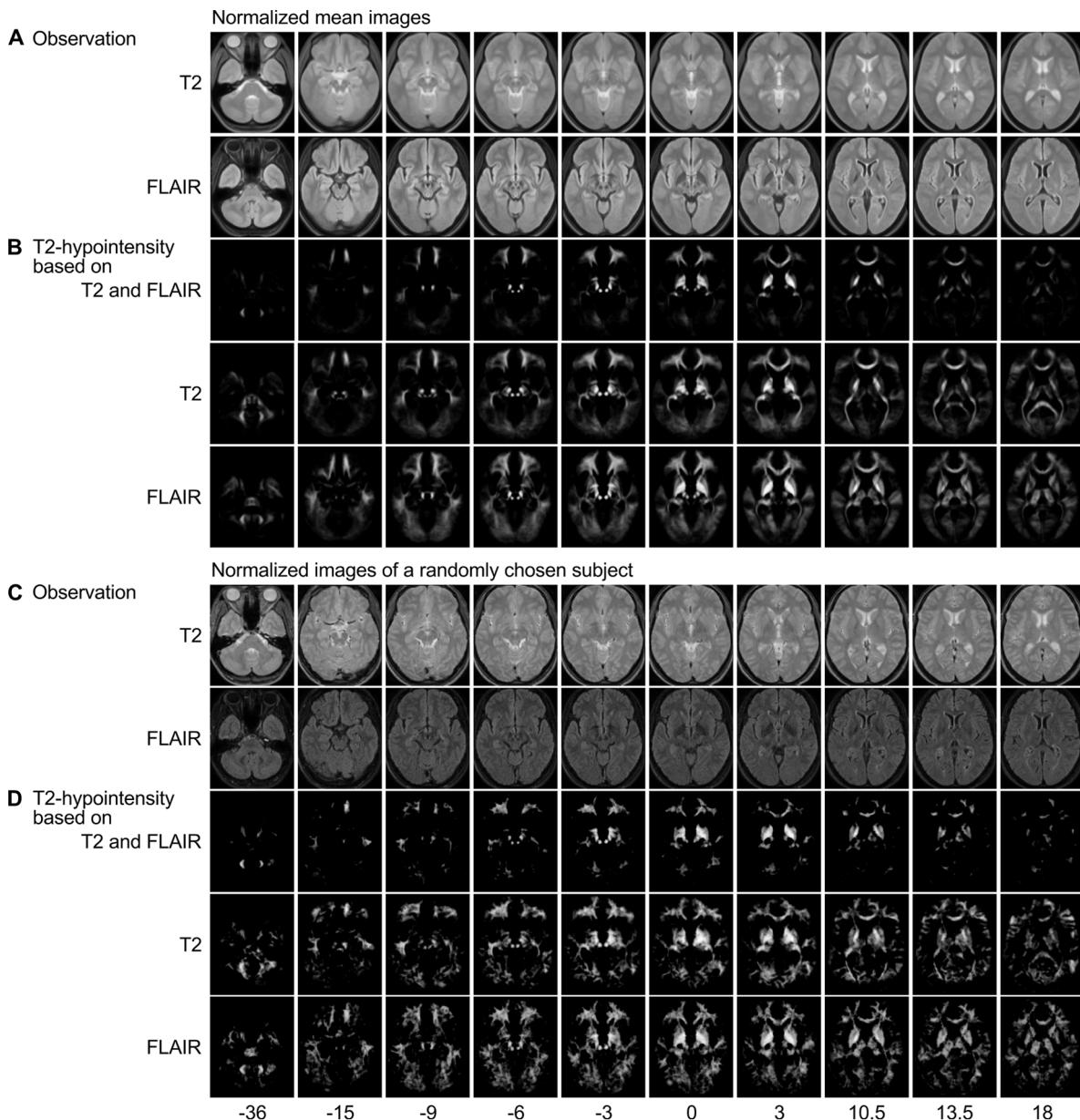
Figure 6. Segmented T2-hypointensity. A) Axial slices of normalized mean images (T2-weighted and FLAIR) are shown. B) Corresponding axial slices of segmented T2-hypointensity are shown (upper row, based on both T2-weighted and FLAIR images; middle row, based only on T2-weighted images; lower row, based only on FLAIR images). C and D) Information of a randomly chosen subject is given in analogy to panels A and B; for better illustration, normalized images are shown, although the algorithm operates in the original (native) space. See text for details.
doi:10.1371/journal.pone.0068196.g006

corresponding unknown regression coefficient. In most applications, the first covariate is $\mathbf{x}_1 = 1_m$ and thus β_{i1} is the intercept in the model for the ith voxel. In the study presented here the linear predictor consists of an intercept and main effects of age and sex.

