A LAG FUNCTIONAL LINEAR MODEL FOR PREDICTION OF MAGNETIZATION TRANSFER RATIO IN MULTIPLE SCLEROSIS LESIONS

Gina-Maria Pomann,

Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina 27710, USA

Ana-Maria Staicu,

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, USA

Edgar J. Lobaton,

Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, North Carolina 27695, USA

Amanda F. Mejia,

Department of Statistics, Indiana University Bloomington, Bloomington, Indiana 47405, USA

Blake E. DEWEY,

National Institute of Neurological Disorders and Stroke NIH, Bethesda, Maryland 20892, USA

Daniel S. Reich,

National Institute of Neurological Disorders and Stroke NIH, Bethesda, Maryland 20892, USA

Elizabeth M. Sweeney,

Statistics Department, Rice University, Houston, Texas 77005, USA

Russell T. Shinohara

Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatisti Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Abstract

We propose a lag functional linear model to predict a response using multiple functional predictors observed at discrete grids with noise. Two procedures are proposed to estimate the regression parameter functions: (1) an approach that ensures smoothness for each value of time using generalized cross-validation; and (2) a global smoothing approach using a restricted maximum likelihood framework. Numerical studies are presented to analyze predictive accuracy in many realistic scenarios. The methods are employed to estimate a magnetic resonance imaging (MRI)-

gp43@duke.edu .

based measure of tissue damage (the magnetization transfer ratio, or MTR) in multiple sclerosis (MS) lesions, a disease that causes damage to the myelin sheaths around axons in the central nervous system. Our method of estimation of MTR within lesions is useful retrospectively in research applications where MTR was not acquired, as well as in clinical practice settings where acquiring MTR is not currently part of the standard of care. The model facilitates the use of commonly acquired imaging modalities to estimate MTR within lesions, and outperforms cross-sectional models that do not account for temporal patterns of lesion development and repair.

Keywords

Functional data analysis; functional linear model; magnetization transfer ratio; image analysis

1. Introduction.

Multiple sclerosis (MS) is an inflammatory disease of the central nervous system in which the myelin sheaths around the axons of the neurons in the brain and spinal cord are damaged. Focal areas of tissue damage in the brain and spinal cord are known as "lesions." These lesions appear in various locations in the brain at different time points. Magnetic resonance imaging (MRI) is sensitive to these lesions and is the most important tool used by clinicians to diagnose and monitor the disease. One benchmark diagnostic criterion that is observed via MRI is the number of white matter lesions [McDonald et al. (2001), Polman et al. (2005)]. Conventional MRI modalities that are used in routine clinical practice to observe lesions are T1-weighted (T1w), T2-weighted (T2w) and T2-weighted fluid attenuated inversion recovery (FLAIR) images. However, it has been argued that rudimentary measures of lesion burden, including the number and volume of lesions, only moderately correlate with disability [Barkhof (2002), Brex et al. (2002), Hawkins et al. (1990), Schmierer et al. (2004)]. Therefore, much research in the field has centered on investigating the use of images that quantify tissue damage (e.g., white matter) and are believed to correlate with disability, such as magnetization transfer ratio (MTR) maps, which are sensitive to the degree of demyelination as well as lesion remyelination. Axons of the neurons in the white matter of the brain are insulated with a myelin sheath, which increases the speed at which impulses propagate along the axon. Demyelinating diseases, such as MS, destroy this layer of myelin, in a process called demyelination. Remyelination is the process of creating new myelin sheaths along these axons.

Here, we estimate MTR within lesions after lesion incidence using only the T1w and FLAIR images. MTR is not typically acquired in clinical settings, yet it would be a valuable addition to the typically acquired images and is believed to contain information about the disease process in MS [Chen et al. (2007, 2008)]. While MTR is not typically collected in clinical practice, the FLAIR and T1w volumes are routinely collected. Prediction of MTR in lesions will be useful in research applications where MTR was not acquired, as well as in clinical practice settings where acquiring MTR is not currently part of the standard of care. In practice, MTR is also an imaging sequence known to contain much noise [Reich et al. (2015)]. Predicting MTR within lesions with the longitudinal information in the FLAIR and T1w volume could therefore potentially provide a less noisy measurement of the MTR.

We propose a voxel-level model that uses sparsely sampled functional measurements of the T1w and FLAIR images over a fixed window of time to predict MTR within lesions. The model yields improvement over cross-sectional models that do not account for temporal patterns of early lesion evolution. This will ultimately allow researchers to study the effect of treatments on tissue damage in patients with MS.

We analyze data that were obtained from a group of 53 MS patients at the National Institute of Neurological Disorders and Stroke (NINDS). Each patient was scanned between 9 and 38 times over a period of 5.5 years. In this longitudinal study, the number of visits varies for each patient; MTR maps, T1w, and FLAIR images were acquired at each of the visits. For our analysis, the voxels in all of the images are temporally registered to obtain voxel-level intensity trajectories over time. The images are also segmented to obtain the set of voxel trajectories that correspond to areas of white matter containing MS lesions. Our objective is to relate current MTR intensity in a given voxel (within a lesion) to the T1w and FLAIR intensities in the same voxel observed over time. The observed voxel-level trajectories are discrete observations arising from a continuous process over time. Therefore, to predict MTR, we employ a function-on-function regression framework.

We propose a model for prediction of the functional response (MTR) beginning 6 months after lesion incidence, using the prior six months of the observed intralesional functional predictors (T1w and FLAIR) at the voxel level. As documented by Van Den Elskamp et al. (2008), MTR values may decrease subtly in areas of future lesion formation within a six-month period. Current methods in the functional data analysis literature do not model the relationship between a functional response and more than one functional predictor when it is assumed that only the recent past of the predictors affect the current value of the response. We develop a model that is able to use only the intensity of the T1w and FLAIR images observed in the previous six months of historical data to predict MTR at the current time point.

Functional data analysis has been a very active research area due to a surge of applications; see Besse and Ramsay (1986), Ferraty, Vieu and Viguier-Pla (2007), Horváth and Kokoszka (2012), Ramsay and Silverman (2005), Rice and Silverman (1991), to name a few. In particular, function-on-function regression has attracted increasing interest, starting with the functional linear model introduced in Ramsay and Silverman (2005) for which Scheipl and Greven (2015) address identifiability issues. Meyer et al. (2015) proposes a multilevel, wavelet-based, Bayesian function-on-function regression framework. Additionally, Scheipl, Staicu and Greven (2015) proposed a framework that incorporates additive regression models in penalized function-on-function regression. Recently, Ivanescu et al. (2015) developed a penalized regression method for association models between functional responses and multiple functional predictors, where the current response is allowed to depend on the entire trajectory of the predictor. Until recently, the work related to function-on-function regression has been able to account for only one functional predictor; see He et al. (2010), Ramsay and Dalzell (1991), Ramsay and Silverman (2005), Wu, Fan and Müller (2010), Yao, Müller and Wang (2005b). For a complete review of the current literature in this area, see Morris (2015).

