HHS Public Access

Author manuscript

Stat Biosci. Author manuscript; available in PMC 2020 April 01.

Published in final edited form as:

Stat Biosci. 2019 April ; 11(1): 22–46.

A Spatio-Temporal Model for Longitudinal Image-on-Image Regression

Arnab Hazra[1], **Brian J. Reich**[1], **Daniel S. Reich**[2], **Russell T. Shinohara**[3], and **Ana-Maria Staicu**[1]

[1]North Carolina State University, Raleigh, NC, USA

[2]National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA

[3]University of Pennsylvania, Philadelphia, PA, USA

Abstract

Neurologists and radiologists often use magnetic resonance imaging (MRI) in the management of subjects with multiple sclerosis (MS) because it is sensitive to inflammatory and demyelinative changes in the white matter of the brain and spinal cord. Two conventional modalities used for identifying lesions are T1-weighted (T1) and T2-weighted fluid-attenuated inversion recovery (FLAIR) imaging, which are used clinically and in research studies. Magnetization transfer ratio (MTR), which is available only in research settings, is an advanced MRI modality that has been used extensively for measuring disease-related demyelination both in white matter lesions as well across normal-appearing white matter. Acquiring MTR is not standard in clinical practice, due to the increased scan time and cost. Hence, prediction of MTR based on the modalities T1 and FLAIR could have great impact on the availability of these promising measures for improved patient management. We propose a spatio-temporal regression model for image response and image predictors that are acquired longitudinally, with images being co-registered within the subject but not across subjects. The model is additive, with the response at a voxel being dependent on the available covariates not only through the current voxel but also on the imaging information from the voxels within a neighboring spatial region as well as their temporal gradients. We propose a dynamic Bayesian estimation procedure that updates the parameters of the subject-specific regression model as data accummulates. To bypass the computational challenges associated with a Bayesian approach for high-dimensional imaging data, we propose an approximate Bayesian inference technique. We assess the model fitting and the prediction performance using longitudinally acquired MRI images from 46 MS patients.

Keywords

spatio-temporal regression model; longitudinal imaging study; dynamic Bayesian updating; multiple sclerosis; magnetization transfer ratio; T1-weighted; T2-weighted fluid-attenuated inversion recovery; composite likelihood

1 Introduction

In modern medicine, neuroimaging is increasingly critical in the management of patients with neurological diseases. Magnetic resonance imaging (MRI) is non-invasive and sensitive

to both the structural and functional changes in brain disorders. In this work, we will focus on structural MRI (see [49]). Common structural MRI modalities include T1-weighted (T1) and T2-weighted fluid attenuated inversion recovery (FLAIR) imaging, which give complementary information about brain structure. These measures, however, are not specific for the disease-related changes in the white matter brain; thus, significant effort in the neuroimaging community has focused on more specifically measuring changes in structure including demyelination. Of the modalities that purport to measure myelin, magnetization transfer ratio (MTR) imaging has the longest history and has been used extensively to study differences across the brain. We are interested in studying the relationship between MTR and the conventional T1 and FLAIR imaging modalities for monitoring changes in the white matter of subjects with multiple sclerosis (MS).

MS is an immune-mediated inflammatory disease that affects the central nervous system, and is associated with damage to the myelin sheath that insulates neurons. This occurs both in focal white matter lesions as well as across normal-appearing white matter, which causes disruption in the information connections between the various part of the brain and the rest of the body. Its diagnosis is rather complex and relies heavily on structural MRI of the brain and spinal cord ([24]; [30]). Two conventional modalities commonly used for this purpose are T1 and FLAIR. These images are easy to acquire and relatively inexpensive; however there has been criticism on the ground that they have lower specificity ([22]; [4]; [37]). On the other hand, MTR is thought to better quantify and identify brain tissue image, and thus has been proposed as a better alternative to aid with the diagnosis and quantifying tissue damage in MS ([17]; [14]). Nevertheless acquiring MTR is not standard practice, due to the increased time and financial cost involved in the process.

In this article, we propose to fill in these gaps by proposing a novel prediction model of MTR from the conventionally acquired modalities, T1 and FLAIR. This problem has been recently considered by [31], who restricted their study to the lesions only and proposed a voxel-wise approach. While this analysis is helpful in gaining understanding of the relationship between the three modalities, it has limited applicability. In contrast, we focus on the normal-appearing white matter, which consists of myelinated nerve cells projections outside of focal lesions as apparent on MRI. It is well documented that normal-appearing white matter in subjects with MS is often the site of diffuse changes associated with the disease, and MTR has been used in the past to study these changes ([7], [16], [26]). Like [31] our goal is to use past imaging information; however unlike them, we want to account for and incorporate the dependence of the neighboring voxels.

Imaging regression has attracted recent increasing interest, where the images are used as predictors or response or both. When the predictor is an image and the response is scalar, recent literature on scalar-on-image regression includes [15] and [21] - who use a Bayesian framework - and [35] who considers a frequentist viewpoint. When the response is an image, a popular approach has been to consider a mass univariate analysis; see for example [13], [1], [44]. This technique typically consists of two-stage parts: first, a voxel-wise regression model is fitted and second, the voxel-dependent coefficients are analyzed. Modern approaches borrow information from the spatial dependence across brain locations by either

modeling the spatial dependence directly (see [38]) or employing appropriate penalty criterions (see [8], [50]).

Image regression when both the response and predictors are images is somewhat more limited; a few are [28], [47], [36], [19], [20]. Past work on predicting lesional activity focused on voxel-level regression modeling ([41], [32]). More recently [25] proposed Quantitative MR Estimation Employing Normalization (QuEEN) regression approach for estimating quantitative T1 maps based on conventional imaging, and [46] implemented the QuEEN model for estimating four diffusion tensor imaging (DTI) measures (mean diffusivity, fractional anisotropy, radial diffusivity and axial diffusivity) based on conventional MRI images. In related work, [9] proposed a simple voxel-wise regression method for predicting the recovery of white matter lesions by integrating imaging and clinical information observed at lesion incidence in MS patients. [3] proposed a longitudinal multivariate nonlinear mixed effects model considering spatial correlation; though they applied their method to lower-dimensional (30,000 voxels for each subject on an average) brain images. Nevertheless, neither one of these approaches accounts for the spatial dependence across voxels in the brain or considers very high-dimensional spatial datasets. Additionally, to the best of the authors' knowledge, this is the first modeling attempt to incorporate neighboring covariate information along with spatial correlation for very high-dimensional spatial datasets.

Motivated by a longitudinal MRI study of MS we propose a spatio-temporal additive model, where the response intensity at a current voxel is related to the intensities of the image covariates as well as to the changes in the intensities since the most recent visit at the current and neighboring voxels. Our model assumes that the effect of the covariates at nearby voxels varies smoothly with the distance to these voxels. One important challenge when working with images is registration. In our study, the MS subjects are observed over many hospital visits, and at each visit, they are imaged using multiple modalities (such as T1, FLAIR, and MTR). The scans are linearly co-registered across modalities, as well as within subject; however they are not registered across subjects. Not registering the images across subjects is appealing for studies in which structural pathologies are present, as registration may cause distortion of lesions which could potentially result in misleading imaging signatures ([11]). To accommodate this, it is assumed that the model parameters vary across subjects. We consider a Bayesian framework to describe the variability of the subject-specific smooth parameters. A major disadvantage of Bayesian modeling is that the computation is often infeasible for high-dimensional MRI datasets, as this one. We consider an approximate Bayesian alternative that has computational advantages. We propose to estimate the subject-specific parameters dynamically; the results with this approach are contrasted with the counterparts based on a static estimation. Prediction of the response is readily obtained and quantification of the prediction uncertainty is based on the posterior distribution.

The paper is organized as follows. We describe the MRI-based study of MS in Section 2. In Section 3 we present the proposed spatio-temporal image-on-image regression model. Parameter estimation and prediction using both dynamic and static approaches is discussed in Section 4. Section 5 includes the analysis of the MRI-based study of MS: it argues the various model fitting choices and compares the proposed model with simpler alternatives

that do not incorporate the spatial dependence or the temporal dependence or both, or do not incorporate the anatomical structural information of white matter. A simulation study which justifies numerically the approximate Bayesian inference technique is provided in Section 6. Final remarks are included in Section 7.

2 MRI dataset

The motivating application is a longitudinal study of MS carried out by the National Institute of Neurological Disease and Stroke: MS patients are imaged using structural MRIs at baseline and many follow up visits. Details concerning the study and acquisition parameters are available in [48], and we summarize them briefly here. Subjects were scanned approximately once per month between 2000 and 2008 with standard deviation (SD) 52.3 days, with an average of 21 (SD 8) scans per subject. The earliest and the most recent scans were performed in 2000 and 2008 respectively. Whole-brain MTR, T1 and FLAIR volumes were acquired in a 1.5T GE scanner using clinically optimized scanning parameters. All the modalities were interpolated to a voxel size of 1 mm^3 yielding images of dimension 182×218×182. All images were registered longitudinally and across the modalities and rigidly aligned to the Montreal Neurological Institute standard space ([12]); images were otherwise not registered across the subjects. Extracerebral voxels were removed using a skull-stripping procedure ([5]), and normal-appearing white matter (NAWM) was identified automatically using the T1 and FLAIR images ([39]).

We use the ICBM-DTI-81 white-matter atlas ([27]) to identify the anatomical regions of white matter of a person. We project the atlas labels to a person's white matter space using the R packages ANTsR ([2]) and extrantsr ([29]). There are 50 labels in the atlas. Details about the labels are available at http://www.loni.usc.edu/ICBM/Downloads/Downloads_DTI-81.shtml. Some of the neighboring regions, mostly in the lower part of the brain, are very small and we merge them; we obtain 32 labels. Specifically we combine the labels 1–2, 7–16, 31–32 and 47–48 together (into label 1), labels 5–6 (into label 4), labels 37 and 39 (into label 23) and labels 38 and 40 (into label 24).

The study contains information on 60 subjects; all images were inspected visually for quality, and excluded any low-quality scans due to subject motion or preprocessing errors from analysis. After quality assurance, the remaining dataset consists of structural MRIs from a total of 46 subjects observed at 802 hospital visits. For the included patients, the average number of scans per subject is 18 with SD 7.0 and range [7, 34]. The average length of the study periods is 21 months (SD 12 months, range [252 days, 5.5 years]). The average lag between two consecutive scans varies between 27 days and 71 days across the subjects. The number of NAWM voxels vary between 164,070 and 587,768. The histograms of the visit times, the study periods, the average lag and number of NAWM voxels for each subject are provided in Figure 1.

For every scan, we normalize T1 and FLAIR images across the NAWM voxels following the *z*-score normalization method ([40], [42]) where we estimate the mean and variance of the predictor intensities across the NAWM voxels and then normalize the predictors at each voxel by subtracting the mean and dividing by the standard deviation. The MTR, normalized

T1 and FLAIR profiles for a representative subject on 28-th day (2nd visit), 105-th day (5th visit), 161-th day (8th visit) and 273-th day (11th visit) at the brain slice of Z-dimension fixed at $Z = 84$ are provided in Figure 2.

3 Statistical framework

Let $Y_{ij}(\mathbf{v})$ denote the response for subject i at visit j in voxel $\mathbf{v} \in \mathbf{V}_i$, where $\mathbf{V}_i$ denotes the set of voxels for subject i and t_{ij} denotes the j-th visit time (in days) of that subject since the first visit, thus $t_{i1} = 0$. The visit times are irregularly spaced and vary across subjects. Denote the number of scans available from subject i by J_i and assume there are P modalities, $M_{ijp}(\mathbf{v})$, $p = 1, \ldots, P$, whose effect on the response we are interested to study. In our application, Y_{ij} is the MTR intensity for MS subject i at time t_{ij}, $P = 2$, and M_{ij1} and M_{ij2} correspond to T1 and FLAIR respectively. The objective is to predict the current response image for a subject, given all their imaging history, currently available modalities, and data from other subjects.