By defining $\mathbf{y} = (\mathbf{y}'_1, \ldots, \mathbf{y}'_n)'$ and $\boldsymbol{\beta}_k = (\beta_{1k}, \ldots, \beta_{nk})'$ for $k = 1, \ldots, p$, the $mn \times n$ matrix

$$\mathbf{X}_k = \mathbf{I}_n \otimes \mathbf{x}_k$$

and the $mn \times n$ diagonal precision matrix

$$\mathbf{Q_y} = \mathrm{diag}(\kappa_1, \ldots, \kappa_n) \otimes \mathbf{I}_m$$

the complete model can be written in compact matrix notation as

$$\mathbf{y} \sim \mathrm{N}(\boldsymbol{\eta}, \mathbf{Q_y}^{-1})$$
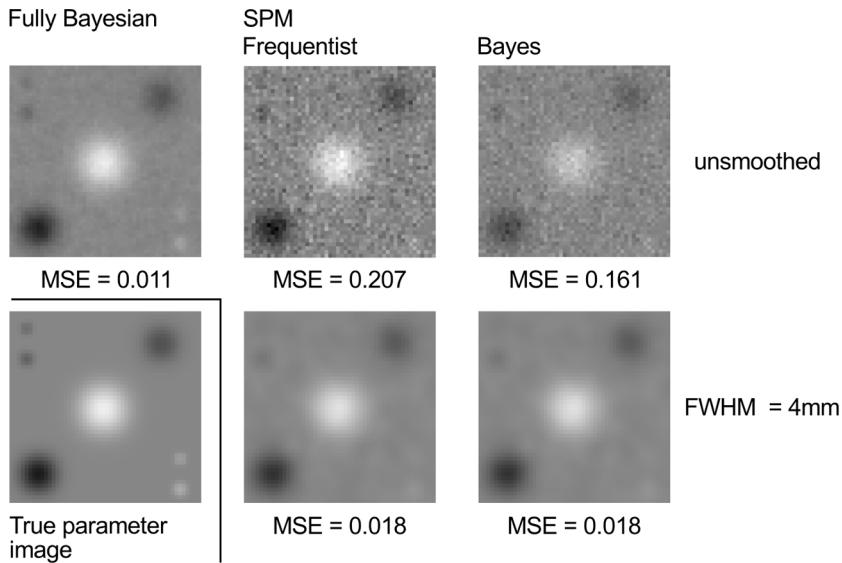
Figure 7. Estimated regression coefficients of the simulated data. Posterior mean image for unsmoothed data of the approach proposed in this paper is shown in the upper left corner. Results of SPM's frequentist and Bayesian implementation are shown in the second and third column for unsmoothed (upper row) and smoothed (lower row) data, respectively. The true parameter image is shown in the lower left corner. The approach proposed in this paper performs best as demonstrated by the MSE and by visual inspection.
doi:10.1371/journal.pone.0068196.g007

with

$$\boldsymbol{\eta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \ldots + \mathbf{X}_p\boldsymbol{\beta}_p. \tag{8}$$

We use independent Gaussian Markov Random Field (GMRF, [24]) priors for the regression coefficients as they are commonly used in neuroimaging in order to account for the spatial structure of images, see for example [25], [10] and [26]. We therefore have

$$p(\boldsymbol{\beta}_k|\lambda_k) \propto \exp\{-\frac{\lambda_k}{2}\boldsymbol{\beta}'_k\mathbf{K}\boldsymbol{\beta}_k\}.$$

Here, λ_k is a precision parameter (inverse variance) and $\mathbf{K}$ is a structure matrix. Whereas λ_k operates as a smoothness parameter that is estimated by the data, the matrix $\mathbf{K}$ accounts for spatial dependencies between the regression coefficients. The elements of $\mathbf{K}$ are

$$K_{il} = \begin{cases} -1 & \text{if } i \sim l, \\ n_i & \text{if } i = l, \\ 0 & \text{otherwise,} \end{cases}$$

where the number of voxels in the neighborhood of voxel i is denoted by n_i and $i \sim l$ stands for all voxels l that share a common border with voxel i, that is we use a first order neighborhood consisting of the six nearest neighbors. One advantage of such a prior is that it acts like a smoothness prior. To show this, the full conditional of β_{ik}, given all the other values of $\boldsymbol{\beta}_k$, can be written as

$$\beta_{ik}|\boldsymbol{\beta}_{-ik} \sim \mathrm{N}\left(\frac{1}{n_i}\sum_{l \sim i}\beta_{lk}, \frac{1}{n_i\lambda_k}\right).$$

Thus, the conditional prior corresponds to a normal distribution with expectation equal to the mean of the effects of neighboring voxels and precision proportional to the number of neighboring voxels and precision λ_k.

To perform fully Bayesian inference, priors for the precision parameters $\kappa_i, i = 1, \ldots, n$ and $\lambda_k, k = 1, \ldots, p$ have to be chosen. We chose independent Gamma distributions with hyperparameters $a_{\mathbf{y}}$ and $b_{\mathbf{y}}$ for the precisions of $\mathbf{y}$ and a_λ and b_λ for the precisions of the regression coefficients. By adopting small values for the hyperparameters, for example 0.1, 0.01 or 0.001, one obtains 'diffuse' priors for the precision parameters.