Motivated by the medical application, we focus on the scientific problem of estimating the functional response at a specific time point using only a fixed window of history of the functional predictors. To address this problem when there is only one covariate, Malfait and Ramsay (2003) introduced the Historical Functional Linear Model (HFLM). In their model, they represent the coefficient function using a triangular basis function that is estimated using information from the predictor, and estimate a coefficient function at each observation point. Harezlak et al. (2007) present a penalized approach that allows for varying lags for the HFLM. Kim, entürk and Li (2011) propose a least-squares approach to fitting this model by estimating the coefficient function at each timepoint known as the Recent HFLM. However, their methods do not account for more than one predictor.

In this paper we study function-on-function regression where the response and predictors are functions defined on the same domain, multiple functional predictors are considered, and the response depends solely on a fixed window of the functional predictors. To estimate the regression parameter functions, we propose two methods: (1) a semi-local smoothing approach; and (2) a global smoothing approach.

We introduce the modeling framework in Section 2. Two estimation procedures are presented in Sections 2.1 and 2.2. Medically relevant extensions to sparse and noisy predictors are discussed in Section 2.3. We evaluate the proposed methods in a simulation experiment in Section 3. In Section 4, the proposed estimation approaches are then applied for predicting MTR trajectories using the recent past of T1w and FLAIR voxel-level intensities.

2.   Methodology.

Suppose we observe data, $[Y_{ij}, t_{ij} : i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_{0i}\}]$, $[X_{1ij}, r_{1ij} : i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_{1i}\}]$ and $[X_{2ij}, r_{2ij} : i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_{2i}\}]$, where $t_{ij}, r_{1ij}, r_{2ij} \in \mathcal{T}$, a bounded and closed interval. For example, Y_{ij} corresponds to the MTR measurement at voxel i at time t_{ij}. In our data, there are n voxels under consideration and m_{0i} observations for voxel i. The variables X_{1ij} and X_{2ij} correspond to the smooth T1w and FLAIR intensities for voxel i at times r_{1ij} and r_{2ij}, respectively. Our data analysis relies on the working assumption that the measurements are independent across voxels. Additionally, we assume the voxels in the scans of patients are properly registered over time. Throughout the rest of this section we consider $r_{1ij} = t_{ij}$ and $r_{2ij} = t_{ij}$, and make use of r_{1ij} and r_{2ij} again in Section 3 where we present simulations.

Let $X_{1ij} = X_{1i}(t_{ij})$ and $X_{2ij} = X_{2i}(t_{ij})$, and assume that $X_{1i}(\cdot)$ and $X_{2i}(\cdot)$ are independent and square-integrable random smooth functions over $\mathcal{T}$. Without loss of generality, let us assume $E[X_{1i}(t)] = E[X_{2i}(t)] = 0$. Assume that the predictor functions are observed on a dense grid of points and without noise, and the response is observed on a dense and equidistant grid of points $t_{ij} = t_j$; more realistic sampling designs and scenarios will be considered in Section 2.3. We assume the fixed-lag HFLM with multiple covariates for the response Y_{ij},

$$Y_{ij} = f_i(t_j) + \varepsilon_i(t_j),$$

$$f_i(t_j) = \beta_0(t_j) + \int_0^{\Delta_1} \beta_1(s, t_j) X_{1i}(t_j - s) ds$$

$$+ \int_0^{\Delta_2} \beta_2(s, t_j) X_{2i}(t_j - s) ds,$$

(2.1)

where $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, m_i\}$, $\beta_0 : \mathscr{T} \to \mathbb{R}$ and $\beta_1, \beta_2 : [0, \Delta] \times \mathscr{T} \to \mathbb{R}$ are continuous functions, and $\varepsilon_i(\cdot)$ is an independent (across i) measurement error with zero-mean and covariance σ_ε^2. $\varepsilon_i(\cdot)$ is also assumed to be independent of $X_{1i}(\cdot)$ and $X_{2i}(\cdot)$.

This model (2.1) is used throughout the paper. For intuition, notice that $f_i(t_j) = \beta_0(t_j) + \int_{t_j - \Delta_1}^{t_j} \beta_1(t_j - s, t_j) X_{1i}(s) ds + \int_{t_j - \Delta_2}^{t_j} \beta_2(t_j - s, t_j) X_{2i}(s) ds$. The model assumes that, given the entire predictor trajectories, the response is affected by the predictor values over only recent time windows of length Δ_1 and Δ_2, respectively; these lag parameters are chosen based on scientific knowledge. For simplicity of illustration, we let $\Delta = \Delta_1 = \Delta_2$. The time-varying coefficient functions, $\beta_1(\cdot, \cdot)$ and $\beta_2(\cdot, \cdot)$, weigh the predictor values over the Δ lag-window, and quantify the effect of the predictors on the current response. While historical functional linear models have been presented in the current literature [Kim, Şentürk and Li (2011), Malfait and Ramsay (2003)] they have not accounted for multiple covariates as model (2.1) does.

We discuss two approaches for estimating the model parameters $\beta_0(\cdot)$, $\beta_1(\cdot, \cdot)$ and $\beta_2(\cdot, \cdot)$. The first approach considers a pointwise (PW) estimation procedure for each observed time point t_{ij}. The second approach considers a basis function expansion using prespecified bases for each model parameter and conceptually controls the smoothness over $\mathscr{T}$ globally (GB). The two approaches are detailed in the next two sections.

2.1. PW model representation.

Model assumption.—As is common in the nonparametric regression literature, we assume that smooth coefficient functions can be represented using a basis function expansion. Specifically, let $\{B_{s1,k}(s)\}_k$ and $\{B_{s2,k}(s)\}_k$ be two prespecified functional bases on $[0, \Delta]$. Then $\beta_1(s, t)$ is assumed to be represented as a tensor product of the form $\beta_1(s, t) = \sum_{k=1}^{K_1} B_{s1,k}(s) \alpha_{1k}(t)$, where K_1 is large enough to capture the flexibility of the model and the $\alpha_{1k}(t)$'s are unknown time-varying coefficient functions defined on $\mathscr{T}$. We use a similar basis expansion for $\beta_2(s, t) = \sum_{k=1}^{K_2} B_{s2,k}(s) \alpha_{2k}(t)$, where $\alpha_{2k}(t)$'s are unknown functions defined on $\mathscr{T}$.