We consider a mean model of the form $E[Y_{ij}(\mathbf{v})] = \mu_{ij}(\mathbf{v})$ for $j \geq 2$, where

$$\mu_{ij}(\mathbf{v}) = \alpha_{i0} + \alpha_i' h(\mathbf{v}) + \sum_{p=1}^{P} \sum_{d \in \mathcal{D}} \beta_{ip}(d) \overline{M}_{ijp}^{(d)}(\mathbf{v}) + \sum_{p=1}^{P} \sum_{d \in \mathcal{D}} \gamma_{ip}(d) \Delta \overline{M}_{ijp}^{(d)}(\mathbf{v}) . \quad (1)$$

Here $h(\mathbf{v}) = (\mathbf{v} - \mathbf{V}_i^{min}) \oslash (\mathbf{V}_i^{max} - \mathbf{V}_i^{min})$ is the rescaled voxel location where $\mathbf{V}_i^{max}$ and $\mathbf{V}_i^{min}$ are the dimension-wise maximum and minimum coordinate 3-tuples of $\mathbf{V}_i$ ($\oslash$ stands for element-wise division of two vectors); $h(\mathbf{v})$ acts as an inhomogeneity correction when analyzing images that are not registered across subjects. The set $\mathbf{V}_i$ has a grid-like structure and the set of neighboring points of $\mathbf{v} = (v_1, v_2, v_3)$ are $\{(v_1 + x, v_2 + y, v_3 + z) : x, y, z = 0, \pm 1, \pm 2, \ldots\}$; thus the possible Euclidean distances are 0, 1, $\sqrt{2}$, $\sqrt{3}$, 2, $\sqrt{5}$, $\sqrt{6}$ etc. We restrict our study to neighborhoods of the form $\mathbf{v} \pm (x, y, z)$ where $x, y, z \in \{0, \pm 1, \pm 2\}$, which leads to a set of voxel-distances $\mathcal{D} = \{0, 1, \sqrt{2}, \sqrt{3}, 2, \sqrt{5}, \sqrt{6}, 2\sqrt{2}, 3, 2\sqrt{3}\}$. This choice is made mainly for computational convenience; however we also noticed that larger neighborhoods run the risk of including high variance inflation factors in some cases. Using the selected neighborhood, most of values are considerably small. The term $\overline{M}_{ijp}^{(d)}(\mathbf{v}) = \sum_{\mathbf{v}' \in N_{id}(\mathbf{v})} M_{ijp}(\mathbf{v}') / |N_{id}(\mathbf{v})|$ is the average intensity for voxels that are at distance d from voxel $\mathbf{v}$ corresponding to the pth image modality and $N_{id}(\mathbf{v})$ is the set of available voxels with this property. The term $\Delta \overline{M}_{ijp}^{(d)}(\mathbf{v})$ is obtained similarly by averaging the temporal gradients $\Delta M_{ijp}(\mathbf{v}') = [M_{ijp}(\mathbf{v}') - M_{i(j-1)p}(\mathbf{v}')]/(t_{ij} - t_{i(j-1)})$, where $\Delta M_{ijp}(\mathbf{v}')$ describes the relative change of the image predictor at voxel v' over time. Here α_{i0} is a subject-specific intercept, α_i is the subject-specific vector of regression coefficients corresponding to the rescaled coordinates; it can be interpreted as an inhomogeneity correction ([43]; [23]). Finally $\beta_{ip}(\cdot)$ and $\gamma_{ip}(\cdot)$ quantify the subject-specific effects of the image covariates at the neighboring voxels; these effects are assumed to vary smoothly with the distance between voxels.

It is reasonable to assume that the effect of the predictors on the response may vary across the brain (e.g. different anatomical regions of white matter in case of MRI dataset). To account for this reasonable assumption, we consider a partition of the brain and allow for the model parameters to vary across the partition sub-regions. Let $\mathbf{V}_i = \cup_{k=1}^{k=K} \mathbf{V}_{ik}$ and use subscript k for the model parameters to indicate their restriction to the sub-region $\mathbf{V}_{ik}$. This leads us to posit the following model, for $\mathbf{v} \in \mathbf{V}_{ik}$ for $j \geq 2$:

$$Y_{ij}(\mathbf{v}) = \mu_{ijk}(\mathbf{v}) + \varepsilon_{ij}(\mathbf{v}), \quad j = 2, \ldots, J_i$$

$$\mu_{ijk}(\mathbf{v}) = \alpha_{i0k} + \alpha'_{ik} h(\mathbf{v}) + \sum_{p=1}^{P} \sum_{d \in \mathscr{D}} \beta_{ipk}(d) \overline{M}_{ijp}^{(d)}(\mathbf{v}) + \sum_{p=1}^{P} \sum_{d \in \mathscr{D}} \gamma_{ipk}(d) \Delta \overline{M}_{ijp}^{(d)}(\mathbf{v}), \quad (2)$$

where $k = 1, \ldots, K$, the subject-specific random effects, α_{i0k} and α_{ik}, vary across sub-regions, and the subject-specific smooth functions $\beta_{ipk}(\cdot)$ and $\gamma_{ipk}(\cdot)$ quantify the local effect of the image predictors. Here, the dependence of $\mu(\mathbf{v})$ on subscript k is redundant as $\mathbf{v} \in \mathbf{V}_{ik}$ and we drop it later in the paper; however we use it above for clarity, as some of the terms on the right hand side depend solely on k and do not involve $\mathbf{v}$. We assume that given the set of available covariates, the residual term $\varepsilon_{ij}(\mathbf{v})$ is independent across visits j; see also [31].

We assume that the error terms are independent across partitions and consider a Matérn correlation ([45]) within partition. Specifically, we assume that $\varepsilon_{ij}(\cdot)$ is a spatially correlated Gaussian process with mean zero and covariance for $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V}_{ik}$ described by the following:

$$cov\{\varepsilon_{ij}(\mathbf{v}_1), \varepsilon_{ij}(\mathbf{v}_2)\} = \frac{\sigma_{ik}^2}{\Gamma(\nu_{ik}) 2^{\nu_{ik}-1}} \left(\frac{h}{\rho_{ik}}\right)^{\nu_{ik}} K_{\nu_{ik}}\left(\frac{h}{\rho_{ik}}\right) + \tau_{ik}^2 I(\mathbf{v}_1 = \mathbf{v}_2) \quad (3)$$

where $h = \|\mathbf{v}_1 - \mathbf{v}_2\|$ is the Euclidean distance between $\mathbf{v}_1$ and $\mathbf{v}_2$, σ_{ik}^2, ρ_{ik}, ν_{ik} and τ_{ik}^2 are subject-specific, time-invariant partial sill, range, smoothness, and nugget parameters respectively. Also $K_{\nu_{ik}}$ is the Modified Bessel function of degree ν_{ik} and $I(\mathbf{v}_1 = \mathbf{v}_2) = 1$ if $\mathbf{v}_1 = \mathbf{v}_2$ and 0 otherwise.

The proposed model specified by (2) and (3) describes the variability of the response at the subject-level; we refer to this part by *stage I* of the hierarchical model. The model specification will be completed with a second stage that will describe how all the subject-specific random effects vary in the population; these assumptions describe how the subject-specific random effects vary across sub-regions k and modalities p.

Denote by $\mathbf{\Phi}_{ik} = (\log(\sigma_{ik}^2), \log(\rho_{ik}), \log(\nu_{ik}), \log(\tau_{ik}^2))'$ the subject-specific covariance-related random effects on a log scale and assume a normal distribution for how they vary across

subjects and sub-regions. For simplicity of exposition, in order to describe the variability of the subject-specific mean random effects in a way that accounts for dependence across modalities' effects as well as dependence across partitions, we first describe the modeling of the smooth random effects and then specify their variability across subjects.

The smooth effects, $\beta_{ipk}(\cdot)$ and $\gamma_{ipk}(\cdot)$, are modeled using regression splines; this is a common practice to reduce the problem dimensionality in semi-parametric regression literature ([34]). For exposition convenience we consider the same basis functions for both $\beta_{ipk}(\cdot)$ and $\gamma_{ipk}(\cdot)$ throughout the paper and model them as linear combinations of L B-spline basis functions, $\beta_{ipk}(d) = \sum_{l=1}^{L} \theta_{ilpk} B_l(d)$ and $\gamma_{ipk}(d) = \sum_{l=1}^{L} \theta_{il(P+p)k} B_l(d)$. The choice of L and the order of splines are problem-specific; a large value of L increases computational burden significantly.

By combining the modalities with the basis functions, let $X^{\overline{M}}_{ijlp}(\mathbf{v}) = \sum_{d \in \mathscr{D}} B_l(d) \overline{M}^{(d)}_{ijp}(\mathbf{v})$ and $X^{\Delta\overline{M}}_{ijlp}(\mathbf{v}) = \sum_{d \in \mathscr{D}} B_l(d) \Delta\overline{M}^{(d)}_{ijp}(\mathbf{v})$; the mean simplifies to:

$\mu_{ij}(\mathbf{v}) = \alpha_{i0k} + \alpha'_{ik} h(\mathbf{v}) + \sum_{p=1}^{P} \sum_{l=1}^{L} \theta_{ilpk} X^{\overline{M}}_{ijlp}(\mathbf{v}) + \sum_{p=1}^{P} \sum_{l=1}^{L} \theta_{il(P+p)k} X^{\Delta\overline{M}}_{ijlp}(\mathbf{v})$ for $\mathbf{v} \in \mathbf{V}_{ik}$. Model (2) can be represented simpler using matrix notation. By an abuse of notation, denote $\mathbf{X}_{ijk}$ for the design matrix with a number of rows equal to the cardinality of the set $\mathbf{V}_{ik}$, let $\mathbf{\Theta}_{ik} = (\alpha_{i0k}, \alpha'_{ik}, \theta_{i11k}, \ldots, \theta_{i1Lk}, \ldots, \theta_{i(2P)Lk})'$ the $(4 + 2PL)$-dimensional vector parameter corresponding to the i-th subject and kth sub-region, and let $\mathbf{E}_{ijk}$ be the vector of residuals $\varepsilon_{ij}(\mathbf{v})$ for $\mathbf{v} \in \mathbf{V}_{ik}$. Model (2), with specification (3), is approximated by:

$$\mathbf{Y}_{ijk} = \mathbf{X}_{ijk} \mathbf{\Theta}_{ik} + \mathbf{E}_{ijk}; \quad (4)$$

$\mathbf{Y}_{ijk}$ is the vector of responses $Y_{ij}(\mathbf{v})$ for $\mathbf{v} \in \mathbf{V}_{ik}$ and $\mathbf{E}_{ijk}$'s are mutually independent and have a multivariate normal distribution with zero-mean and Matérn covariance matrix described by $\mathbf{\Phi}_{ik}$, denoted $\Sigma_{ik}(\mathbf{\Phi}_{ik})$.

Next we consider the variability of the subject-specific random effects in population; to focus ideas we start with the mean random effects. Let k and k' correspond to two brain sub-regions and let $\Theta^{(q)}_{ik}$ and $\Theta^{(q')}_{ik'}$ two arbitrary components of the vectors $\mathbf{\Theta}_{ik}$ and $\mathbf{\Theta}_{ik'}$ respectively. We assume that their dependence is described by $Cov(\Theta^{(q)}_{ik}, \Theta^{(q')}_{ik'}) = \Sigma_{1kk'} \Sigma_{2qq'}$, where $\Sigma_{1kk'}$ and $\Sigma_{2qq'}$ denote the (k, k')-th and the (q, q')-th elements of a $K \times K$ covariance matrix Σ_1 and $Q \times Q$ covariance matrix Σ_2 respectively, where $Q = (4 + 2PL)$. This assumption is equivalent to saying that if $\mathbf{\Theta}_i = (\mathbf{\Theta}^T_{i1}, \ldots, \mathbf{\Theta}^T_{iK})^T$ is the vector of all the mean random effects $\mathbf{\Theta}_{ik}$'s then $Var\{\mathbf{\Theta}_i\} = \Sigma_1 \otimes \Sigma_2$; in other words we assume a separable covariance matrix for the mean effects; see [18]. This choice is primarily made for computational convenience as it greatly helps to reduce the parameter dimensionality. However, it also makes sense from an interpretation perspective: it implies that the dependence structure is preserved across partitions by some factor. Following similar logic,

we assume that the subject-specific covariance-related random effects, $\mathbf{\Phi}_i = (\mathbf{\Phi}_{i1}^T, \ldots, \mathbf{\Phi}_{iK}^T)^T$, are dependent according to a separable covariance structure. In addition we assume that the mean and covariance random effects are mutually independent. We propose to model the variability of the subject-specific random terms in the population as

$$\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_I \overset{IID}{\sim} N(\Lambda_{\mathbf{\Theta}}, \Sigma_1 \otimes \Sigma_2)$$

$$\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_I \overset{IID}{\sim} N(\Lambda_{\mathbf{\Phi}}, \Omega_1 \otimes \Omega_2) \quad (5)$$

$\mathbf{\Theta}_i, \mathbf{\Phi}_{i'}$ are mutually independent

where $\Lambda_{\mathbf{\Theta}}$ and $\Lambda_{\mathbf{\Phi}}$ are the unknown mean vectors and Ω_1 and Σ_1 are unknown covariance matrices of dimensions $K \times K$, and Σ_2 and Ω_2 are unknown covariances of dimensions $Q \times Q$ and 4×4 respectively; here 4 is the number of the spatial covariance parameters. We refer to $\Lambda_{\mathbf{\Theta}}$, Σ_1, and Σ_2 as mean-related parameters and to Λ_{Φ}, Ω_1, and Ω_2 as covariance-related parameters.

The specification (5) is referred by as *stage II* and completes the description of the proposed model. As expected, the separable covariance assumption has identifiability issues in the sense that the scales of the matrices Σ_1 and Σ_2 are identifiable only up to their product; the situation is similar for Ω_1 and Ω_2. However this identifiability issue does not affect our analysis, as the inference about the model parameters $\mathbf{\Theta}_i$'s and implicitly of the parameter functions $\beta_{ikp}(\cdot)$ and $\gamma_{ikp}(\cdot)$ depends only on the Kronecker product and not on its individual components.