In our study, response values of the voxel-wise regression model are the segmented hypointensities. Besides an intercept and the effect of age, sex is included as a dummy-coded factor (0 = male, 1 = female) yielding the model

$$\boldsymbol{\eta}_i = \beta_{i1} + \mathbf{age}\cdot\beta_{i2} + \mathbf{sex}\cdot\beta_{i3} \tag{9}$$

for voxel i. Here, age and sex are the vectors of age and sex, respectively.

Parameter estimation. To obtain samples from the joint posterior

$$p(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p, \boldsymbol{\kappa}, \boldsymbol{\lambda}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p, \boldsymbol{\kappa}) \prod_{i=1}^{n} p(\kappa_i) \prod_{k=1}^{p} p(\boldsymbol{\beta}_k|\lambda_k)p(\lambda_k) \tag{10}$$

a Gibbs sampler can be used. As this part is not crucial for understanding the adoption of our voxel-wise linear regression model, the reader may skip to the next subsection.
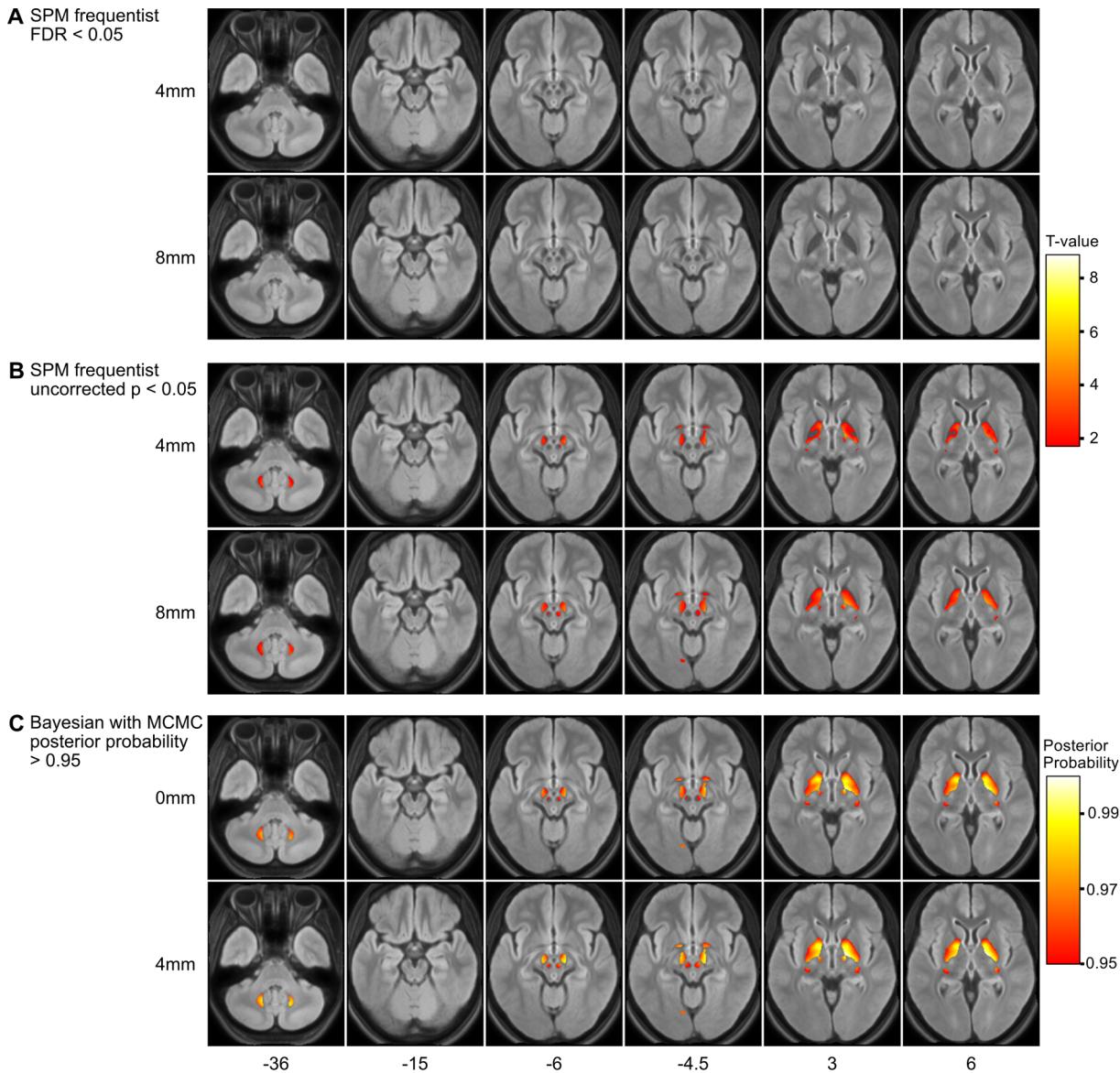
Figure 8. Effect of age on T2-hypointensity. Increasing T2-hypointensity with increasing age is projected onto the mean normalized FLAIR image. Axial slices are indicated in the upper row. Significance is color-coded according to the T-value (Panels A and B) and posterior probability (Panel C) as indicated by the bars on the right. A–B) Results derived from the frequentist approach as implemented in SPM8 are shown after application of different statistical thresholds (Panel A, false-discovery rate <0.05; Panel B, uncorrected p-value <0.05) and different smoothing kernels (upper rows, 4 mm; lower rows, 8 mm. C) Fully Bayesian inference could not only identify the globus pallidus, substantia nigra, and red nucleus but also the dentate nucleus. This result was largely independent of smoothing although more voxels were identified after smoothing with 4 mm.