Without loss of generality, we assume that $K_1 = K_2 = K$. Although many choices for basis functions are possible, we use B-spline functions of degree 4 with 6 equally spaced interior knots over $[0, \Delta]$, which correspond to $K = 10$, as the sets $\{B_{s1,k}(s)\}_k$ and $\{B_{s2,k}(s)_k\}$. B-splines are determined by the degree of the polynomial, and the number and location of the knots. The degree usually does not have a strong effect on the model performance [Wood

(2006)]. Typically the knots are placed at equally spaced locations or equally spaced quantile locations. Furthermore, it is common practice to use a large number of basis functions and use a smoothness parameter to determine the smoothness of the fit [Ruppert, Wand and Carroll (2003)]. The value of K in our application was selected empirically. Substituting the expressions for the coefficient functions into equation (2.1) yields

$$f_i(t_j) = \beta_0(t_j) + \sum_{k=1}^{K} \alpha_{1k}(t_j) \int_0^{\Delta} X_{1i}(t_j - s) B_{s1,k}(s)\,ds$$
$$+ \sum_{k=1}^{K} \alpha_{2k}(t_j) \int_0^{\Delta} X_{2i}(t_j - s) B_{s2,k}(s)\,ds. \tag{2.2}$$

Furthermore, by denoting $\widetilde{X}_{1k,i}(t_j) = \int_0^{\Delta} X_{1i}(t_j - s) B_{s1,k}(s)\,ds$ and $\widetilde{X}_{2k,i}$ similarly, equation (2.2) reduces to

$$f_i(t_j) = \beta_0(t_j) + \sum_{k=1}^{K} \widetilde{X}_{1k,i}(t_j)\alpha_{1k}(t_j) + \sum_{k=1}^{K} \widetilde{X}_{2k,i}(t_j)\alpha_{2k}(t_j). \tag{2.3}$$

Fitting.—This representation captures the dependence between the precomputed covariates ($\widetilde{X}_{1k,i}$ and $\widetilde{X}_{2k,i}$) and the response using a varying-coefficient model. Nevertheless, when the functional predictors are observed at sparse sampling points for each i and possibly contaminated with error, these derived quantities may not be directly computable. In this section, we develop an estimation procedure of the HFLM model that depends on the continuous second moments of the additive predictors. It does not require direct computation of $\widetilde{X}_{pk,i}(t)$ and accounts for the correlation between the functional predictors. A working independence assumption over time for the error process $\varepsilon_i(\cdot)$ is assumed for fitting. This representation of the function-on-function regression model will allow us to obtain smooth pointwise estimates for the coefficient functions. A similar approach was considered by Kim, entürk and Li (2011) for the recent history functional linear model with one functional predictor. Here we present estimation of the coefficient functions and prediction of the response for the dense case; and extensions to more realistic settings are considered in Section 2.3.

Fix t_j, and define $a_{1kj} = a_{1k}(t_j)$ and $a_{2kj} = a_{2k}(t_j)$. Notice that $\beta_0(t_j) = E[Y_{ij}]$, and thus $\beta_0(\cdot)$ can be estimated by smoothing Y_{ij}'s and using a working independence assumption. By an abuse of notation, in the rest of the subsection, let Y_{ij} denote $Y_{ij} - \hat{\beta}_0(t_j)$, where $\hat{\beta}_0(\cdot)$ is a smooth estimator of $\beta_0(\cdot)$. We estimate a_{1kj}'s and a_{2kj}'s by minimizing the following penalized criterion:

$$\sum_i \left\{ Y_{ij} - f_i(t_j) \right\}^2 + \lambda_1 \sum_k \alpha_{1kj}^2 + \lambda_2 \sum_k \alpha_{2kj}^2, \tag{2.4}$$

where $\lambda_1 > 0$ and $\lambda_2 > 0$ are the regularization parameters. Let $a_{pj} = [a_{p1j} \cdots a_{pKj}]^{\top}$ for $p = 1, 2$. Then the minimizer is

$$
\begin{bmatrix} \hat{\alpha}_{1j} \\ \hat{\alpha}_{2j} \end{bmatrix} = \left(Z_j^\top Z_j + \begin{bmatrix} \lambda_1 I_K & 0 \\ 0 & \lambda_2 I_K \end{bmatrix} \right)^{-1} \left(Z_j^\top (\mathbf{Y}_j - \mu_Y(t_j)) \right),
\tag{2.5}
$$

where I_K is the $K \times K$ identity matrix, $\mathbf{Y}_j$ is the vector of all observations at time t_j, $\mu_Y(t) = E[Y(t)]$, and

$$
Z_j = \begin{bmatrix} \widetilde{X}_{11,1}(t_j) & \cdots & \widetilde{X}_{1K,1}(t_j) & \widetilde{X}_{21,1}(t_j) & \cdots & \widetilde{X}_{2K,1}(t_j) \\ \vdots & & \vdots & \vdots & & \vdots \\ \widetilde{X}_{11,n}(t_j) & \cdots & \widetilde{X}_{1K,n}(t_j) & \widetilde{X}_{21,n}(t_j) & \cdots & \widetilde{X}_{2K,n}(t_j) \end{bmatrix}.
\tag{2.6}
$$

For fixed t_j, the analytic solution (2.5) is the solution in a ridge regression framework for fixed values of the tuning parameters λ_1 and λ_2. In our case, a_{1kj}'s and a_{2kj}'s represent the evaluations of smooth functions $a_{1k}(t)$ and $a_{2k}(t)$ at times $t = t_j$. However, selecting optimal values of λ_1 and λ_2 for each t could lead to high variability of the tuning parameters over t. One approach to accommodate this situation is to assume that the tuning parameters are constant across t.

The parameter functions $a_{1k}(t)$ and $a_{2k}(t)$ can be estimated using the probability limits of the standardized block submatrices, $Z_j^\top Z_j$ and $Z_j^\top Y_j$, that are included in expression (2.5). Specifically, for arbitrary t,

$$
\begin{bmatrix} \hat{a}_1(t) \\ \hat{a}_2(t) \end{bmatrix} = \left(\begin{bmatrix} \hat{G}_{11}(t) & \hat{G}_{12}(t) \\ \hat{G}_{21}(t) & \hat{G}_{22}(t) \end{bmatrix} + \begin{bmatrix} \frac{\lambda_1}{n} I_K & 0 \\ 0 & \frac{\lambda_2}{n} I_K \end{bmatrix} \right)^{-1} \begin{bmatrix} \hat{G}_{1Y}(t) \\ \hat{G}_{2Y}(t) \end{bmatrix},
\tag{2.7}
$$

where $\hat{G}_{pY}(t) = \left[\hat{G}_{pY1}(t) \cdots \hat{G}_{pYK}(t) \right]^\top$ is a vector, $\hat{G}_{pYl}(t)$ is an estimator of $G_{pYl}(t) = \mathrm{cov}\left(\widetilde{X}_{pl,i}(t), Y_i(t) \right)$, $\hat{G}_{pq}(t) = \left[\hat{G}_{\widetilde{X}_{pl}, \widetilde{X}_{qk}}(t) \right]_{kl}$ is a $K \times K$ matrix, and $\hat{G}_{\widetilde{X}_{pl}, \widetilde{X}_{qk}}(t)$ is an estimator of $G_{\widetilde{X}_{pl}, \widetilde{X}_{qk}}(t) = \mathrm{cov}\left(\widetilde{X}_{pl,i}(t), \widetilde{X}_{qk,i}(t) \right)$ for $p, q \in \{1, 2\}$. Although the model framework and estimation are discussed for two functional predictors, it is straightforward to extend it to accommodate three or more functional predictors.