The subject-specific effects' dependence implies a far more complex dependence among the random effect functions; for example

$$Cov\{\beta_{ikp}(d), \beta_{ik'p'}(d')\} = \sum_l \sum_{l'} B_l(d) B_{l'}(d') Cov(\theta_{ilpk}, \theta_{il'p'k'}) = \sum_{1kk'} \mathbf{B}'(d) \sum_2^{(p,p')} \mathbf{B}(d')$$

where $\mathbf{B}(d) = (B_1(d), \ldots, B_L(d))'$ and $\sum_2^{(p,p')}$ is the $L \times L$ matrix formed by the elements at the intersection of the columns $5 + (p-1)L, \ldots, 4 + pL$ and the rows $5 + (p'-1)L, \ldots, 4 + p'L$ of the matrix Σ_2. Using similar algebra calculations we have $Cov\{\beta_{ikp}(d), \gamma_{ik'p'}(d')\} = \sum_{1kk'} \mathbf{B}'(d) \sum_2^{(p,P+p')} \mathbf{B}(d')$. One immediate consequence of the separability assumption is that the dependence structure is similar across partitions, up to a constant term that varies across sub-regions.

4 Parameter estimation and prediction

Estimation of the subject-specific random effects is typically done through conditional expectation; prediction of the response is obtained directly by plugging in these estimates in the systematic component of (4). Later, we contrast two approaches: a *static* approach that does not use the past imaging information of a patient in estimating the model parameters and a *dynamic* that accounts for the subject's past information, when such information is available. In general, if the subject random effects $\mathbf{\Phi}_{ik}$'s are known, simple algebra yields the following analytical expression for the conditional distribution of $\mathbf{\Theta}_i$ given all the data for the ith subject, the model parameters $\Lambda_{\mathbf{\Theta}}$, Σ_1 and Σ_2:

$$\mathbf{\Theta}_i | Rest_i, \sim N\left[\left(\sum_1^{-1} \otimes \sum_2^{-1} + \Delta_X^{(i)}\right)^{-1}\left(\sum_1^{-1} \otimes \sum_2^{-1}\Lambda_{\mathbf{\Theta}} + \Delta_Y^{(i)}\right), \left(\sum_1^{-1} \otimes \sum_2^{-1} + \Delta_X^{(i)}\right)^{-1}\right] \quad (6)$$

where $\Delta_X^{(i)} = diag(C_{i1}, \ldots C_{iK})$ with $C_{ik} = \sum_{j=1}^{J_i} \mathbf{X}'_{ijk} \sum_{ik} (\mathbf{\Phi}_{ik})^{-1} \mathbf{X}_{ijk}$ and $\Delta_Y^{(i)} = (d'_{i1}, \ldots d'_{iK})'$, where $d_{ik} = \sum_{j=1}^{J_i} \mathbf{X}'_{ijk} \sum_{ik} (\mathbf{\Phi}_{ik})^{-1} \mathbf{Y}_{ijk}$ for $k = 1, \ldots, K$. Also $Rest_i$ includes the data for the ith subject and the mean-repated population-level parameters $\Lambda_{\mathbf{\Theta}}$, Σ_1 and Σ_2. Thus prediction of $\mathbf{\Theta}_i$ in this way requires estimation of these unknown quantities. The conditional distribution of $\mathbf{\Phi}_{ik}$ given the data for the ith subject and remaining parameters does not have a close form.

We propose a Bayesian approach for parameter estimation and inference, and ultimately for estimation of the current subject-level image given all their past imaging information. We consider conjugate non-informative priors for the model parameters:

$$\Lambda_{\mathbf{\Theta}} | \Sigma_1, \Sigma_2 \sim N(\eta_{\mathbf{\Theta}}, c^{-1} \Sigma_1 \otimes \Sigma_2)$$

$$\sum_1 \sim IW(\nu_{\sum_1}, \Psi_{\sum_1}) \text{ and } \sum_2 \sim IW(\nu_{\sum_2}, \Psi_{\sum_2}) \text{ mutually independent} \quad (7)$$

and $\Lambda_{\mathbf{\Phi}} | \Omega_1, \Omega_2 \sim N(\eta_{\mathbf{\Phi}}, c^{-1}\Omega_1 \otimes \Omega_2)$, $\Omega_1 \sim IW(\nu_{\Omega_1}, \Psi_{\Omega_1})$ and $\Omega_2 \sim IW(\nu_{\Omega_2}, \Psi_{\Omega_2})$ are mutually independent, where IW is the inverse Wishart distribution and hyper-parameters are set to give uninformative priors: $\eta_\Theta = \mathbf{0}_{KQ}$, $\eta_\Phi = \mathbf{0}_{4K}$, $c = 10^{-4}$, $\nu_{\sum_1} = \nu_{\sum_2} = \nu_{\Omega_1} = \nu_{\Omega_2} = 10^{-2}$, $\Psi_{\sum_1} = \Psi_{\Omega_1} = 10^{-2} I_K$, $\Psi_{\sum_2} = 10^{-2} I_Q$, $\Psi_{\Omega_2} = 10^{-2} I_4$.

The full conditional distributions for the mean-related parameters $\Lambda_{\mathbf{\Theta}}$, Σ_1 and Σ_2 given the data and all the remaining parameters on which it is conditioned are

$$\Lambda_{\boldsymbol{\Theta}}|Data, \boldsymbol{\Theta}_i\text{'s}, \sum_1, \sum_2 \sim N\left(\frac{I\overline{\boldsymbol{\Theta}} + c\eta_{\boldsymbol{\Theta}}}{I+c}, \frac{1}{I+c}\sum_1 \otimes \sum_2\right) \quad (8)$$

$$\sum_1|Data, \boldsymbol{\Theta}_i\text{'s}, \Lambda_{\boldsymbol{\Theta}}, \sum_2 \sim IW\left[\nu_{\sum_1} + Q(I+1), \boldsymbol{\Psi}_{\sum_1} + c(\Lambda_{\boldsymbol{\Theta}} - \eta_{\boldsymbol{\Theta}})'\sum_2^{-1}(\Lambda_{\boldsymbol{\Theta}} - \eta_{\boldsymbol{\Theta}}) + S_1\right]$$

$$\sum_2|Data, \boldsymbol{\Theta}_i\text{'s}, \Lambda_{\boldsymbol{\Theta}}, \sum_1 \sim IW\left[\nu_{\sum_2} + K(I+1), \boldsymbol{\Psi}_{\sum_2} + c(\Lambda_{\boldsymbol{\Theta}} - \eta_{\boldsymbol{\Theta}})\sum_1^{-1}(\Lambda_{\boldsymbol{\Theta}} - \eta_{\boldsymbol{\Theta}})' + S_2\right]$$

where $\overline{\boldsymbol{\Theta}} = \frac{1}{I}\sum_{i=1}^{I}\boldsymbol{\Theta}_i$, $\boldsymbol{\Theta}_{i,mat} = [\boldsymbol{\Theta}_{i1}| \ldots |\boldsymbol{\Theta}_{iK}]$ is $Q \times K$ matrix for $i = 1, \ldots, I$, $\Lambda_{\Theta,mat} = E(\boldsymbol{\Theta}_{i,mat})$, $S_1 = \sum_{i=1}^{I}(\boldsymbol{\Theta}_{i,mat} - \Lambda_{\boldsymbol{\Theta},mat})'\sum_2^{-1}(\boldsymbol{\Theta}_{i,mat} - \Lambda_{\boldsymbol{\Theta},mat})$, and $S_2 = \sum_{i=1}^{I}(\boldsymbol{\Theta}_{i,mat} - \Lambda_{\boldsymbol{\Theta},mat})\sum_1^{-1}(\boldsymbol{\Theta}_{i,mat} - \Lambda_{\boldsymbol{\Theta},mat})'$. Estimation of the mean-related parameters is carried via an iterative approach together with (6) and involves sampling from these distribution, which can be done using Gibbs sampling. However, the conditional distributions of $\boldsymbol{\Lambda}_{\boldsymbol{\Phi}}$, Ω_1 and Ω_2, like the one of $\boldsymbol{\Phi}_i$'s, given the data and the appropriate set of parameters cannot be obtained analytically. Thus, sampling from these distributions would require Metropolis-Hastings (MH) algorithm. While such an approach is feasible for moderately large datasets, it becomes increasingly impractical as the dimensionality of the data grows larger. The MH steps get very challenging because of cumbersome covariance matrix calculations. In such cases, empirical Bayesian method, which relies on using point estimates of some parameters and ignoring their uncertainty, is one commonly employed alternative.

Maximum likelihood estimation is used separately for each subject using model (4), and the maximum likelihood estimates (MLE) $\widehat{\boldsymbol{\Phi}}_i$; $i = 1, \ldots, I$ are plugged-in for $\boldsymbol{\Phi}_i$'s; see [6]. For very-high dimensional data, such as our MRI data, obtaining the MLEs is unfeasible, as the dimensions of the spatial domains are in the range 0.3 – 0.5 million. To bypass this challenge, we propose estimation based on a block composite likelihood structure ([10]) approach within each partition. Once MLE estimates for the covariance parameters $\boldsymbol{\Phi}_{ik}$'s are available, $\widehat{\boldsymbol{\Phi}}_{ik}$, the estimated covariance matrices obtained by using these plug-in empirical estimates, $\sum_{ik}(\widehat{\boldsymbol{\Phi}}_{ik})$, are used to calculate C_{ik}'s and d_{ik}'s and ultimately in the Gibbs sampling steps in (8) that are used to estimate $\boldsymbol{\Theta}_i$'s and the mean-related parameters $\Lambda_{\boldsymbol{\Theta}}$, Σ_1 and Σ_2.

The approach outlined above is based on all the subjects' data. As is often the case, in our study the objective is to predict the current response for a subject using their available past information, which includes both the response and covariates at previous times. We discuss two slight modifications of the above approach that explore the advantage of incorporating available subject information in estimating the mean parameters, $\boldsymbol{\Theta}_i$. First, we introduce

additional notation. Let i index an arbitrary subject for which we are interested to predict their response and denote by $Data_{-i}$ the data obtained from all the subjects except subject i. Assume the ith subject has been observed at visits 2, …, $(j-1)$, denote by $Data_i^{2:(j-1)}$ the information comprised of the response and covariates from all these visits. We use the convention that if $j = 2$ then $Data_i^{2:(j-1)}$ is the void set. Thus, predicting $\boldsymbol{\Theta}_i$ using the conditional distribution described in (6) is only defined for $j \geq 3$.

Dynamic estimation

Assume $j \geq 3$. Parameter estimation involves iterative sampling from the conditional distributions described by (6) and (8) until convergence. There are few key differences. First, the data $Data_i^{2:(j-1)}$ is used in step (6) and the covariance-related random effects for the ith subject $\boldsymbol{\Phi}_i$ are estimated based on the subject's available data. In particular $C_{ik} = \sum_{j'=2}^{j-1} \mathbf{X}'_{ij'k} \Sigma_{ik} (\widehat{\boldsymbol{\Phi}}_{ik})^{-1} \mathbf{X}_{ij'k}$ and $d_{ik} = \sum_{j'=2}^{j-1} \mathbf{X}'_{ij'k} \Sigma_{ik} (\widehat{\boldsymbol{\Phi}}_{ik})^{-1} \mathbf{Y}_{ij'k}$ for $k = 1, \ldots, K$. Ideally, the estimates of the covariance-related subject-specific effects $\boldsymbol{\Phi}_{ik}$'s are obtained using MLE based on the data $Data_i^{2:(j-1)}$; however such approach entails heavy computational burden. An appealing alternative is to obtain the MLE estimates at each visit j', $\widehat{\boldsymbol{\Phi}}_{ik}^{j'}$, for $2 \leq j' \leq (j-1)$; the subscript j' reflects the dependence of these terms on the respective visit. Thus $C_{ik} \approx \sum_{j'=2}^{j-1} \mathbf{X}'_{ij'k} \Sigma_{ik} (\widehat{\boldsymbol{\Phi}}_{ik}^{j'})^{-1} \mathbf{X}_{ij'k}$ and similarly $d_{ik} \approx \sum_{j'=2}^{j-1} \mathbf{X}'_{ij'k} \Sigma_{ik} (\widehat{\boldsymbol{\Phi}}_{ik}^{j'})^{-1} \mathbf{Y}_{ij'k}$ Second, the conditional distributions detailed in (8) are based solely on $Data_{-i}$; they are considered an approximation to the respective posterior distribution given both $Data_{-i}$ and $Data_i^{2:(j-1)}$ that has computational advantages. To recognize the dependence of the estimates of the mean-related subject effects on the current visit we denote by $\widehat{\boldsymbol{\Theta}}_{ij}$ the posterior mean based on (8) and the post burn-in samples from the posterior distributions (8). The predicted responses in this fashion are denoted by $\widehat{\mathbf{Y}}_{ij} = \mathbf{X}_i \widehat{\boldsymbol{\Theta}}_{ij}$. Because this approach updates the prediction of $\boldsymbol{\Theta}_i$ after each visit we call it *dynamic parameter estimation*.