doi:10.1371/journal.pone.0068196.g008

To obtain the full conditionals for $\boldsymbol{\beta}_k$, let $\tilde{\mathbf{y}} = \mathbf{y} - \boldsymbol{\eta}_{-k}$, where $\boldsymbol{\eta}_{-k}$ is the linear predictor (8) without the kth term. Then, the full conditional for $\boldsymbol{\beta}_k$ is given by

$$p(\boldsymbol{\beta}_k|\cdot) \propto p(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\kappa})p(\boldsymbol{\beta}_k|\lambda_k)$$

$$\propto \exp\{-\frac{1}{2}(\tilde{\mathbf{y}} - \mathbf{X}_k\boldsymbol{\beta}_k)'\mathbf{Q}_\mathbf{y}(\tilde{\mathbf{y}} - \mathbf{X}_k\boldsymbol{\beta}_k) - \frac{\lambda_k}{2}\boldsymbol{\beta}'_k\mathbf{K}\boldsymbol{\beta}_k\}$$

$$\propto \exp\{-\frac{1}{2}(\boldsymbol{\beta}'_k(\mathbf{X}'_k\mathbf{Q}_\mathbf{y}\mathbf{X}_k + \lambda_k\mathbf{K})\boldsymbol{\beta}_k - 2\boldsymbol{\beta}'_k\mathbf{X}'_k\mathbf{Q}_\mathbf{y}\tilde{\mathbf{y}})\}.$$

By completing squares, we obtain

$$\boldsymbol{\beta}_k \sim \mathrm{N}(\boldsymbol{\mu}_k, \mathbf{Q}_k^{-1}) \qquad (11)$$

with

$$\mathbf{Q}_k = \mathbf{X}'_k\mathbf{Q}_\mathbf{y}\mathbf{X}_k + \lambda_k\mathbf{K} \quad \text{and} \quad \boldsymbol{\mu}_k = \mathbf{Q}_k^{-1}\mathbf{X}'_k\mathbf{Q}_\mathbf{y}\tilde{\mathbf{y}}.\,s \qquad (12)$$

The full conditional for smoothness parameter λ_k are obtained by

$$p(\lambda_k|\boldsymbol{\beta}_k) \propto p(\boldsymbol{\beta}_k|\lambda_k)p(\lambda_k)$$

$$\propto \lambda_k^{\mathrm{rk}(\mathbf{K})/2} \exp\{-\frac{\lambda_k}{2}\boldsymbol{\beta}'_k\mathbf{K}\boldsymbol{\beta}_k\}\lambda_k^{a_\lambda-1}\exp\{-b_\lambda\lambda_k\}$$

$$= \lambda_k^{a_\lambda + \mathrm{rk}(\mathbf{K})/2 - 1}\exp\{-(b_\lambda + \frac{1}{2}\boldsymbol{\beta}'_k\mathbf{K}\boldsymbol{\beta}_k)\lambda_k\}.$$

Thus, the full conditional for λ_k is a Gamma distribution with updated parameters

$$\tilde{a}_\lambda = a_\lambda + \frac{\mathrm{rk}(\mathbf{K})}{2} \quad \text{and} \quad \tilde{b}_\lambda = b_\lambda + \frac{1}{2}\boldsymbol{\beta}'_k\mathbf{K}\boldsymbol{\beta}_k. \tag{13}$$

With a similar calculation, it can be shown that the full conditional for the precision parameter of voxel i, κ_i, follows a Gamma distribution with updated parameters

$$\tilde{a}_{\mathbf{y}} = a_{\mathbf{y}} + \frac{m}{2} \quad \text{and} \quad \tilde{b}_{\mathbf{y}} = b_{\mathbf{y}} + \frac{1}{2}(\mathbf{y}_i - \boldsymbol{\eta}_i)'(\mathbf{y}_i - \boldsymbol{\eta}_i) \tag{14}$$

thus, the precision parameters $\kappa_1, \ldots, \kappa_n$ can be updated for each voxel independently.

1. Initialize the precision parameters $\kappa_1^{(0)}, \ldots, \kappa_n^{(0)}$ and $\lambda_1^{(0)}, \ldots, \lambda_p^{(0)}$ as well as the regressions coefficients $\boldsymbol{\beta}_1^{(0)}, \ldots, \boldsymbol{\beta}_p^{(0)}$.