To obtain the necessary quantities, observe that

$$
\begin{aligned}
&\mathrm{cov}\left(\widetilde{X}_{pl,i}(t), \widetilde{X}_{qk,i}(t) \right) \\
&= \int_0^\Delta \int_0^\Delta E\left[X_{pi}(t - s_1) X_{qi}(t - s_2) \right] B_{sp,l}(s_1) B_{sq,k}(s_2) \, ds_1 \, ds_2 \\
&= \int_0^\Delta \int_0^\Delta G_{X_p X_q}(t - s_1, t - s_2) B_{sp,l}(s_1) B_{sq,k}(s_2) \, ds_1 \, ds_2,
\end{aligned}
\tag{2.8}
$$

where $G_{X_p X_q}(s_1, s_2)$ is the covariance between $X_{pi}(s_1)$ and $X_{qi}(s_2)$. Similarly, we also have that $\mathrm{cov}\big(\widetilde{X}_{pl, i}(t), Y(t)\big) = \int_0^{\Delta} G_{X_p Y}(t - s, t) B_{sp, l}(s)\, ds$, where $G_{X_p Y}(s_1, s_2)$ is the covariance between $X_{pi}(s_1)$ and $Y_i(s_2)$.

To use expression (2.7), we first need to obtain estimates of $G_{11}(t)$, $G_{12}(t)$, $G_{21}(t)$, $G_{22}(t)$, $G_{1Y}(t)$ and $G_{2Y}(t)$. Once estimates of $a_1(t)$ and $a_2(t)$ are available, estimates for the coefficient functions are given by $\hat{\beta}_1(s, t) = \sum_{k=1}^{K_1} B_{s1, k}(s) \hat{a}_{1kj}(t)$ and similarly for $\hat{\beta}_2(s, t)$. We note that this estimation is restricted to the class of regression parameter functions that can be expanded using $B_{s1, k}(s)$ and $B_{s2, k}(s)$ for $k \in \{1, \ldots, K\}$. It follows that the response for the ith voxel can be predicted at any time point t by

$$\hat{Y}_i(t) = \int_0^{\Delta_1} \hat{\beta}_1(s, t) X_{1i}(t - s)\, ds + \int_0^{\Delta_2} \hat{\beta}_2(s, t) X_{2i}(t - s)\, ds. \tag{2.9}$$

Selection of the smoothing parameters λ_1 and λ_2 is very important. To ensure that the regularization parameters are the same for all t_j, these parameters can be selected using cross-validation (CV) to minimize an error-based criterion that we call the Normalized Prediction Error,

$$NPE_{(\lambda_1, \lambda_2)} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{m} \left[\sum_{j=1}^{m} |\hat{Y}_{ij} - Y_{ij}| \right] \Big/ \left[\frac{1}{m} \sum_{j=1}^{m} |Y_{ij}| \right] \right\}; \tag{2.10}$$

a similar criterion is used in Kim, Şentürk and Li (2011). The algorithm for fitting of the model using the *PW* approach is included in the Supplementary Material [Pomann et al. (2016)].

2.2. GB model representation.

Model assumption.—In this section, we propose the BG method, which takes a global smoothing approach to estimate the regression coefficient functions of the HFLM. This method estimates all the parameter functions $\beta_0(\cdot)$, $\beta_1(\cdot, \cdot)$ and $\beta_2(\cdot, \cdot)$ simultaneously. To do this, we assume an expansion of the coefficient functions of the form $\beta_0(t) = \sum_{l=1}^{M_0} B_{t0, l}(t) b_{0l}$, $\beta_1(s, t) = \sum_{l=1}^{M_1} B_{1, l}(s, t) b_{1l}$, and $\beta_2(s, t) = \sum_{l=1}^{M_2} B_{2, l}(s, t) b_{2l}$, where M_0, M_1 and M_2 are large enough constants to capture the variability of the coefficient functions. The discussion on the choice of K in the PW approach is applicable for M_0, M_1 and M_2. $\{B_{t0, l}(\cdot)\}$, $\{B_{1, l}(\cdot, \cdot)\}$ and $\{B_{2, l}(\cdot, \cdot)\}$ are preselected basis functions; and b_{0l}, b_{1l} and b_{2l} are unknown basis coefficients. For this study, the bivariate thin plate regression splines (TPRS) are selected as the sets $\{B_{p, l}(\cdot, \cdot)\}$ with $M_1 = M_2 = 30$, and one-dimensional TPRS are chosen as the set $\{B_{t0, l}(\cdot)\}$ with $M_0 = 10$. These basis functions are commonly used in the literature, and they are the default choices for the package *mgcv* [Wood (2006, 2011)] in R that is used in our study. As in the PW approach, other choices of basis functions could also be used with this framework.

Similar to Section 2.1, the conditional mean, $f_i(t_j)$, can be represented as

$$f_i(t_j) = \sum_{l=1}^{M_0} B_{0t,l}(t_j)b_{0l} + \sum_{l=1}^{M_1} \breve{X}_{1l,i}(t_j)b_{1l} + \sum_{l=1}^{M_2} \breve{X}_{2l,i}(t_j)b_{2l}, \tag{2.11}$$

where $\breve{X}_{1l,i}(t_j) = \int_0^{\Delta} X_{1i}(t_j - s)B_{1,l}(s, t_j)ds$ and $\breve{X}_{2l,i}(t_j)$ is defined analogously. This representation is inspired by Ivanescu et al. (2015), which presents a penalized function-on-function regression framework that uses the entire range for the predictor functions. Equation (2.11) looks similar to equation (2.3); the main difference is that the unknown coefficients, b_{0k}, b_{1k} and b_{2k} are not time varying. This modeling approach ensures smoothness of the regression parameter functions $\beta_1(s, t)$ and $\beta_2(s, t)$ over both s and t.