Static estimation

When there is no past information for the subject ($j = 2$), we need to use a different way to estimate the subject specific mean-related random effects. We use the same iterative approach that involves sampling from the conditional distributions (6) and (8) until convergence, except we only rely on the data $Data_{-i}$. Then we estimate $\boldsymbol{\Theta}_i$ by $\widehat{\boldsymbol{\Theta}}_i = \widehat{\Lambda}_{\boldsymbol{\Theta}}$. This approach is in fact a viable estimation approach, even if $j > 2$. In case the response is costly like ours and unavailable while the cheaper predictor images are available for a new subject, this approach can be used to obtain the predicted responses longitudinally. As it does not use the available subject's past information to estimate the mean-related random effects for the ith subject we call it *static parameter estimation*.

The calculations for the data analysis are based on generating 10,000 posterior samples for each of the parameters and removing the first 5,000 as burn-in. Among the remaining samples, we perform thinning by keeping one out of each 5 consecutive samples. Thus, the posterior inference about the main parameters of interest, the pointwise mean, standard deviation and the z-score calculated for the terms $\Lambda_{\boldsymbol{\Theta}}$, Σ_1 and Σ_2 is obtained based on the 1000 post-burn-in samples. The flowchart of the steps in inference and prediction are provided in Figure 3.

An advantage of the proposed Bayesian methodology is that it could be also used to gain insights about how the effect of the modalities and their temporal gradients vary across the spatial distance. More specifically, one could estimate the posterior mean and variance profiles of the smooth effects $\beta_{ikp}(d)$ and $\gamma_{ikp}(d)$ using the MCMC samples of the parameters $\Lambda_{\boldsymbol{\Theta}}$, Σ_1 and Σ_2. Pointwise posterior z-scores (i.e., the ratio of the posterior mean and standard deviation) are easily calculated and can be used for inference by contrasting them to the appropriate standard normal critical values. Pointwise inference is quite common when the parameters are smooth functions ([33]). More importantly the method allows to construct credible intervals for the prediction. For example, voxe-lwise normal-based prediction intervals are obtained as $\hat{Y}_{ijk}(\mathbf{v}) \pm z_{\alpha/2}\sqrt{V_{ijk}(\mathbf{v})}$, where $V_{ijk}(\mathbf{v}) = \mathbf{X}_{ijk}(\mathbf{v})'\widehat{\sum}_{1kk}\sum_2\mathbf{X}_{ijk}(\mathbf{v}) + \hat{\sigma}^2_{ik} + \hat{\tau}^2_{ik}$. Here $\hat{\sigma}^2_{ik}$ and $\hat{\tau}^2_{ik}$ are estimates of the subject-specific sill and nugget and are determined based on $\widehat{\boldsymbol{\Phi}}_{ik}$; for instance $\hat{\sigma}^2_{ik} = \exp\{\widehat{\boldsymbol{\Phi}}_{ik1}\}$ if $\log\{\sigma^2_{ik}\}$ is the first element of $\boldsymbol{\Phi}_{ik}$ and so on. Also $z_{\alpha/2}$ denotes the $(1 - \alpha/2)$-th quantile of standard normal density.

5 MRI Study of MS

5.1 Model fitting-related choices

In this section we apply the proposed methods to the MRI study of MS. Since the proposed model uses the relative change in the medical images over time, we discard the response at the baseline visit of the subjects and refer to the "baseline response" the response image corresponding to $j = 2$. Our objective is to predict the MTR image of a subject based on their current FLAIR and T1 images if this is their baseline hospital visit, or based on their past medical images - MTR, FLAIR and T1 - acquired at the previous hospital visits for $j > 2$. Let $Y_{ij}(\mathbf{v})$ be the MTR intensity at vovel $\mathbf{v}$ for the ith subject observed at jth hospital visit; we assume that $j \geq 2$. Also let $M_{ij1}(\mathbf{v})$ and $M_{ij2}(\mathbf{v})$ be the FLAIR and T1 intensities at the same voxel and corresponding to the ith subject and jth visit. As stated in Section 2 we restrict the analysis to the NAWM region that we further partition into a total of 32 anatomical sub-regions. For each such region $\mathbf{V}_{ik}$, the MTR intensity is related to the other two images and their gradients using the association described in (2). The model involves making some informed choices about the set of distances $\mathscr{D}$, the type of spatial correlation, and the basis for the smooth effects; we discuss these choices next.

Neighboring set—We use the commonly used coefficient of determination in regression analysis, R-square, to inform about the optimal choice of the neighboring set. Specifically, the best neighboring set results in a distance for which R-square plateaus. Figure 4 (second

panel) illustrates how the relative R^2 varies across different distances; the relative R^2 is calculated as $(R^2 - R_0^2)/(R_{full}^2 - R_0^2)$ where R_0^2 is the R-square corresponding to the model that does not consider any neighbor, with $\mathscr{D} = \{0\}$, and R_{full}^2 is the counterpart for a model that considers the neighboring set of the form $\{\mathbf{v} + \mathbf{u} : \mathbf{u} = (u_1, u_2, u_2); u_1, u_2, u_3 \in \{0, \pm 1, \pm 2\}\}$. We focus on this set, as considering larger neighboring sets increases the computation time and storage space requirements. All the R-square values are obtained under an independent and identically distributed (IID) error structure. The maximal distance corresponding to R_{full}^2 is $d = 2\sqrt{3}$mm while the distance is $d = 0$ for R_0^2; recall the set of possible distances is $\mathscr{D} = \{0, 1, \sqrt{2}, \sqrt{3}, 2, \sqrt{5}, \sqrt{6}, 2\sqrt{2}, 3, 2\sqrt{3}\}$. The relative R^2 profiles seem to stabilize for most of the partitions after a distance of 3 mm (1 voxel = 1× 1 × 1 mm^3).

Covariance structure—For modeling the residual covariance structure, we calculate the variograms based on sampling 1000 voxels randomly for each scan and finally average across subjects and scans corresponding to each partition. The plot of the averaged variograms is provided in the right panel of Figure 4. The averaged variograms indicate that a Matérn spatial correlation, as described in (3), is appropriate to describe the dependence of the residuals. Preliminary analysis - by fitting a Matérn covariance within each image - seems to indicate that the covariance function is smooth (i.e. $\nu_{ik} = \infty$); this allows to simplify the covariance model and assume a squared exponential covariance structure instead. The preliminary results indicate that the partial sill and nugget corresponding to the superior part of the brain are lower than those of the inferior part.

B-spline selection—We use a B-spline basis for modeling the smooth functions $\beta_{ikp}(\cdot)$ and $\gamma_{ikp}(\cdot)$. The set $\mathscr{D}$ includes 10 elements and thus we consider only four linear splines with equally-spaced knots. This choice results in reducing the dimensionality of the parameters and thus has important computational advantages.

5.2 Dynamic estimation versus static estimation

The model fitting is done as described in the previous section. The estimates of the covariance-related subject specific parameters $\mathbf{\Phi}_{ik}$ are obtained using a composite likelihood approach using blocks of 100 voxels per partition. We estimate the mean-related subject specific parameters $\mathbf{\Theta}_{ik}$ using the two approaches: dynamic and static.

To better understand the difference between the two methods, consider the subject-level predictions at few hospital visits, for the subject illustrated in Figure 2. For each visit and slice considered in Figure 2, the MTR image is predicted using $\widehat{\mathbf{\Theta}}_{ik}$ estimated dynamically (Figure 5) as well as statically (Figure 6). In both the cases, we provide the posterior predictive mean, prediction error, i.e. the voxel-wise difference between the posterior predictive mean and the true MTR profile, and the voxel-wise posterior standard deviation for each of the four visits considered in Figure 2. The prediction for the second visit (Day 28) using static and dynamic approaches are same by model specification. Since the dynamic estimation uses the past history to update parameter estimation, it is expected that it would show improved performance for the later visits prediction and the two figures visually

confirm this tendency. Moreover, Table 1 presents the root mean squared error (RMSE) based on the difference between the predicted response and the observed response, separately for the two approaches. It shows that, as there is sufficient imaging history for the subject, their response predictions are getting more accurate (lower mean and standard deviation) with the dynamic approach that with the static one. The sites with maximum posterior standard deviation correspond to the labels 31 and 32 representing Tapatum right (TAP-R) and Tapatum left (TAP-L) regions of white matter respectively.

To assess the performance of the proposed model and compare it with simpler alternatives, we take a step back and examine its four key parts. First, the mean model (2) includes the inhomogeneity correction term $h(\mathbf{v})$ (C). Second, the proposed mean model accounts for spatial association (S) between the response intensity at a voxel and image covariates intensity at neighboring voxels through terms like $\beta(\cdot)$'s. Third, the model (2) uses as predictors the relative change in the image covariates, thus accounting for a temporal association (T). Fourth, the model (2) assumes the effect of each predictor varies across the brain tissue and considers partitions of the NAWM into anatomical regions or blocks (B). In this section, we evaluate the importance of each of these main components of the mean model via the predictive likelihood, based on the assumed model distribution.

For each i at scan $j \geq 2$, we use the data obtained from the rest of the subjects, $Data_{-i}$, to estimate the model components in static estimation and $Data_{-i}$ along with $Data_i^{2:(j-1)}$, the historical information from subject i, to estimate the model components in dynamic estimation. We then calculate the predictive mean log-likelihood ($logPL_{ij}$) by

$$logPL_{ij} = \sum_{k=1}^{K} \log f(\mathbf{Y}_{ijk}; \widehat{\boldsymbol{\Theta}}_{ik}, \widehat{\boldsymbol{\Phi}}_{ik})/|\mathbf{V}_i|,$$

where $f(Y; \boldsymbol{\Theta}, \boldsymbol{\Phi})$ is the multivariate normal density with the mean and covariance matrix described in Section 3, $\widehat{\boldsymbol{\Theta}}_{ik}$, $\widehat{\boldsymbol{\Phi}}_{ik}$ are the subject/region-specific components predicted using a dynamic or static approach and $|\mathbf{V}_i|$ is the number of the NAWM voxels of the ith subject. Then we consider the average of the predictive mean log-likelihood across the subjects by $logPL_j^{(avg)} = \sum_i^I logPL_{ij}/I$. When the dependence of the residual error is not trivial the computations are very heavy. Thus, for simplicity we first carry out the comparisons under a working assumption of independent residual error structure and use the proposed dependence structure (squared exponential) only for the final model that includes all the components.

The results for visits $j = 2, 3, \ldots, 11$ are depicted in Figure 7, separately for dynamic and static estimation of the mean-related subject specific effects $\widehat{\boldsymbol{\Theta}}_{ik}$'s. The base model, denoted by 0, involves only the intercept. The results indicate a clear advantage of the dynamic estimation of the subject-specific mean-related effects especially for the later visits, irrespective of the assumption made on the residual error dependence. Under the a working independence assumption for the errors, it appears that accounting for the inhomogeneity correction (C), the spatial covariates (S) and the partitioning into blocks (B) lead to the

larger predictive likelihood. When using all the terms (STBC) and accounting for the spatial dependence of the residuals using a squared exponential (SE) correlation, the predictive likelihood is considerably larger for both static and dynamic approaches. This finding is not surprising, as our exploratory analysis indicated voxel dependence of the residuals. The averaged variograms in the right panel of Figure 4 show that there is a strong spatial correlation among the residuals at nearby voxels. For the dynamic approach, the $logPL_j^{(avg)}$ values increase with the number of visits, confirming that using historical information improves prediction.

Finally, we assess the prediction performance of the proposed methods using the root mean squared error (RMSE) based on all the subjects. Using the same set up as before, for each subject we estimate the subject-specific parameters based on the remaining subjects data and the available subject information, if the case. RMSE is calculated for each visit, in a similar fashion as before, except now is averaged over the subjects observed at each visit j for $j = 2, 3, \ldots, 11$. Results are presented in Table 2: mean and standard deviation are shown separately for each visit and for both dynamic and static approaches. The result support the finding that the dynamic estimation improves prediction accuracy once the subject have sufficient medical imaging history.

5.3 Overall associations between T1, FLAIR, and MTR

In this section we analyze the results obtained with a dynamic estimation approach. The key components of the mean model (2) are the smooth subject-specific effects and we focus on them here, but in a population-level sense. Figure 8 and Figure 9 illustrate the posterior mean estimates of the smooth effects of the various modalities and their gradients and provide insights into their significance.