2. For $t = 1, \ldots, n_{\mathrm{sim}}$ repeat the following steps:

   - For current $\boldsymbol{\kappa}^{(t-1)}$ calculate $\mathbf{Q}_k^{(t)}$ and $\boldsymbol{\mu}_k^{(t)}$ for $k = 1, \ldots, p$ according to (12) and draw $\boldsymbol{\beta}_k^{(t)}$ from a multivariate normal distribution with mean $\boldsymbol{\mu}_k^{(t)}$ and precision matrix $\mathbf{Q}_k^{(t)-1}$.
   - For $i = 1, \ldots, n$ draw $\kappa_i^{(t)}$ from a Gamma distribution with shape and rate parameters according to (14).
   - For $k = 1, \ldots, p$ draw $\lambda_k^{(t)}$ from a Gamma distribution with shape and rate parameters according to (13).

3. After discarding the realizations of an initial burn-in phase the remaining samples can be considered as dependent realizations of the joint posterior (10).

By sampling $\boldsymbol{\beta}_k$ from its full conditional, we make use of the independence structure that is imposed by the voxel layout. To be more precise, we split all voxels in two sets of independent voxels according to the first order neighborhood. This has the advantage that, conditioned on each other, the precision matrix of the full conditional for each set is diagonal. Thus, calculating the corresponding Cholesky triangle is not necessary anymore and sampling of $\boldsymbol{\beta}_k$ becomes feasible while still considering the full covariance structure [24].

Hyperparameters for the precision and smoothness parameters are set in accordance with [7] to $a_{\mathbf{y}} = 0.001$ and $b_{\mathbf{y}} = 0.001$ and to $a_\lambda = 1$ and $b_\lambda = 10$.

To fit model (9) to the data, three parallel chains of length 1500 were calculated. Starting points were generated randomly in the interval $[0,1]$ for precision parameters and in the interval $[-1,1]$ for regression coefficients. Trace plots of the precision parameters λ_1, λ_2 and λ_3 of one chain are shown in the left panel of Figure 5 and of two selected voxels of $\boldsymbol{\beta}_2$ in the right panel of this figure. MNI coordinates of voxels are $-16, -63, -36$ (top) and $-18, 38.5, 7$ (bottom). As for the mixture model, the initial 500 draws were discarded and additional 500 draws of each chain were

saved. Again, we calculated Gelman and Rubin's potential scale reduction factor for the mean of the precision parameters. In all cases, it can be assumed that the simulation converged to the target distribution.

Calculation of posterior probability maps. Results for the regression coefficients can be displayed in different ways. In order to compare the results with those derived from an already existing implementation, we calculated posterior probability maps. Usually, this is achieved by computing p-values based on the analytical marginal posterior distributions [27]. Here, we estimate the probability p-values of a positive or negative effect of the predictors by the proportion of the corresponding MCMC draws that lie above or below zero, respectively.

Validation. First, we validated our voxel-wise regression model by a simulation study. In accordance with Penny et al. [10], we generated a two-dimensional 50×50 pixel image of regression coefficients that contains Gaussian blobs. It was created by placing circular effect patterns with heights ranging from -5 to $+5$ on seven different locations. Radii of effects ranged from 1 to 4 pixels. Gaussian blobs were obtained by smoothing these effects individually with Gaussian kernels having different full width at half maximum (FWHM) ranging from 1 to 8 pixels. In order to simulate observation images, we generated values for one metric covariate at random between 0 and 1 and multiplied the coefficient image with those values. Finally, we randomly generated a precision parameter for each pixel using a gamma distribution with shape and scale parameters set to 2. Gaussian noise with precision set to these parameters was added to the multiplied images of regression coefficients. This way, we generated 30 'fake' observations. The parameter image was estimated by the presented approach and by SPM8 (both standard frequentist and Bayesian implementation) after smoothing the observation images with Gaussian kernels of 0 and 4 pixels. Results were compared by visual inspection and by calculating the mean squared error (MSE) between the true and the estimated coefficient images.

Second, we biologically validated our model by analyzing our normalized segmented T2-hypointensity images for age-related effects. This validation is justified as it is commonly accepted that the loss in the T2-weighted signal within the most T2-hypointense GM areas is due to an increased iron content, which is not only related to neurodegeneration but also to normal aging [15,28,29]. We compared the results derived from our approach to those derived from SPM8. Yet, we will not report the results of the Bayesian approach implemented in SPM8, which yielded implausible results. This could be replicated with simulated data by drastically increasing the ratio between voxels without an effect and those with an effect. We reported this problem, which is intended to be fixed. To compare our approach to the frequentist approach in SPM8 we applied different smoothing kernels, namely a Gaussian kernel of 0, 4, and 8 mm FWHM. As significance thresholds, we chose a posterior probability of 0.95 or, correspondingly, a false discovery rate (FDR) of 0.05 [27]. In all cases, the effect size threshold was set to zero. We restricted our analyses to voxels with a mean hypointensity score of greater than 0.25.

Software

We implemented both presented approaches in pure MATLAB (http://www.mathworks.de/products/matlab/) code. Segmentation of one subject took about 20 minutes with a 3.2 GHz processor. On the same machine, one chain for the voxel-wise regression model could be obtained within six hours requiring about 20 GB RAM.