Fitting.—The coefficient vectors, $b_0 = \begin{bmatrix} b_{01} \cdots b_{0M_0} \end{bmatrix}^{\top}$, $b_1 = \begin{bmatrix} b_{11} \cdots b_{1M_1} \end{bmatrix}^{\top}$ and $b_2 = \begin{bmatrix} b_{21} \cdots b_{2M_2} \end{bmatrix}^{\top}$, are estimated by minimizing the following penalized criterion:

$$\sum_{i,j} \{Y_i(t_j) - f_i(t_j)\}^2 + \lambda_0 P_0(b_0) + \lambda_1 P_1(b_1) + \lambda_2 P_2(b_2), \tag{2.12}$$

where λ_0, λ_1 and λ_2 are smoothing parameters and P_0, P_1 and P_2 are the penalty terms [Ivanescu et al. (2015), Wood (2011)]. In particular, the integral of the square of the second derivative is a common measure for smoothness, in which case $P_0(b_0) = \int_{\mathcal{T}} \{\beta_0''(t)\}^2 dt = b_0^{\top} S_0 b_0$ and $P_p(b_p) = \int_{\mathcal{T}} \int_0^{\Delta} \left[\{\partial^2 \beta_p(s, t)/\partial s^2\}^2 + \{\partial^2 \beta_p(s, t)/\partial s \, \partial t\}^2 + \{\partial^2 \beta_p(s, t)/\partial t^2\}^2 \right] ds \, dt = b_p^{\top} S_p b_p$, where S_0 and S_p are penalty matrices specified by the choice of basis functions for $p = 0, 1, 2$ [Wood (2006)]. A working independence assumption over time for the error process $\varepsilon_t(\cdot)$ is assumed for fitting.

Selection of the optimal regularization parameters can be done via generalized cross-validation (GCV). However, empirical evidence suggests that GCV leads to undersmoothing [Reiss and Ogden (2009), Wood (2011)]. Instead, the problem can be posed as a mixed model of the form $\mathbf{X}_f b_f + \mathbf{Z} b_r$, where b_f contains the fixed effect parameters and b_r is a random effect vector [Wood (2006, 2011)]. In this context, λ_0, λ_1 and λ_2 represent variance components of the random effect covariance structure. This framework allows us to use a restricted maximum likelihood (REML) to estimate λ_0, λ_1 and λ_2 [Reiss and Ogden (2009), Wood (2006, 2011)]. This approach is also more robust to nonindependent error assumptions [Ivanescu et al. (2015), Krivobokova and Kauermann (2007)]. The algorithm for fitting of the model using the *GB* approach is included in the Supplementary Material [Pomann et al. (2016)].

Predictions of the functional response can be obtained using the same methodology presented for the *PW* method. In particular, given that $\hat{\beta}_0(t) = \sum_{l=1}^{M_0} B_{t0,l}(t)\hat{b}_{0l}$ and $\hat{\beta}_1(s, t) = \sum_{l=1}^{M_1} B_{1,l}(s,t)\hat{b}_{1l}$ with a similar expression for $\hat{\beta}_2(s, t)$, then $\hat{Y}_i(t) = \hat{\beta}_0(t) + \int_0^{\Delta_1} \hat{\beta}_1(s, t)X_{1i}(t-s)ds + \int_0^{\Delta_2} \hat{\beta}_2(s, t)X_{2i}(t-s)\,ds.$

2.3.   Noisy and sparsely observed data.

The estimation approach presented above is applicable to smooth predictors observed on a fine grid of time points. However, the response MTR and both covariates T1w and FLAIR are observed at few time points for each voxel. Additionally, the measurements are contaminated with noise. To accommodate this situation, let $W_{pi}(t)$ $X_{pi}(t)$ $e_{pi}(t)$ for $p = 1$, 2, where $\varepsilon_{pi}(\cdot)$ is a zero mean white noise process with variance σ_p^2. Assuming the same sampling design for the functional covariates, we distinguish 4 different scenarios according to whether the response is sparse or dense, and whether the functional predictors are sparse or dense. Namely, the scenarios are as follows: sparse/sparse, dense/sparse, sparse/dense and dense/dense. In the following we discuss in detail the case sparse/sparse which is the most relevant for our application; the remaining cases are detailed in the Supplementary Material [Pomann et al. (2016)].

PW method.—This approach requires smooth estimators of $G_{X_1 X_1}(s_1, s_2)$, $G_{X_2 X_2}(s_1, s_2)$, $GX_1 X_2(s_1, s_2)$, $G_{X_1 Y}(s_1, s_2)$ and $G_{X_2 Y}(s_1, s_2)$, which can be obtained using functional principal component analysis (FPCA). Mercer's theorem yields the spectral decomposition of the covariance $G_{X_p X_p}(s_1, s_2) = \sum_{l=1}^{\infty} \omega_{pl}\psi_{pl}(s_1)\psi_{pl}(s_2)$ in terms of non-negative eigenvalues ω_{p1} ω_{p2} $\cdots$ 0 and orthogonal eigenfunctions $\psi_{pl}(t)$ with $\int_{\mathcal{T}} \psi_{pl}(t)\psi_{pl'}(t)\,dt = 1(l = l')$, where $1(l = l')$ is the indicator function which equals 1 when $l = l'$ and 0=otherwise [Bosq (2000)]. Given that $E[X_{pi}(t)] = 0$, the above decomposition of the covariance implies that $X_{pi}(t)$ can be represented via the Karhunen–Loève (KL) expansion as $X_{pi}(t) = \sum_{l=1}^{\infty} \zeta_{pl,i}\psi_{pl}(t)$, where $\zeta_{pl,i} = \int_{\mathcal{T}} X_{pi}(t)\psi_{pl}(t)\,dt$ are commonly called the functional principal component (FPC) scores and are uncorrelated random variables with zero mean and variance equal to ω_{pl}. For practical and theoretical reasons, the KL expansion is often truncated [Di et al. (2009), Hall, Müller and Wang (2006), Yao, Müller and Wang (2005a)]; let $X_{1i}^{L_1} = \sum_{l=1}^{L_1} \zeta_{1l,i}\psi_{1l}(t)$ and $X_{2i}^{L_2} = \sum_{2i}^{L_2} \zeta_{2l,i}\psi_{2l}(t)$ be the truncated KL expansions of $X_1(t)$ and $X_2(t)$, respectively, for finite L_1 and L_2. The spectral decomposition of the smooth estimates of $G_{X_1 X_1}(s_1, s_2)$ and $G_{X_2 X_2}(s_1, s_2)$ yields the pairs of estimated eigenfunctions and eigenvalues $\{\hat{\psi}_{1l}, \hat{\omega}_{1l}\}_l$ and $\{\hat{\psi}_{2l}, \hat{\omega}_{2l}\}_l$, respectively. Then one can obtain consistent estimators of the scores $\hat{\zeta}_{1l,i} = \int_{\mathcal{T}} X_{1i}(t)\hat{\psi}_{1l}\,dt$ and $\hat{\zeta}_{2l,i} = \int_{\mathcal{T}} X_{2i}(t)\hat{\psi}_{2l}\,dt$ [Hall, Müller and Wang (2006), Zhang and Chen (2007)].