First consider the average effect of T1 onto the mean MTR at a current voxel. Figure 8 top left panel shows that the mean response is positively correlated to the T1 intensity at the current voxel, but it is negatively associated at other neighboring voxels. Not surprisingly, the magnitude of this association decreases as the distance between the voxels increases. The top middle panel displays the posterior standard deviation profile across partitions and the top rightmost panel gives the pointwise ratio between the posterior mean and the posterior standard deviation - what is also referred to as z-score profile - for the effect of the T1 intensity; this graph is helpful in assessing the significance of the effect. It shows that for the current voxel and for voxels that are further apart, the effect is not negligible. This observation is consistent across all the anatomical regions of the brain, though the magnitude of the effect varies slightly. The bottom panels of Figure 8 focus on the effect of the rate of change in T1 on the current mean MTR intensity. On average, it seems that the effect is larger in magnitude at the current voxel and decays for far away voxels though the actual tendency varies considerably across the brain regions. When examining the z-score profiles we see that the effect is significant for some anatomical regions: highly significant in Posterior limb of internal capsule right (PLIC-R) and Posterior limb of internal capsule left (PLIC-L) and other few significant cases correspond to Anterior limb of internal capsule right (ALIC-R), Anterior limb of internal capsule left(ALIC-L), Superior fronto-occipital fasciculus right (SFO-R), Superior fronto-occipital fasciculus left (SFO-L) etc. For some

brain regions, it seems that the effect of the rate of change in T1 is non-negligible for farther apart voxels (voxel distance of 2mm in some cases), e.g. Superior longitudinal fasciculus right (SLF-R), Superior longitudinal fasciculus left (SLF-L), Cingulum (cingulate gyrus) right (CGC-R), Cingulum (cingulate gyrus) left (CGC-L) etc.

Figure 9 shows the average effect of FLAIR and of its rate of change on the MTR intensity at a voxel. The effect is very similar across the brain. FLAIR has a negative effect at the current voxel indicating that a positive value on the FLAIR image at the voxel is realted to a negative value for the MTR intensity at the same voxel. The effect decays at a constant rate as the distance between voxels increases; the distance at which the effect becomes negligible seems to vary across the brain regions. The bottom panels show a positive association with the temporal rate change in the T1 intensity at the current voxels and close neighboring voxels. Overall these results imply that while a positive intensity in FLAIR is associated with a negative MTR value on average at the current voxels, an increase in the FLAIR intensity is associated with an increase in the MTR intensity while a decrease in the FLAIR intensity is associated with a decrease in the average MTR intensity. This tendancy is valid across the various anatomical brains and it seems to reduce as the distance between the voxels increases.

6 Simulation study

We further investigate the proposed methodology in a numerical simulation. Specifically, we consider a setting mimicking the data application and assess the frequentist properties of the proposed approximate Bayesian methodology.

Due to the large dimensionality of the data, working with the original image covariates is prohibitive in a simulation study. Nevertheless we would like to use a setting that matches as much as possible the data application setting. In designing the simulation setting, the most important factors are: the number of subjects, the number of repeated hospital visits, the structure of the image covariates, and the use of partitions. To address this challenge we consider a reduced number of hospital visits per subject, according to the rule: the number of hospital visits per subject is capped at 10. For each scan, only 10 anatomical regions are used, and in each anatomical region the selected set of voxels is an arbitrarily chosen subset of size capped at 1000 from the original set of voxels. Once the set of covariates is determined, the response intensity per subject and visit is determined using the model (2), (3), and (5).

The parameters $\Lambda_{\boldsymbol{\Theta}}$, Σ_1, Σ_2 are set to the values estimated from the full MRI dataset analysis. The values for the covariance-related parameters, Λ_{Φ}, Ω_1 and Ω_2, are also set to their estimates from the data analysis. Specifically, $\Lambda_{\boldsymbol{\Phi}}$ is equal to $\widehat{\Lambda}_{\boldsymbol{\Phi}} = \frac{1}{I}\sum_{i=1}^{I} \widehat{\boldsymbol{\Phi}}_i$. Using the *second stage* model specification (5) and $\widehat{\boldsymbol{\Phi}}_i$'s we compute a pseudo likelihood function, $p\ell(\Omega_1, \Omega_2) = \ell(\Omega_1, \Omega_2, \Lambda_{\boldsymbol{\Theta}}; \widehat{\boldsymbol{\Phi}}_i)$, where $\ell(\cdot; \widehat{\boldsymbol{\Phi}}_i)$ is the log-likelihood corresponding to (5). We estimate Ω_1, Ω_2 by the maximizers of $p\ell(\Omega_1, \Omega_2)$.

For each simulated data set, we fit our method and calculate pointwise 95% credible regions of the population means $\beta_{kp}(\cdot)$ and $\gamma_{kp}(\cdot)$ based on the MCMC samples from Λ_{Θ}. Then we determine the pointwise coverage as the average of the points at which the true population mean is included in the corresponding credible interval. In a similar way, we construct pointwise 95% credible regions of the population standard deviations of $\beta_{ikp}(\cdot)$'s and $\gamma_{ikp}(\cdot)$'s based on the MCMC samples from Σ_1 and Σ_2. The results based on 100 simulated data sets are presented in Tables 3 and 4.

The actual coverages are close to 95% for the effect of both T1 and T1-gradient, while they fall slightly short corresponding to the effects of each FLAIR and FLAIR-gradient. In case of Table 4, the coverage of the intervals for all the effects is less than the nominal level. We observe similar pattern for the pair of T1 and T1-gradient and that of FLAIR and FLAIR-gradient as in Table 3. It appears that plugging-in the MLEs of the covariance parameters instead of sampling from them using Metropolis-Hastings algorithm affects the standard deviation profiles more than the mean profiles. Considering the huge computational challenges involved in implementing the Metropolis-Hastings steps for high dimensional datasets, the simulation study confirms numerically that the proposed approximation approach performs reasonably well.

7 Final remarks

We propose a population-level two-stage spatio-temporal regression model for longitudinal imaging studies, where each subject is imaged at many hospital visits and the images are co-registered but are not registered across subjects. The proposed method accounts for past subject imaging information, spatial association, effects that differ across the brain, and spatial dependence of the residual errors. We discuss a Bayesian approach to fit the model that has important inferential advantages. Our methodology consists of two stages. In the first stage, we extract maximum likelihood estimates separately from each subject. The fitted model allows for pooling information across nearby voxels to construct predictors and different effects in different anatomical regions. We also account for residual spatial dependence using a composite likelihood method. In the second stage, we combine the results across subjects to estimate the population distribution, which is then used as a prior distribution in a dynamic model to make predictions for a new patient. Our simulation study shows that the resulting statistical inference is well calibrated, and when applied to a database of imaging in subjects with MS we show the proposed model outperforms several simpler alternatives.

The proposed modeling approach has a few drawbacks. In case of large number of image covariates, implementation of our method could be computationally challenging even after the approximation techniques we have discussed. The mean model discussed here assumes an additive and linear relationship. The linearity assumption can be dropped by considering polynomials of the predictors or a penalized regression splines approach following [25]. For the inhomogeneity correction, we consider the spatial coordinates as predictors keeping the computational burden in mind, but higher-order polynomials of the spatial coordinates could be more appropriate. Within the covariance structure, we keep same covariance parameters corresponding to all voxels within a partition but a partition may not remove the non-

stationarity fully. Also, we consider the smaller blocks within the partitions to be of small size which makes the cross-block correlation zero which may not be accurate.

The proposed modeling is used to predict advanced and less widely available MTR imaging based on the conventional T1 and FLAIR modalities. As acquiring MTR is not standard practice due to the increased acquisition time, our model could be used for prediction of MTR based on the spatio-temporal image predictors. The correlation between MTR and the conventional modalities is low, and thus a flexible spatio-temporal random effect model is a better choice than a simple cross-sectional model and our model serves this purpose. Additionally, MTR is a significantly more noisy image and the predicted values based on our model are less noisy. We notice that the average log-predictive likelihood increases on average by 44.04% in the case of dynamic updating and by 52.08% in the case of static approach to the simplest cross-sectional model with subject-varying random effects. Based on our case study, the coverage probabilities of the credible regions corresponding to the posterior predictive distributions are quite high which indicates that our method serves well in predicting MTR and hence, facilitating additional future studies of MS.

Acknowledgments

The project described was supported in part by RO1NS085211, R21NS093349 and RO1MH 086633 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. This research was partially supported by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke. The authors would like to acknowledge useful suggestions of Joseph Guinness and two anonymous reviewers which helped to improve the paper in several ways.

References

1. Ashburner J, Friston KJ. Voxel-based morphometrythe methods. Neuroimage. 11(6):805–821.2000; [PubMed: 10860804]
2. Avants BB, Kandel BM, Duda JT, Cook PA, Tustison NJ, KL S. ANTsR: ANTs in R: quantification tools for biomedical images. 2016
3. Bilgel M, Prince JL, Wong DF, Resnick SM, Jedynak BM. A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. Neuroimage. 134:658–670.2016; [PubMed: 27095307]
4. Brex PA, Ciccarelli O, O'Riordan JI, Sailer M, Thompson AJ, Miller DH. A longitudinal study of abnormalities on mri and disability from multiple sclerosis. New England Journal of Medicine. 346(3):158–164.2002; [PubMed: 11796849]
5. Carass, A, Wheeler, MB, Cuzzocreo, J, Bazin, PL, Bassett, SS, Prince, JL. Biomedical Imaging: From Nano to Macro, 2007 ISBI 2007 4th IEEE International Symposium on Biomedical Imaging. IEEE; 2007. A joint registration and segmentation approach to skull stripping; 656–659.
6. Casella G. An introduction to empirical bayes data analysis. The American Statistician. 39(2):83–87.1985;
7. Catalaa I, Grossman RI, Kolson DL, Udupa JK, Nyul LG, Wei L, Zhang X, Polansky M, Mannon LJ, McGowan JC. Multiple sclerosis: Magnetization transfer histogram analysis of segmented normal-appearing white matter 1. Radiology. 216(2):351–355.2000; [PubMed: 10924552]
8. Chen Y, Wang X, Kong L, Zhu H. Local region sparse learning for image-on-scalar regression. arXiv preprint arXiv:1605.08501. 2016
9. Dworkin JD, Sweeney EM, Schindler MK, Chahin S, Reich DS, Shinohara RT. Prevail: Predicting recovery through estimation and visualization of active and incident lesions. NeuroImage: Clinical. 12:293–299.2016; [PubMed: 27551666]

10. Eidsvik J, Shaby BA, Reich BJ, Wheeler M, Niemi J. Estimation and prediction in spatial models with block composite likelihoods. Journal of Computational and Graphical Statistics. 23(2):295–315.2014;
11. Eloyan A, Shou H, Shinohara RT, Sweeney EM, Nebel MB, Cuzzocreo JL, Calabresi PA, Reich DS, Lindquist MA, Crainiceanu CM. Health effects of lesion localization in multiple sclerosis: spatial registration and confounding adjustment. PloS one. 9(9):e107263.2014; [PubMed: 25233361]
12. Fonov VS, Evans AC, McKinstry RC, Almli C, Collins D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage. 47:S102.2009;
13. Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RS. Statistical parametric maps in functional imaging: a general linear approach. Human brain mapping. 2(4):189–210.1994;
14. Gass A, Barker G, Kidd D, Thorpe J, MacManus D, Brennan A, Tofts P, Thompson A, Mc-Donald W, Miller D. Correlation of magnetization transfer ratio with clinical disability in multiple sclerosis. Annals of Neurology. 36(1):62–67.1994; [PubMed: 8024264]
15. Goldsmith J, Huang L, Crainiceanu CM. Smooth scalar-on-image regression via spatial bayesian variable selection. Journal of Computational and Graphical Statistics. 23(1):46–64.2014; [PubMed: 24729670]
16. Griffin C, Parker GJ, Barker G, Thompson A, Miller D. Mtr and t1 provide complementary information in ms nawm, but not in lesions. Multiple Sclerosis Journal. 6(5):327–331.2000; [PubMed: 11064442]
17. Grossman RI. Magnetization transfer in multiple sclerosis. Annals of Neurology. 36(S1):S97–S99.1994; [PubMed: 8017897]
18. Hoff P. Separable covariance arrays via the Tucker product, with applications to multivariate relational data. Bayesian Analysis. 6:179–196.2011;
19. Jog A, Carass A, Roy S, Pham DL, Prince JL. Mr image synthesis by contrast learning on neighborhood ensembles. Medical image analysis. 24(1):63–76.2015; [PubMed: 26072167]
20. Jog A, Carass A, Roy S, Pham DL, Prince JL. Random forest regression for magnetic resonance image synthesis. Medical image analysis. 35:475–488.2017; [PubMed: 27607469]
21. Kang J, Reich BJ, Staicu AM. Scalar-on-image regression via the soft-thresholded gaussian process. arXiv preprint arXiv:1604.03192. 2016
22. Kappos L, Moeri D, Radue EW, Schoetzau A, Schweikert K, Barkhof F, Miller D, Guttmann CR, Weiner HL, Gasperini C, et al. Predictive value of gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a meta-analysis. The Lancet. 353(9157):964–969.1999;
23. Lai SH, Fang M. A new variational shape-from-orientation approach to correcting intensity inhomogeneities in magnetic resonance images. Medical Image Analysis. 3(4):409–424.1999; [PubMed: 10709704]
24. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, McFarland HF, Paty DW, Polman CH, Reingold SC, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. Annals of Neurology. 50(1):121–127.2001; [PubMed: 11456302]
25. Mejia AF, Sweeney EM, Dewey B, Nair G, Sati P, Shea C, Reich DS, Shinohara RT. Statistical estimation of t1 relaxation times using conventional magnetic resonance imaging. NeuroImage. 133:176–188.2016; [PubMed: 26732403]
26. Moll NM, Rietsch AM, Thomas S, Ransohoff AJ, Lee JC, Fox R, Chang A, Ransohoff RM, Fisher E. Multiple sclerosis normal-appearing white matter: Pathology–imaging correlations. Annals of neurology. 70(5):764–773.2011; [PubMed: 22162059]
27. Mori S, Wakana S, Van Zijl PC, Nagae-Poetscher L. MRI atlas of human white matter Elsevier. 2005
28. Morris JS, Baladandayuthapani V, Herrick RC, Sanna P, Gutstein H. Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data. The Annals of Applied Statistics. 5(2A):894.2011; [PubMed: 22408711]
29. Muschelli J. extrantsr: Extra Functions to Build on the ANTsR Package R package version 2.8.