Results

Segmentation of T2-hypointensities

Visual inspection of the segmented T2-hypointensities of simulated data showed that T2-hypointense regions were reliably detected. Spurred false positives occurred at the border to CSF. The highest DC value (0.754) was observed for a δ of 1.4 and for the binary threshold of 0.45. This excellent similarity measure [30,31] was robust as indicated by DC values greater than 0.7 after changing $\delta \in [1.2, 1.65]$ and the binary threshold in [0.1, 0.65]. Selected slices of the hypointensity score for the simulated images and $\delta = 1.4$ are shown in Figure 4. For further analyses, we chose a δ value of 1.4.

Segmentation of T2-hypointensities in healthy controls is displayed in Figure 6. The first five rows show the mean images derived from the whole group, the last five rows show the images derived from a randomly chosen subject. Structures known to be T2-hypointense are clearly visible, that is the globus pallidus, substantia nigra, red nucleus, and dentate nucleus. Further, the use of both T2-weighted sequences (T2-weighted and FLAIR) resulted in more accurate segmentation than the use of only one sequence.

Voxelwise regression model

Estimated regression coefficients of our simulated data are shown in Figure 7. The approach proposed in this paper performs best as demonstrated by the MSE and by visual inspection. While all blobs of the true coefficient image are visible in our estimation, both SPM's frequentist and Bayesian implementation fail to detect smaller effects.

Within the GM of our healthy controls, we observed only T2-hypointensity that increased with increasing age. The results derived from different multiple linear regression models yielded different results, which will be described in correspondence to the number of identified voxels from low to high. The conventional frequentist approach did not yield any meaningful results neither at the pre-defined significant threshold nor at the voxel threshold of 0.05 family-wise error corrected (Figure 8, Panel A). Only after relaxing the voxel threshold to 0.05 uncorrected, we observed all expected GM regions, namely globus pallidus, substantia nigra, red nucleus, and dentate nucleus (Figure 5, Panel B). By the use of our fully Bayesian approach, we could not only identify the globus pallidus, substantia nigra, and red nucleus but also the dentate nucleus. This result was largely independent of smoothing although more voxels were identified after smoothing with 4 mm (Figure 5, Panel D).

Discussion

In this work, we have developed and validated algorithms based on fully Bayesian inference to preprocess and statistically analyze structural MRI data. Separately for preprocessing and statistical analysis, we will discuss the rationale, realization and validation of our approaches. We will also acknowledge limitations of our work and outline room for improvement.

In the first part of our study, we developed a tool for segmentation of T2-hypointensity, which, to the best of our knowledge, is the first that utilizes PPC for outlier detection in the context of neuroimaging. The concept of PPC derives its flexibility from the possibility that any scalar summary of the data can be chosen for the discrepancy measure T and that it can be applied to every model that has been fitted in a fully Bayesian way. In principle, simulation based model checking techniques can also be applied in the framework of non-Bayesian estimation methods [32] given that (asymptotic) distributions of model parameters can

be obtained, for example, by standard errors of parameters. However, commonly used iterative algorithms, such as the expectation maximization algorithm, need to be extended to estimate standard errors, which has been regarded technically challenging [33]. Therefore, we decided to address the segmentation problem by PPC based on a fully-Bayesian approach. The resulting segmentation algorithm was fully automatic and operated across the whole brain. Influence of adjacent voxels during segmentation can be controlled by the δ-parameter. Further, no thresholds had to be chosen and user-defined regions of interest did not have to be defined. Moreover, the flexibility of the proposed mixture model enabled the incorporation of two different T2-weighted sequences, which clearly improved the precision of T2-hypointensity segmentation. As a result, all GM regions known to be T2-hypointense in healthy subjects were segmented reliably and almost exclusively in both simulated data and real data even at the single subject level. As T2-hypointense GM areas display an increased T1-weighted signal similar to that of WM, we were unable to clearly attribute T2-hypointensity to one of the two tissue classes through our model. Therefore, we included all brain parenchyma, namely GM and WM, in our segmentation. Accordingly, our tool also segmented WM areas. These areas, primarily the corpus callosum and frontal forceps, are known to contain tightly packed fibers so that segmentation of these WM areas can be attributed to the lowest T2-weighted WM signal of these regions [34]. Hence, segmentation of WM is inherent to our approach and biologically plausible. Moreover, the proposed mixture model may be extended in many further ways. For instance, different distributions can be chosen for the mixture components to obtain a better fit to particular structures of the intensity distribution. Further, prior information for the location of T2-hypointensity may be constructed and used within the segmentation step. Although our tool accounts for the information of three different MRI sequences, the approach is still hierarchical as information of the T1-weighted image constitutes the basis for the segmentation of the two T2-weighted images. However, a truly multimodal segmentation that simultaneously accounts for all available information is likely to be advantageous over our algorithm.