Since the response is observed on a sparse grid of points, bivariate kernel smoothing [Yao, Müller and Wang (2005a)] or spline-based smoothing [Wood (2006)] can be used to estimate the covariance functions. In our implementation, we make use of B-spline-based smoothing techniques [Crainiceanu, Staicu and Di (2009), Di et al. (2009)]. To account for the measurement error, the diagonal elements are left out when smoothing is carried out. The resulting covariance estimator is adjusted to be symmetric and positive semidefinite. Then the variance of the noise, σ_1^2 and σ_2^2, can be estimated based on the difference between the empirical pointwise variance of the corresponding observed predictors and the estimated pointwise variance $\hat{G}_{X_1 X_1}(s_1, s_2)$, $\hat{G}_{X_2 X_2}(s_1, s_2)$, [Staniswalis and Lee (1998)].

In this setting, the conditional expectation formula should be employed to estimating the scores, $\hat{\zeta}_{1l,i} = \hat{E}[\zeta_{1l,i} \mid \mathbf{W}_{1i}] = \hat{\eta}_{1l}\hat{\psi}_{1l}^{\top}(\hat{G}_{X_1,X_1} + \hat{\sigma}_1^2 I_{m_{1i} \times m_{1i}})^{-1}\mathbf{W}_{1i}$ [Yao, Müller and Wang (2005a)]. Here $\mathbf{W}_{1i}$ is a vector of length m_{1i} containing the observed values of the first predictor function, $\hat{\psi}_{1l}^{\top}$ is a vector of length m_{1i} with the jth entry equal to $\hat{\psi}_{1l}(t_{ij})$, $\hat{G}_{X_1,X_1}$ is a $m_{1i} \times m_{1i}$-dimensional matrix with the (j, j')th entry equal to $\hat{G}_{X_1,X_1}(t_{ij}, t_{ij'})$ and $I_{m_{1i} \times m_{1i}}$ is the $m_{1i} \times m_{1i}$ identity matrix. $\hat{\zeta}_{2l,i}$ can be estimated similarly. Yao, Müller and Wang (2005a) showed that, under the assumption that the predictors and their errors are jointly Gaussian, this equation yields the empirical best linear unbiased predictor of the scores.

In this case, there is an alternative approach to predict the response trajectory. Specifically, let $X_{pi}(t) = \sum_{l=1}^{L_p} \zeta_{pl,i} \psi_{pl}(t)$ be the KL expansion for the covariate functions. Then by substituting this into equation (2.2) we obtain

$$
\begin{aligned}
\hat{Y}_i(t) &= \hat{\beta}_0(t) + \sum_{k=1}^{K} \hat{\alpha}_{1k}(t) \int_0^{\Delta} \sum_{l=1}^{L_1} \hat{\zeta}_{1l,i}\hat{\psi}_{1l}(t-s) B_{s1,k}(s)\,ds \\
&+ \sum_{k=1}^{K} \hat{\alpha}_{2k}(t) \int_0^{\Delta} \sum_{l=1}^{L_2} \hat{\zeta}_{2l,i}\hat{\psi}_{2l}(t-s) B_{s2,k}(s)\,ds \\
&= \hat{\beta}_0(t) + \sum_{l=1}^{L_1} \hat{\zeta}_{1l,i}\hat{P}_{1l}(t) + \sum_{l=1}^{L_2} \hat{\zeta}_{2l,i}\hat{P}_{2l}(t),
\end{aligned}
\tag{2.13}
$$

where $\hat{\alpha}_{1k}(t)$ and $\hat{\alpha}_{2k}(t)$ are the estimators of $\alpha_{1k}(t)$ and $\alpha_{2k}(t)$, respectively, $\hat{P}_{1l}(t) = \sum_{k=1}^{K} \hat{\alpha}_{1k}(t) \int_0^{\Delta} \hat{\psi}_{1l}(t-s) B_{s1,k}(s)\,ds$ and $\hat{P}_{2l}(t)$ is defined similarly.

GB method.—Densely sampled covariate functions are needed for the GB method; hence, smooth estimators for the covariates can be obtained using the predicted scores, eigenfunctions and KL expansion; that is, $\hat{X}_{pi}(t) = \sum_{l=1}^{L_p} \hat{\zeta}_{pl,i}\hat{\psi}_{pl}(t)$. The methodology can be applied by using the predicted functional covariates in place of the true smooth functions.

3. Simulation studies.

This section presents results from an extensive simulation study used to evaluate the predictive performance of the two proposed methods, *PW* and *GB*. We consider two settings for data generated as in model (2.1): (I) the response and functional predictors are observed densely so that m_{0i}, m_{1i} and m_{2i} are large and the predictors are observed without measurement error; and (II) the functional predictors and response are observed on a sparse grid of points so that m_{0i}, m_{1i} and m_{2i} are small and the predictors are observed with measurement error.

We first generate the response and predictor functions on a dense and equally spaced grid of points, $t_{ij} \in \mathcal{T} = \{j/99 : j = 0, 1, 2, \ldots, 99\}$. In setting (I), we generate observed data, $\{t_{ij}, Y_{ij}, W_{1ij}, W_{2ij}\}_j$, where Y_{ij} is the response function and $W_{1ij} = X_{1i}(t_{ij})$ and W_{2ij}

$= X_{2i}(t_{ij})$ are the predictor functions measured without error. We define the predictors, $X_{1i}(t) = \sum_{k=1}^{2} u_{ik1}\sin(k\pi t) + u_{ik2}\cos(k\pi t)$ and $X_{2i}(t) = \sum_{k=1}^{3} v_{ik}\cos(2k\pi t)$ for u_{ik1}, u_{ik2} and $v_{ik1}\overset{\text{i.i.d.}}{\sim}N(0, 1/k^4)$. We use regression parameter functions $\beta_0(t) = e^{-(t-.5)^2}$, $\beta_1(s, t) = \cos(2\pi t)\cos(\pi s)$ and $\beta_2(s, t) = \cos(\pi t)\cos(\pi s)$. We take $\quad = 0.4$ so that $\beta_1, \beta_2 : [0, 0.4] \times [0, 1] \to [0, 1]$ and β_0 $[0, 1] \to [0, 1]$. Illustrations of the simulated data and predictions are included in the Supplementary Material [Pomann et al. (2016)].