30. Polman CH, Reingold SC, Edan G, Filippi M, Hartung H-P, Kappos L, Lublin FD, Metz LM, McFarland HF, O'Connor PW, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the mcdonald criteria. Annals of Neurology. 58(6):840–846.2005; [PubMed: 16283615]
31. Pomann GM, Staicu AM, Lobaton EJ, Mejia AF, Dewey BE, Reich DS, Sweeney EM, Shinohara RT, et al. A lag functional linear model for prediction of magnetization transfer ratio in multiple sclerosis lesions. The Annals of Applied Statistics. 10(4):2325–2348.2017;
32. Pomann GM, Sweeney EM, Reich DS, Staicu AM, Shinohara RT. Scan-stratified case-control sampling for modeling blood–brain barrier integrity in multiple sclerosis. Statistics in Medicine. 34(20):2872–2880.2015; [PubMed: 25939401]
33. Ramsay JO. Functional data analysis Wiley Online Library. 2006
34. Ramsay JO, Silverman BW. Applied functional data analysis: methods and case studies. 77
35. Reiss PT, Huo L, Zhao Y, Kelly C, Ogden RT, et al. Wavelet-domain regression and predictive inference in psychiatric neuroimaging. The Annals of Applied Statistics. 9(2):1076–1101.2015; [PubMed: 27330652]
36. Roy S, Carass A, Prince JL. Magnetic resonance image example-based contrast synthesis. IEEE transactions on medical imaging. 32(12):2348–2363.2013; [PubMed: 24058022]
37. Schmierer K, Scaravilli F, Altmann DR, Barker GJ, Miller DH. Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. Annals of Neurology. 56(3):407–415.2004; [PubMed: 15349868]
38. Shi R, Kang J. Thresholded multiscale gaussian processes with application to bayesian feature selection for massive neuroimaging data. arXiv preprint arXiv:1504.06074. 2015
39. Shiee N, Bazin PL, Ozturk A, Reich DS, Calabresi PA, Pham DL. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. NeuroImage. 49(2):1524–1535.2010; [PubMed: 19766196]
40. Shinohara RT, Crainiceanu CM, Caffo BS, Gaitán MI, Reich DS. Population-wide principal component-based quantification of blood–brain-barrier dynamics in multiple sclerosis. NeuroImage. 57(4):1430–1446.2011; [PubMed: 21635955]
41. Shinohara RT, Goldsmith J, Mateen F, Crainiceanu C, Reich DS. Predicting breakdown of the blood-brain barrier in multiple sclerosis without contrast agents. American Journal of Neuroradiology. 33(8):1586–1590.2012; [PubMed: 22442041]
42. Shinohara RT, Sweeney EM, Goldsmith J, Shiee N, Mateen FJ, Calabresi PA, Jarso S, Pham DL, Reich DS, Crainiceanu CM, et al. Statistical normalization techniques for magnetic resonance imaging. NeuroImage: Clinical. 6:9–19.2014; [PubMed: 25379412]
43. Sled, JG, Zijdenbos, AP, Evans, AC. Biennial International Conference on Information Processing in Medical Imaging. Springer; 1997. A comparison of retrospective intensity non-uniformity correction methods for mri; 459–464.
44. Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage. 31(4):1487–1505.2006; [PubMed: 16624579]
45. Stein, M. Statistical Interpolation of Spatial Data. Springer; New York: 1999.
46. Suttner LH, Mejia A, Dewey B, Sati P, Reich DS, Shinohara RT. Statistical estimation of white matter microstructure from conventional mri. NeuroImage: Clinical. 12:615–623.2016; [PubMed: 27722085]
47. Sweeney E, Shinohara R, Shea C, Reich D, Crainiceanu C. Automatic lesion incidence estimation and detection in multiple sclerosis using multisequence longitudinal mri. American Journal of Neuroradiology. 34(1):68–73.2013; [PubMed: 22766673]
48. Sweeney EM, Shinohara RT, Dewey BE, Schindler MK, Muschelli J, Reich DS, Crainiceanu CM, Eloyan A. Relating multi-sequence longitudinal intensity profiles and clinical covariates in incident multiple sclerosis lesions. NeuroImage: Clinical. 10:1–17.2016; [PubMed: 26693397]
49. Symms M, Jäger H, Schmierer K, Yousry T. A review of structural magnetic resonance neuroimaging. Journal of Neurology, Neurosurgery & Psychiatry. 75(9):1235–1244.2004;
50. Zhu H, Fan J, Kong L. Spatially varying coefficient model for neuroimaging data with jump discontinuities. Journal of the American Statistical Association. 109(507):1084–1098.2014; [PubMed: 25435598]

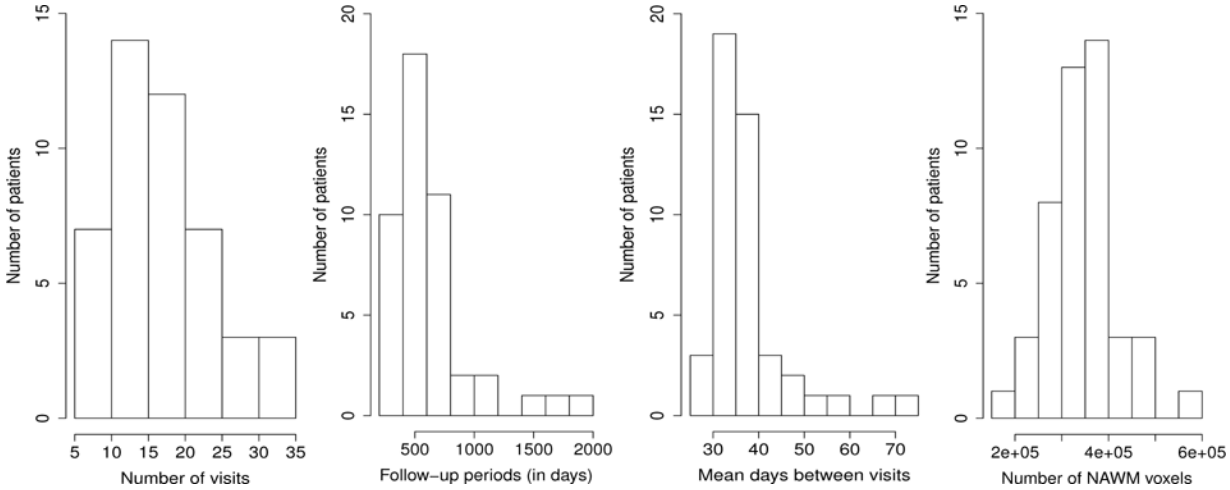

Figure 1.

The histograms of the number of visits (on the first panel), the follow-up period in days (on the second panel), the mean days between two visits (on the third panel) and the number of NAWM voxels (on the fourth panel) for each subject.

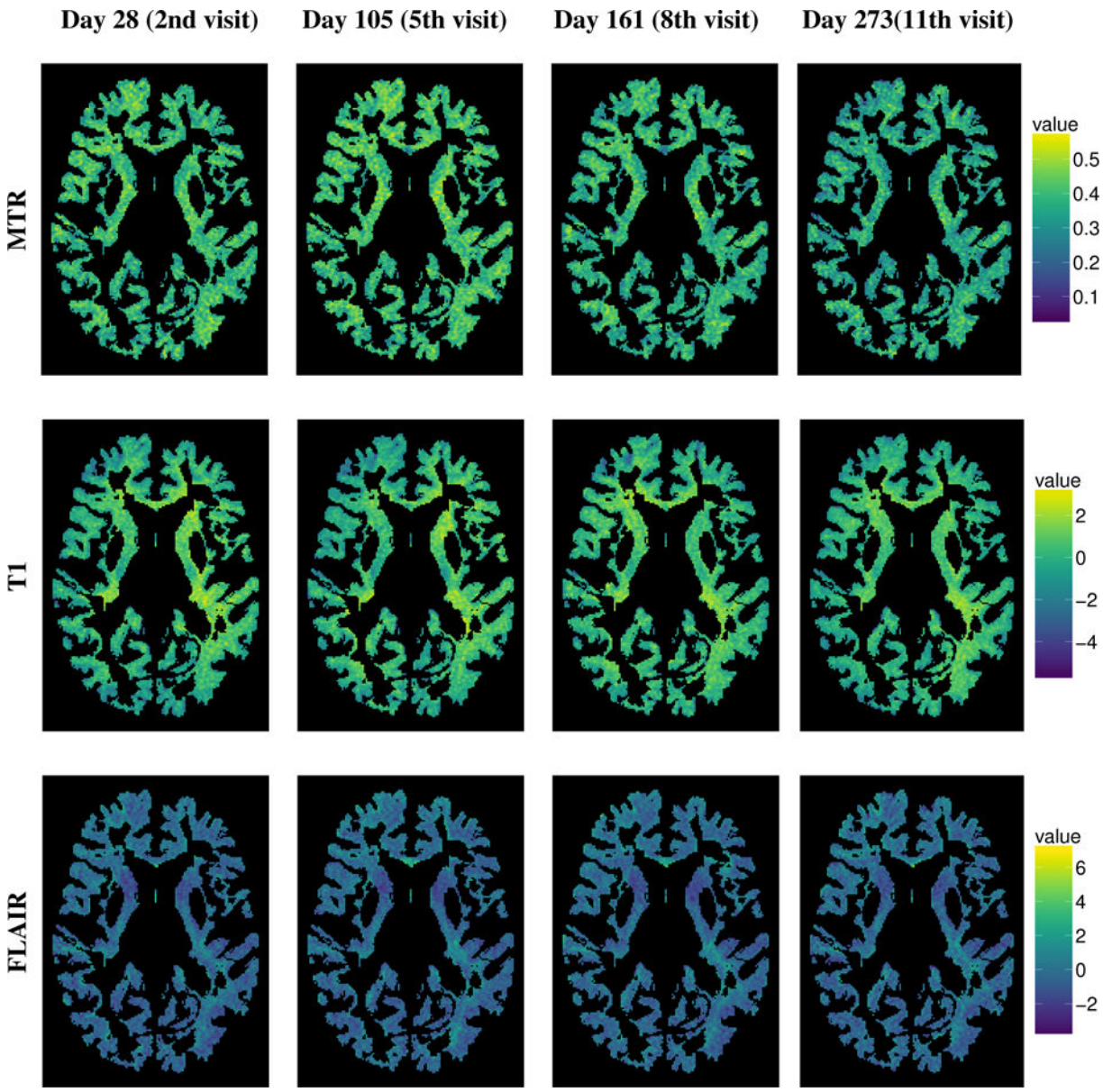

Figure 2.

MTR (first row), normalized T1 (middle row) and normalized FLAIR (last row) profiles for a representative subject on 28-th day, 105-th day, 161-th day and 273-th day for the Z-slice $Z = 84$.