In the second part of our study, we adapted a voxel-wise linear regression model through Bayesian inference. In contrast to SPM8, which applies global shrinkage priors [27], our approach accounts for the spatial dependency of voxels within the estimation procedure by the use of GMRF priors. Further, smoothness parameters are estimated from the data at hand by MCMC methods. We expected our approach to be advantageous over conventional frequentist and available Bayesian approaches for three reasons. First, eliminating the necessity for *post hoc* correction for multiple comparisons should increase statistical power compared to conventional frequentist approaches. Second, accounting for the spatial dependency of voxels within the estimation procedure mitigates the necessity to smooth images in order to increase the signal to noise ratio [10]. The spatial dependency of voxels has not been included in available frequentist approaches (apart from smoothing) and only in some available Bayesian approaches. For example in SPM8, the spatial dependency of voxels is considered within the estimation procedure for analysis of fMRI time series at the first level [10] but not for analyses at the second level [27]. Yet, we did not expect our results to be completely independent of smoothing, since it also compensates for imperfect coregistration. Third, more accurate approximation of posterior distributions by MCMC methods should increase both sensitivity and specificity compared to available Bayesian approaches. Of note, all three assumptions comply with the results of

our validation through simulated data. Further, biological validation by analysis of age-related T2-hypointensity yielded plausible results. Compatible with increasing iron content with increasing age [35–37], we found increasing T2-hypointensity with increasing age exclusively, but in all T2-hypointense GM regions (globus pallidus, substantia nigra, red nucleus, dentate nucleus). The striking pronunciation of age-related increase in T2-hypointensity within the dentate nucleus is well explained by the different kinetics of age-related iron accumulation across different GM areas given that the age of our subjects ranged between 20 and 58. In the basal ganglia and thalamus, a significant increase in iron content was observed only after the age of forty [38], while a considerable increase in iron content beginning as early as the age of 20 years was observed in the dentate nucleus [36]. Of note, our approach identified age-related T2-hypointensity better than both the conventional frequentist approach as implemented in SPM8 whereas the Bayesian approach did not work properly. The frequentist approach did not yield meaningful results at the pre-defined statistical threshold of FDR <0.05. Age-related changes of T2-hypointensity could only be visualized at unacceptably liberal statistical thresholds up to an uncorrected p-value of 0.05. Even though our results showed that the simple GMRF prior clearly improves the estimation of regression coefficients, we note that the specified prior for the regression coefficients can have troublesome features [39] and alternative strategies may be more effective.

Moreover, the proposed voxel-wise regression model can be extended in various ways. For example, better edge preserving properties may be achieved by introducing spatially adaptive interaction weights between adjacent voxels [8]. Further, spatial and non-spatial prior information can be combined in order to separate the control over the variance and the effect of neighboring voxels [40]. The use of Diffusion-based spatial priors [41] may also improve the estimation. With regard to possible non-linear relations, more realistic modeling can be achieved by P-Splines [42] or alternative distributional assumptions for the response variable.

In summary, we have demonstrate that fully Bayesian inference can successfully be applied for preprocessing and statistical analysis of structural MRI data.

Acknowledgments

Author Contributions

Conceived and designed the experiments: DB SB MM. Performed the experiments: AF DB SB. Analyzed the data: CG MM PS VS. Contributed reagents/materials/analysis tools: CG AF. Wrote the paper: MM PS.