The response for voxel $i \in 1, \ldots, n$ is generated from model (2.1) via numerical integration for each $t_{ij} \in \mathcal{T}$. The measurement error in model (2.1) is taken to be $\varepsilon_i(t) = \varepsilon_{a,i}(t) + \varepsilon_{g,i}(t)$, where $\varepsilon_{g,i}(t)$ is a zero-mean i.i.d. normal random variable with variance $\sigma_\varepsilon^2/2$, and $\varepsilon_{a,i}(t)$ is a zero-mean AR(1) process with covariance structure given by $\Sigma(t_j, t_{j'}) = (\sigma_\varepsilon^2/2)\rho^{|t_j - t_{j'}|}$ with $\rho = 0.5$. It follows that $\text{var}(\varepsilon_i(t)) = \sigma_\varepsilon^2$. Note that an AR(1) process is considered in the simulation study to analyze the effect of nonindependent (over time) noise structure in the fitting process. We simulate data under a framework similar to Ivanescu et al. (2015) and McLean et al. (2014) by using a version of the empirical signal to noise ratio ($eSNR_e$), which we define as $eSNR_\varepsilon = \left[\sum_{i,j}\left\{E(Y_{ij}) - \sum_i E(Y_{ij})/n\right\}^2/\left\{\sigma_\varepsilon^2(n-1)m\right\}\right]$. The value of σ_ε^2 is chosen by specifying the $eSNR_e$ for each Monte Carlo (MC) replication. In this section, we present results for $eSNR_e = 5$. It was observed that the predictive error decreases (increases) as $eSNR_e$ is increased (decreased). Results for $eSNR_e = 1$ as well as for responses generated with a white noise error process are included in the Supplementary Material [Pomann et al. (2016)].

In setting (II) the response and predictors are first generated as in setting (I). Then, to generate the sparse data, we evaluate the functions at points $r_{ik} \in \mathcal{R}_i$ for a randomly sampled subset $\mathcal{R}_i \subset \mathcal{T}$ of size m_i. To determine the number of observations for each voxel, we take $m_i\overset{\text{i.i.d.}}{\sim}$ Uniform on $\{10, 11, \ldots, 15\}$. This level of sparsity is consistent with the 13.6 average number of samples observed per curves in the real data (see the Data Analysis Details section in the Supplementary Material [Pomann et al. (2016)]). To generate the predictors that are observed with error, we take $W_{1ik} = X_1(r_{ij}) + \varepsilon_{1ik}$ and $W_{2ik} = X_2(r_{ij}) + \varepsilon_{2ik}$, where $\varepsilon_{1ik} \sim N(0, 0.1)$ and $\varepsilon_{2ik} \sim N(0, 0.05)$. The latter error variances were selected to correspond to a signal-to-noise ratio of 5.

The main simulation factors of interest are the number of observations and the lag parameter used for fitting the model. We present results for $n \in \{20, 50, 100\}$. To understand how the model performs when the lag parameter is misspecified, we let $\quad_F$ be the lag parameter used for fitting the model and present results for $\quad_F \in \{0.4, 0.6\}$. As expected, when $\quad_F < \quad$, both estimation methods have poor predictive accuracy and results in increased bias. Intuitively, this case corresponds to fitting a misspecified model, as it enforces the two regression parameters to be null on domains where they are not. Namely, $\beta_1(s, t) = 0$ and $\beta_2(s, t) = 0$ are enforced for $s \in [\quad_F, \quad]$.

To fit the model, the predictors are first standardized to have a zero mean function and pointwise variance equal to 1. For the model using the PW method, we consider the

bases sets $\{B_{s1,k}\}$ and $\{B_{s2,k}\}$ to be the set of $K = 10$ B-spline basis functions o degree 4 with 6 interior knots over $[0, \quad_F]$. This value of K was selected empirically. We use one tuning parameter λ that is selected using a grid search over $I_\lambda = [0.0001, 0.01]$ and cross-validation. The range was determined empirically. Using a larger range and two tuning parameters produced similar results. To obtain the smooth covariance and cross-covariance functions, $\hat{G}_{X_1X_1}(s_1, s_2)$, $\hat{G}_{X_2X_2}(s_1, s_2)$, $\hat{G}_{X_1X_2}(s_1, s_2)$, $\hat{G}_{X_1Y}(s_1, s_2)$ and $\hat{G}_{X_2Y}(s_1, s_2)$, we use bivariate smoothing, via tensor products of 20 univariate cubic regression splines with 15 equally spaced interior knots. The positive semidefiniteness of the covariance functions is ensured by truncating the eigenvalues to be positive.

For the *GB* method, the functions $B_{p,k}(s, t)$ are chosen to be 30 thin plate spline basis functions with 30 knots equally spaced. The penalty for equation (2.12) is taken to be $P_l(b_l) = b_l^\top S_l b_l$, where S_l is chosen to be the thin plate spline penalty [Wood (2006)]. In setting (II), the reconstructed trajectories are estimated using FPCA with 20 basis functions to smooth the covariance and cross-covariance functions; the percent variance explained is set to 99%. The smooth predictor trajectories are sampled over a grid of 100 equally spaced observation points on [0, 1]. To obtain the necessary smooth estimates, we use *fpca.sc* function of the package *refund* [Di et al. (2009), Goldsmith, Greven and Crainiceanu (2012), Staniswalis and Lee (1998), Yao, Müller and Wang (2005b)]. To obtain predictions, we use the *gam* function of the *mgcv* package [Wood (2006, 2011)].

The computational cost and empirical performance of the *PW* method depends directly on the range of values I_λ used to search for the optimal regularization parameter. Since the estimation of regression parameters for the *PW* method uses only the smooth covariance and cross-covariance estimates at each time t_{ij}, it is more computationally efficient than the *GB* method which uses large matrices that grow linearly with $\sum_{i=1}^{n} m_i$; see the Supplementary Material [Pomann et al. (2016)] for algorithms that outline the implementation details.

3.1. Simulation results.

To evaluate model performance, we present the relative mean square error (reMSE) for $Y(t)$ as $\text{reMSE}(Y) = \left(\sum_i \int_\Delta^T \{ \hat{Y}_i(t) - f_i(t) \}^2 dt \right) / \left(\sum_i \int_\Delta^T \{ f_i(t) \}^2 dt \right)$. Since our main objective is prediction, we report the mean and standard error of the reMSE(Y) values over 500 Monte Carlo (MC) replications. We compare the resulting prediction errors by evaluating improvement of quantity "A" over "B," defined to be $[(B - A)/B]\,100\%$. The discussion in this section focuses on the comparison of the reMSE(Y) values found in the last column of the tables provided.