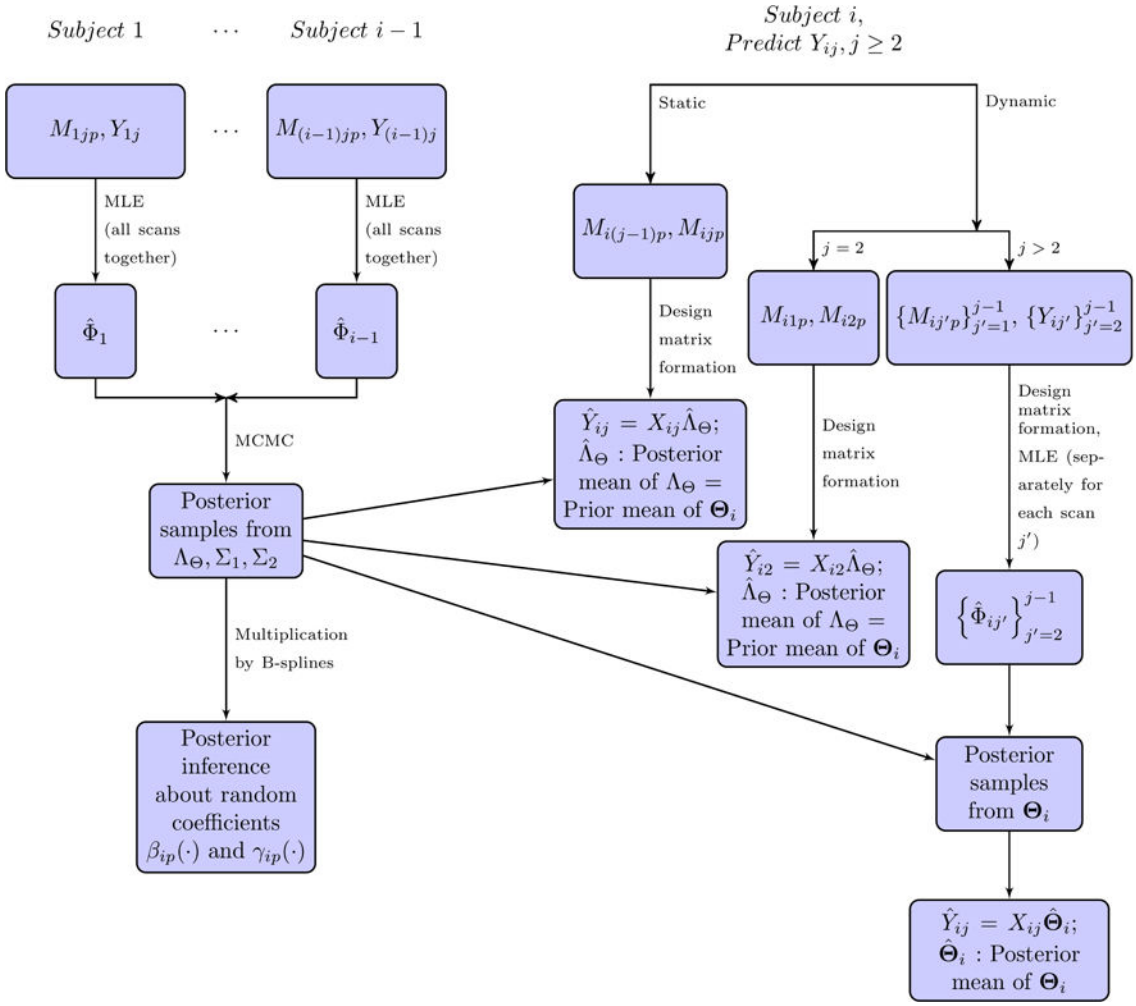

Figure 3.
Flowchart of the computational details.

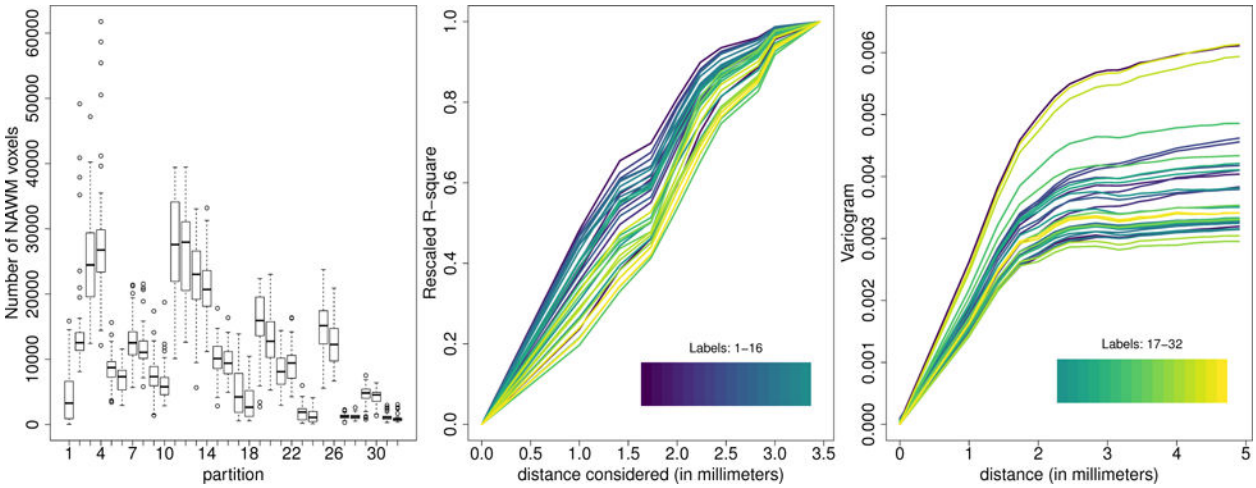

Figure 4.

The boxplot of the partition-wise dimensions of the white matter voxels across the subjects (first panel), the plot of spatial distance of the farthest neighbor considered versus the corresponding rescaled R^2 values across the partitions (second panel) and the variograms of the residuals after fitting 1 but with IID errors for all the partitions (third panel). The last two panels are averaged across subjects and scans. The colors corresponding to the 32 anatomical regions are represented by the color bars in the last two panels.

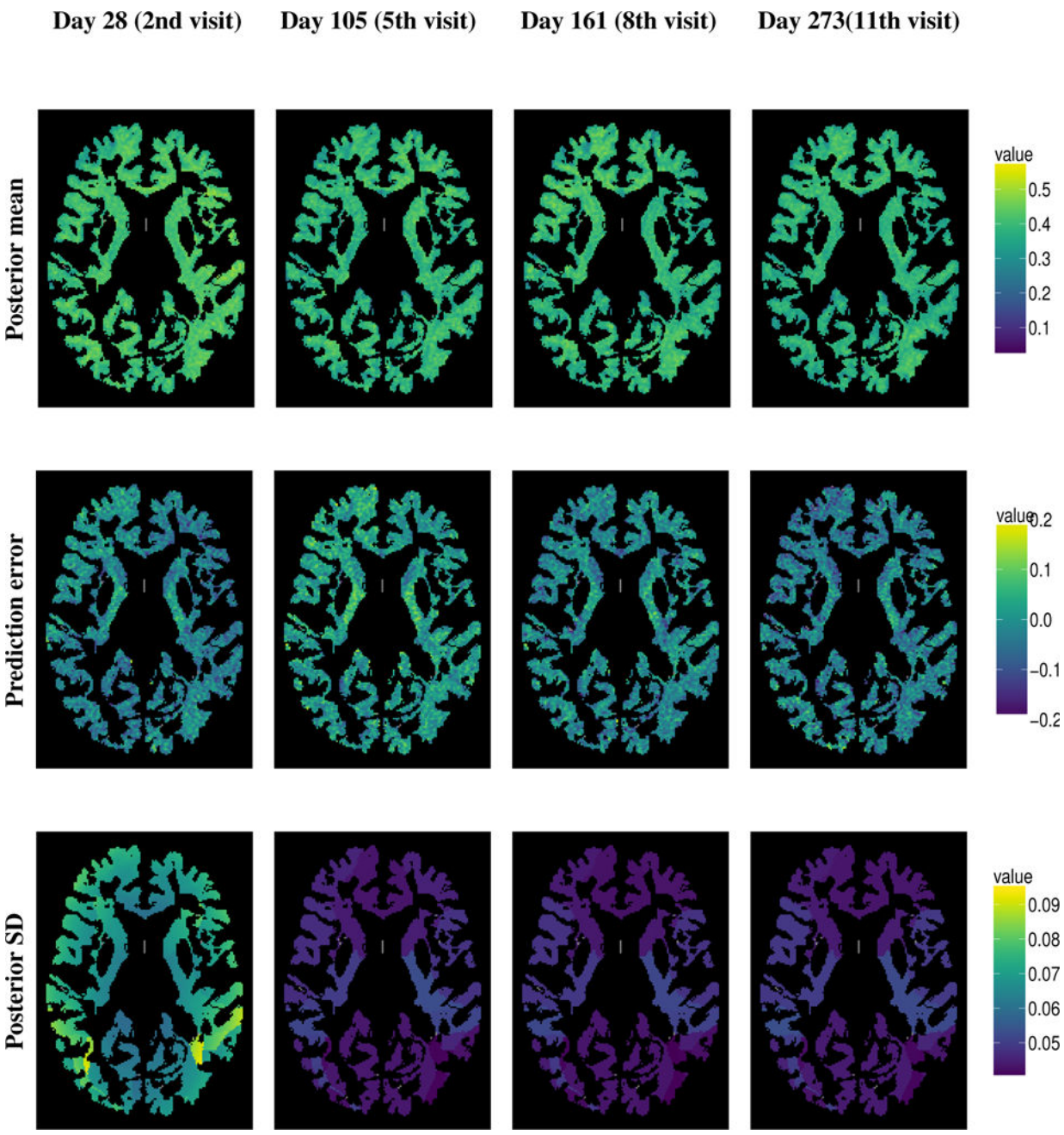

Figure 5.

Posterior predictive mean, prediction error and posterior SD of the MTR profiles for the same representative subject as in Figure 2 on 28-th day, 105-th day, 161-th day and 273-th day for the Z-slice $Z = 84$ in case of dynamic prediction.

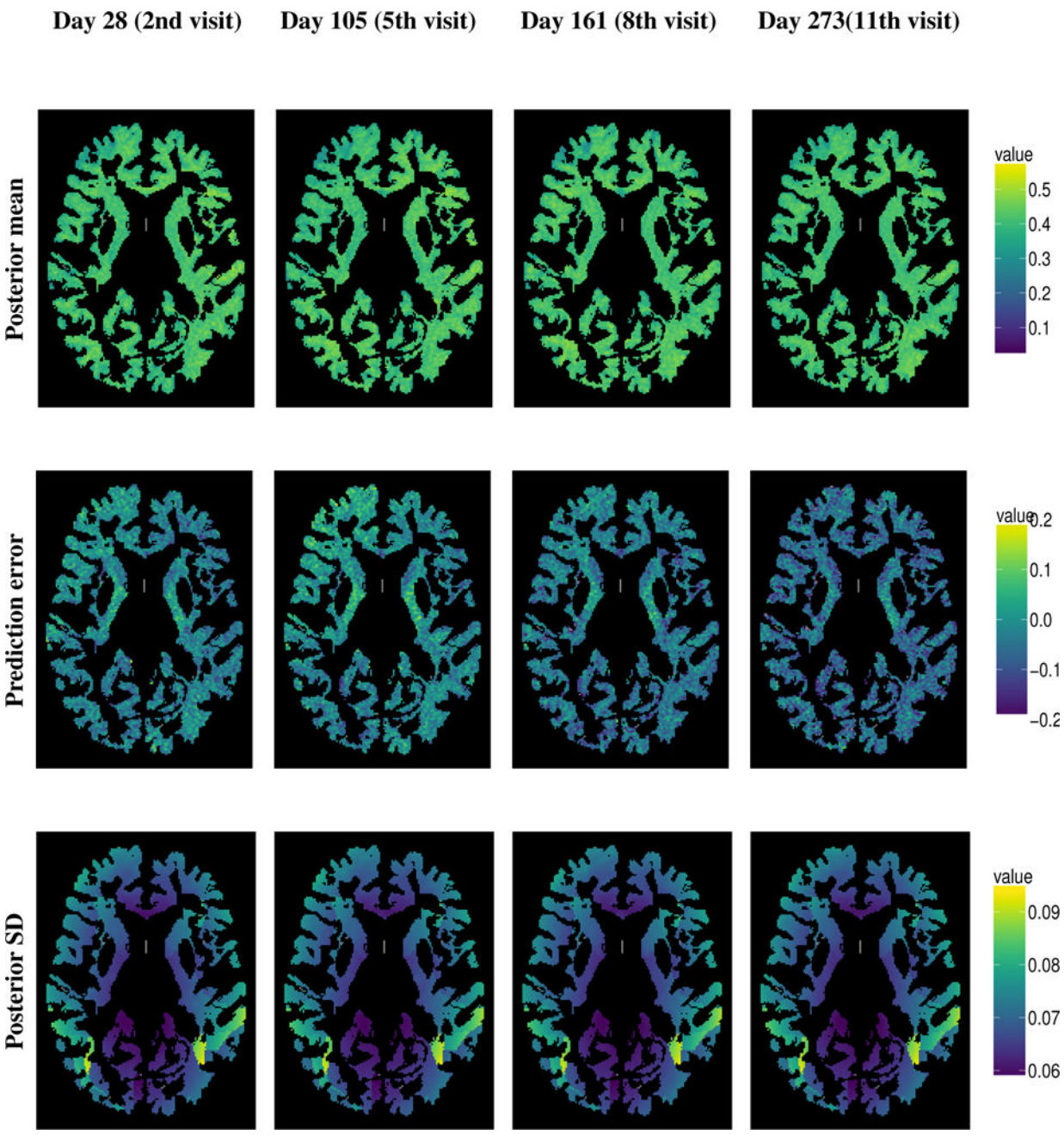

Figure 6.

Posterior predictive mean, prediction error and posterior SD of the MTR profiles for the same representative subject as in Figure 2 on 28-th day, 105-th day, 161-th day and 273-th day for the Z-slice $Z = 84$ in case of static prediction.