References

1. Zecca L, Youdim MBH, Riederer P, Connor JR, Crichton RR (2004) Iron, brain ageing and neurodegenerative disorders. Nature reviews Neuroscience 5: 863–73.
2. Bermel RA, Puli SR, Rudick RA, Weinstock-Guttman B, Fisher E, et al. (2005) Prediction of longitudinal brain atrophy in multiple sclerosis by gray matter magnetic resonance imaging T2 hypointensity. Archives of neurology 62: 1371–6.
3. Schmidt P, Gaser C, Arsic M, Buck D, Förschler A, et al. (2012) An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. NeuroImage 59: 3774–83.
4. Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, et al. (1995) Statistical Parametric Maps in Functional Imaging: A General Linear Approach. Human Brain Mapping.
5. Woolrich MW (2012) Bayesian inference in FMRI. NeuroImage 62: 801–10.
6. Gelman A, Carlin JB, Stern HS (2003) Bayesian Data Analysis. Chapman & Hall, second edition.
7. Gössl C, Auer DP, Fahrmeir L (2001) Bayesian Spatiotemporal Inference in Functional Magnetic Resonance Imaging. Biometrics 57: 554–562.
8. Brezger A, Fahrmeir L, Hennerfeind A (2007) Adaptive Gaussian Markov random fields with applications in human brain mapping. Journal of the Royal Statistical Society: Series C (Applied Statistics) 56: 327–345.
9. Woolrich MW, Jenkinson M, Brady JM, Smith SM (2004) Fully Bayesian spatio-temporal modeling of FMRI data. IEEE transactions on medical imaging 23: 213–31.
10. Penny WD, Trujillo-Barreto NJ, Friston KJ (2005) Bayesian fMRI time series analysis with spatial priors. NeuroImage 24: 350–62.
11. Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuro-Image 38: 95–113.
12. Gelman A, Carlin JB, Stern HS, Rubin DB (2003) Mixture Models. In: Bayesian Data Analysis, Chapman & Hall/CRC. Second edition, 464–480.
13. Stephens M (2000) Dealing with label switching in mixture models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62: 795–809.
14. Diebolt J, Robert CP (1994) Estimation of Finite Mixture Distributions through Bayesian Sampling. Journal of the Royal Statistical Society Series B (Methodological) 56: 363–375.
15. Vymazal J, Hajek M, Patronas N, Giedd JN, Bulte JW, et al. (1995) The quantitative relation between T1-weighted and T2-weighted MRI of normal gray matter and iron concentration. Journal of magnetic resonance imaging: JMRI 5: 554–60.
16. Rajapakse JC, Giedd JN, Rapoport JL (1997) Statistical approach to segmentation of singlechannel cerebral MR images. IEEE transactions on medical imaging 16: 176–86.
17. Gelman A, Rubin DB (1992) lnference from Iterative Simulation Using Multiple Sequences. Statistical Science 7: 457–472.
18. Rubin D (1984) Bayesianly justifiable and relevant frequency calculations for the applies statistician. The Annals of Statistics 12: 1151–1172.
19. Gelman A, Meng Xl, Stern H (1996) Posterior Predictive Assessment of Model Fitness via Realized Discrepancies (with discussion). Statistica Sinica 6: 733 – 807.
20. Collins DL, Zijdenbos AP, Kollokian V, Sled JG, Kabani NJ, et al. (1998) Design and construction of a realistic digital brain phantom. IEEE transactions on medical imaging 17: 463–8.
21. Kwan RKS, Evans AC, Pike GB (1996) An Extensible MRI Simulator for Post-Processing Evaluation. Visualization in Biomedical Computing (VBC'96) Lecture Notes in Computer Science 1131: 135–140.
22. Dice L (1945) Measures of the Amount of Ecologic Association Between Species. Ecology 26: 297–302.
23. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC (1994) Morphometric analysis of white matter lesions in MR images: method and validation. IEEE transactions on medical imaging 13: 716–24.
24. Rue H, Held L (2005) Gaussian Markov Random Fields: Theory and Applications, volume 104. Chapman & Hall/CRC, 263 p.
25. Woolrich MW, Behrens TEJ, Smith SM (2004) Constrained linear basis sets for HRF modeling using Variational Bayes. NeuroImage 21: 1748–61.
26. Quirós A, Diez RM, Wilson SP (2010) Bayesian spatiotemporal model of fMRI data using transfer functions. NeuroImage 52: 995–1004.
27. Friston KJ, Penny W (2003) Posterior probability maps and SPMs. NeuroImage 19: 1240–1249.
28. Dhenain M, Duyckaerts C, Michot JL, Volk A, Picq JL, et al. (1998) Cerebral T2-weighted signal decrease during aging in the mouse lemur primate reects iron accumulation. Neurobiology of aging 19: 65–9.
29. Hallgren B, Sourander P (1958) The effect of age on the non-haemin iron in the human brain. Journal of Neurochemistry 3: 41–51.
30. Anbeek P, Vincken KL, van Osch MJP, Bisschops RHC, van der Grond J (2004) Probabilistic segmentation of white matter lesions in MR imaging. NeuroImage 21: 1037–44.
31. Bartko JJ (1991) Measurement and reliability: statistical thinking considerations. Schizophrenia bulletin 17: 483–9.
32. Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press, 625 p.
33. Baker SG (1992) A Simple Method for Computing the Observed Information Matrix When Using the EM Algorithm with Categorical Data. Journal of Computational and Graphical Statistics 1: 63–76.
34. Barkovich AJ (2000) Concepts of myelin and myelination in neuroradiology. AJNR American journal of neuroradiology 21: 1099–109.
35. Bilgic B, Pfefferbaum A, Rohlfing T, Sullivan EV, Adalsteinsson E (2012) MRI estimates of brain iron concentration in normal aging using quantitative susceptibility mapping. NeuroImage 59: 2625–35.
36. Maschke M, Weber J, Dimitrova A, Bonnet U, Bohrenkämper J, et al. (2004) Age-related changes of the dentate nuclei in normal adults as revealed by 3D fast low angle shot (FLASH) echo sequence magnetic resonance imaging. Journal of neurology 251: 740–6.

37. Pfefferbaum A, Adalsteinsson E, Rohlfing T, Sullivan EV (2009) MRI estimates of brain iron concentration in normal aging: comparison of field-dependent (FDRI) and phase (SWI) methods. NeuroImage 47: 493–500.

38. Haacke EM, Miao Y, Liu M, Habib CA, Katkuri Y, et al. (2010) Correlation of putative iron content as represented by changes in R2* and phase with age in deep gray matter of healthy adults. Journal of magnetic resonance imaging: JMRI 32: 561–76.

39. Paciorek CJ (2012) Combining spatial information sources while accounting for systematic errors in proxies. Journal of the Royal Statistical Society: Series C (Applied Statistics) 61: 429–451.

40. Groves AR, Chappell MA, Woolrich MW (2009) Combined spatial and non-spatial prior for inference on MRI time-series. NeuroImage 45: 795–809.

41. Harrison LM, Penny W, Ashburner J, Trujillo-Barreto N, Friston KJ (2007) Diffusion-based spatial priors for imaging. NeuroImage 38: 677–95.

42. Lang S, Brezger A (2004) Bayesian P-Splines. Journal of Computational and Graphical Statistics 13: 183–212.