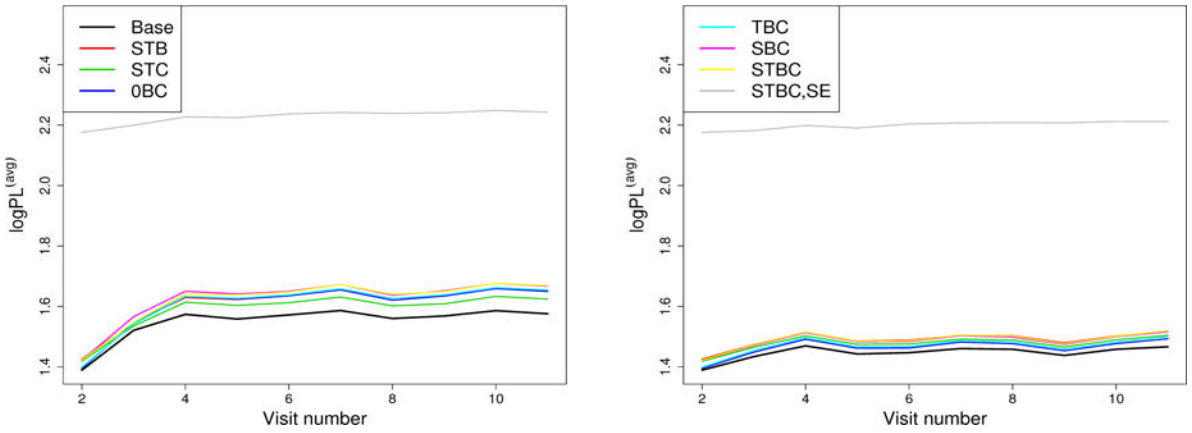

Figure 7.

$logPL_j^{(avg)}$ values for the first 10 scans in case of dynamic updating (left) and static (right).

We use the following notations: 0- base model, S- spatial predictors, T - temporal predictors, C - inhomogeneity correction, B - partitioning the white matter into 32 labels, SE - squared exponential covariance.

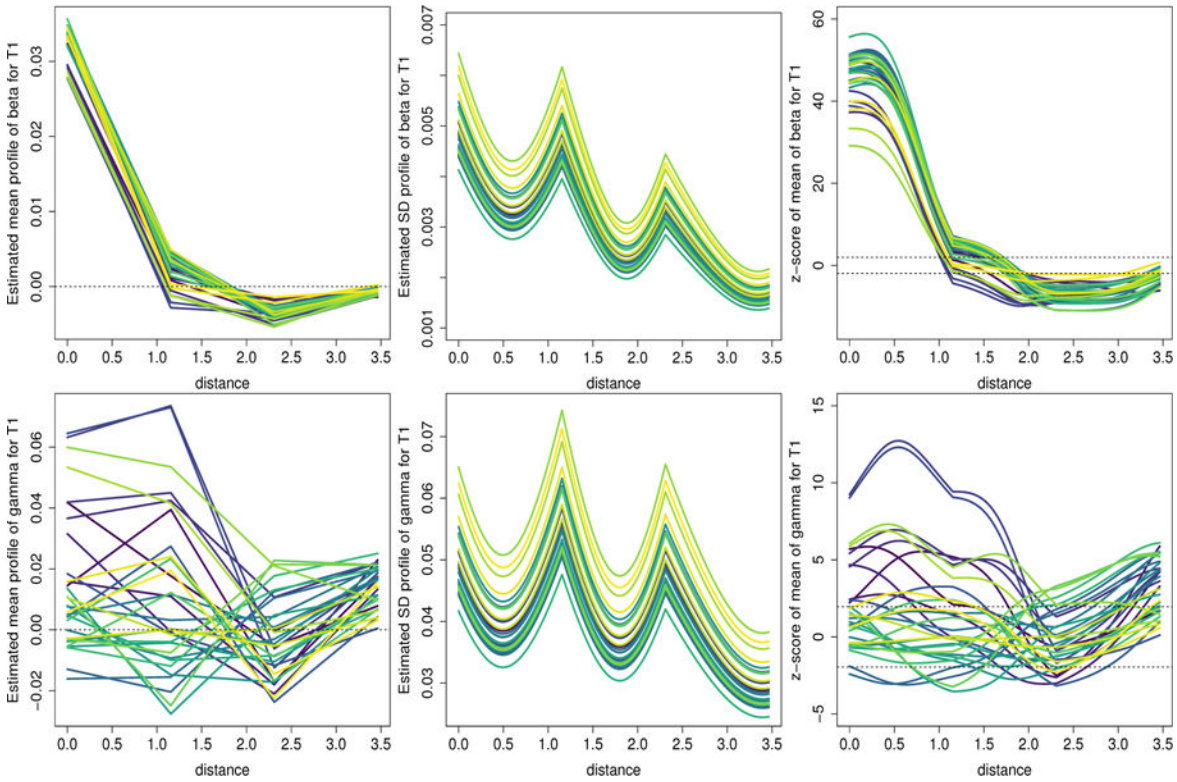

Figure 8.

Pointwise posterior means of the population means and population standard deviations of $\beta_{ikp}(\cdot)$ profiles and the z-score profiles corresponding to T1 (top panel). Similarly, the pointwise posterior means of the population means and population standard deviations of $\gamma_{ikp}(\cdot)$ and the z-score profiles corresponding to T1-gradient (bottom panel). The horizontal lines along with the z-scores denote the 0.025-th and 0.975-th quantiles of standard normal density. The colors for the labels are as in Figure 4.

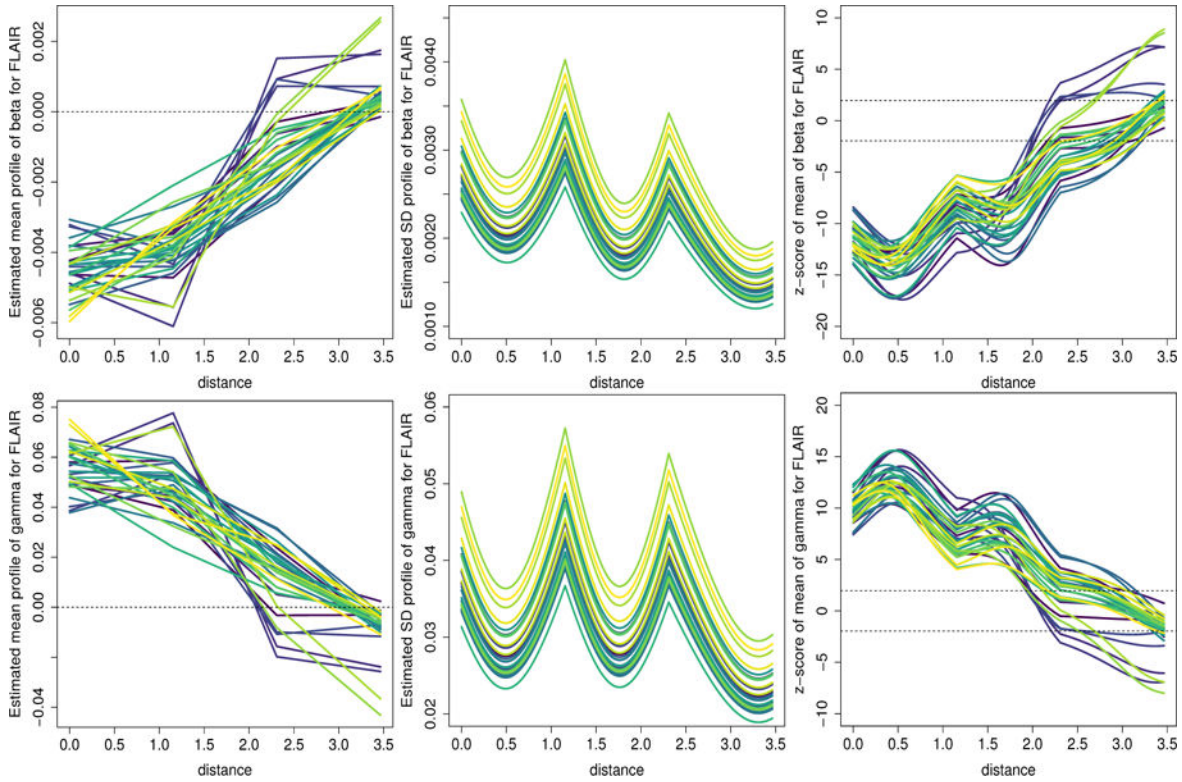

Figure 9.

Pointwise posterior means of the population means and population standard deviations of $\beta_{ikp}(\cdot)$ profiles and the z-score profiles corresponding to FLAIR (top panel). Similarly, the pointwise posterior means of the population means and population standard deviations of $\gamma_{ikp}(\cdot)$ and the z-score profiles corresponding to FLAIR-gradient (bottom panel). The horizontal lines along with the z-scores denote the 0.025-th and 0.975-th quantiles of standard normal density. The colors for the labels are as in Figure 4.

Table 1

Average RMSE the corresponding average posterior standard deviation values (within bracket) for the same representative subject as in Figure 2 on Day 28 (2nd visit), Day 105 (5th visit), Day 161 (8th visit) and Day 273 (11th visit) for the Z-slice $Z = 84$.

	2nd visit	5th visit	8th visit	11th visit
Dynamic	0.0512(0.0702)	0.0487(0.0461)	0.0485(0.0457)	0.0585(0.0464)
Static	0.0512(0.0702)	0.0484(0.0702)	0.0626(0.0702)	0.0795(0.0701)

Table 2

Average RMSE across subjects (standard error $\times 10^{-2}$ are within brackets) for the visits $j = 2,3,\ldots, 11$ using a dynamic parameter estimation and a static one.

Visit	Dynamic	Static
2	0.0568 (0.0364)	0.0567 (0.0364)
3	0.0515 (0.0207)	0.0550 (0.03.64)
4	0.0484 (0.0085)	0.0534 (0.0203)
5	0.0486 (0.0010)	0.0545 (0.0214)
6	0.0483 (0.0083)	0.0546 (0.0242)
7	0.0473 (0.0059)	0.0537 (0.0222)
8	0.0485 (0.0076)	0.0539 (0.0139)
9	0.0484 (0.0062)	0.0550 (0.0230)
10	0.0474 (0.0061)	0.0539 (0.0206)
11	0.0474 (0.0060)	0.0532 (0.0218)

Table 3

Simulation study: Coverage percentages along with standards deviation within brackets corresponding to the population mean of $\beta_{ikp}(\cdot)$ and $\gamma_{ikp}(\cdot)$.

Partition	T1	FLAIR	T1-gradient	FLAIR-gradient
1	93.4 (1.3)	92.0 (1.2)	94.6 (1.2)	92.7 (1.1)
2	94.6 (1.2)	93.6 (1.2)	93.2 (1.2)	92.2 (1.1)
3	93.2 (1.2)	93.9 (1.3)	93.0 (1.2)	91.5 (1.2)
4	95.2 (1.0)	95.3 (0.9)	94.1 (1.0)	93.2 (1.0)
5	94.7 (1.1)	93.6 (1.2)	94.0 (1.1)	92.0 (1.2)
6	94.7 (1.1)	91.5 (1.4)	94.5 (1.0)	91.6 (1.4)
7	94.1 (1.1)	93.4 (1.2)	93.6 (1.2)	91.7 (1.3)
8	94.1 (1.2)	93.4 (1.2)	94.2 (1.2)	93.0 (1.0)
9	94.4 (1.2)	92.5 (1.3)	92.4 (1.3)	92.9 (1.3)
10	95.2 (1.1)	91.1 (1.3)	96.3 (0.9)	92.6 (1.3)

Table 4

Simulation study: Coverage percentages along with standards deviation within brackets corresponding to the population standard deviation of $\beta_{ikp}(\cdot)$'s and $\gamma_{ikp}(\cdot)$'s.

Partition	T1	FLAIR	T1-gradient	FLAIR-gradient
1	92.4 (1.2)	91.2 (1.4)	91.9 (1.1)	89.4 (1.4)
2	91.7 (1.3)	89.0 (1.4)	91.1 (1.2)	89.9 (1.2)
3	92.3 (1.4)	91.0 (1.2)	89.7 (1.2)	90.2 (1.2)
4	92.6 (1.2)	90.2 (1.5)	91.8 (1.1)	89.8 (1.2)
5	90.6 (1.4)	89.4 (1.4)	91.1 (1.1)	89.9 (1.2)
6	91.0 (1.4)	89.4 (1.7)	90.3 (1.3)	89.9 (1.2)
7	92.0 (1.2)	89.5 (1.5)	91.5 (1.1)	89.3 (1.3)
8	90.6 (1.4)	88.7 (1.4)	90.4 (1.3)	89.3 (1.2)
9	92.5 (1.2)	90.5 (1.4)	92.1 (1.0)	89.5 (1.4)
10	92.0 (1.3)	89.0 (1.4)	91.0 (1.2)	89.8 (1.2)