First we describe and compare the results for the two estimation methods under the dense setting. Table 1 reports the simulation results for the *PW* and the *GB* methods under setting (I). On average, the *GB* method yields an improvement of 92% over the *PW* method in reMSE of prediction for this setting. For both the *PW* and *GB* methods, the case $_F=$ has smaller prediction errors than the case when $_F$. As expected, the prediction error decreases in both cases as the sample size increases. For the *PW* method, $n = 50$ yields an average improvement of 47% over the case of $n = 20$, and for $n = 100$ we observe

an improvement over the case when $n = 50$. Using the *GB* method with $n = 50$ yields an average improvement of 56% over the case when $n = 20$, and a clear improvement as the sample size increases.

Next, we compare performance under the dense and sparse settings for each of the estimation methods. As expected, the performance of both approaches is affected by the sparse sampling design of the functional covariates and response. For the *PW* method, prediction errors in setting (I) are on average 70% smaller than in setting (II). Similarly, for the *GB* method the prediction errors in setting (I) are on average 95% smaller than in setting (II).

Last, we describe and compare the results for the two estimation methods under the sparse setting (II) (see Table 2). On average, in these datasets, the *GB* method yields an improvement of 55% over the *PW* method. For *PW*, we observe that the case when $_F =$ has smaller prediction errors than the case when $_F$. For *GB*, on average, having $_F =$ yields a slight improvement over having $_F$. As expected, the prediction error decreases in both cases as the sample size increases.

4. Prediction of Magnetization Transfer Ratio within lesions of MS patients.

Recall that the data consist of MTR maps and T1w and FLAIR images obtained for 53 MS patients who were imaged as part of a natural history study at the National Institute of Neurological Disorders and Stroke. Time to each observed scan is measured from the first scan after lesion incidence. With this definition of observation times, each patient had between 9 and 38 observed scans with the maximum time ranging from 20 weeks to 5.5 years. Our objective is to predict MTR at voxels within lesions using the T1w and FLAIR measurements acquired within the previous six months. This is important as MTR is not typically acquired in clinical settings, yet it would be a valuable addition to the typically acquired images and is believed to contain information about the disease process in MS [Chen et al. (2007, 2008)]. Predicting the MTR in MS lesions from the T1w and FLAIR information will provide information within lesions, without needing to acquire the MTR volume. This information will be of use in research for retrospective analysis where the MTR is not available as well as in clinical practice in settings where the MTR is not acquired. For this model, the window is chosen to be six months because it has been reported that MTR values may decrease subtly in areas of future lesion formation within a six-month period [Van Den Elskamp et al. (2008)]. Previously, Mejia et al. (2016) proposed a cross-sectional model to predict quantitative T_1 maps, which estimate T_1 relaxation times as a function of T_1-weighted, T_2-weighted, *PD*-weighted and FLAIR images; and Suttner et al. (2015) applied this method to predict DTI measures. To the best of our knowledge, this is the only other work in the literature that has been presented to estimate images of quantitative MR values. Jog et al. (2013a) do attempt to estimate average quantitative tissue properties, but not the quantitative images themselves. In addition, work has been done to predict standard nonquantitative clinical images using cross-sectional models [Jog et al. (2013b); Roy, Carass and Prince (2011, 2013)].

It is common practice to use cross-sectional linear models to make other voxel-level predictions [Pomann et al. (2015), Sweeney et al. (2013a, 2013b)]. Therefore, we evaluate the performance of the proposed approaches against two cross-sectional models that do not take temporal dependence of the data into account. Section 4.1 discusses the preprocessing of the data, and Section 4.2 presents implementation details and results for this analysis.

4.1. Data processing.

Details about the data acquisition, preprocessing, image registration and normalization are provided in Sweeney et al. (2015). After image preprocessing, we temporally align the voxel trajectories by aligning the voxels to the time point when it first is identified as belonging to a new or enlarging lesion. We identify this time of incidence for each voxel using the Subtraction-Based Logistic Inference for Modeling and Estimation (SuBLIME) procedure [Sweeney et al. (2013a)]. For each voxel, we obtain the time marker that corresponds to the time when the voxel is first identified as being part of a lesion ("time zero"). We then measure all following time points as the time difference in weeks from this marker. The alignment is appropriate because our goal is to predict MTR in the time window after the lesion develops; see the Supplementary Material [Pomann et al. (2016)] for more details.

4.2. Model implementation and results.

After normalization and spatiotemporal registration, we measure time in weeks, and let Y_{ij} be the MTR at voxel i in the brain of a subject at week $t_{ij} \in [0, 190]$. Similarly, we let the predictors W_{1ij} and W_{2ij} be the normalized T1w and FLAIR intensities, respectively, of voxel i at time t_{ij}, which are assumed to be observed with noise. We consider these two predictors as they are commonly acquired images as part of clinical MRI studies. We omit times after 190 weeks, as there are too few observations beyond this point. After $_F = 26$ weeks (approximately six months), lesion evolution is approximately stable [Meier, Weiner and Guttmann (2007)], which justifies our choice for lag window.

Figure 1 presents images for a lesion for one of the patients with MS in our study. From top to bottom, the FLAIR, T1w and MTR images are displayed for this patient obtained over time in weeks from left to right (indicated below in white text). Time zero corresponds to lesion incidence for each voxel. The lesion is the hypointense region in the center of the image on the T1w and MTR images and the hyperintense on the FLAIR. Figure 2(a), (b) and (a) present the corresponding T1w, FLAIR and MTR voxel-level trajectories for this image.

In order to determine the benefit of the HFLM model for predicting MTR, we compare its performance against a cross-sectional model (CS),

$$Y_{ij} = \beta_0 + \beta_1 W_{1ij} + \beta_2 W_{2ij} + \varepsilon_{ij}. \tag{4.1}$$

Additionally, we compare the performance against a cross-sectional nonlinear model (CSNL),

$$Y_{ij} = \beta_0 + g_1[W_{1ij}] + g_2[W_{2ij}] + \varepsilon_{ij}, \tag{4.2}$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are unknown nonlinear functions. These models do not account for any temporal dependencies that may be present in the data. To fit the cross-sectional models, the predictors are standardized to have a mean zero and standard deviation of one. Predictions for the CSNL model are obtained using the *gam* function of the *mgcv* package in R with 10 thin plate basis functions with 10 equally spaced knots.

To fit the HFLM models, the predictors are first standardized to have a zero mean function and pointwise variance equal to 1. Prediction is performed only for voxels in lesions at times after the corresponding lesion incidence. The analysis is performed on the logit transform of the MTR trajectories since the response values lie in the range [0, 1]. The same analysis is also conducted on the raw data which yields comparable results (see the Supplementary Material [Pomann et al. (2016)]). There are some zero values in the response but no values of 1, and most of the MTR values are concentrated between 0.2 and 0.5. Hence, in order to avoid division by zero, we shift the data by 0.01 when performing the logit transformation; that is, for our analysis we define $\mathrm{logit}(y) = \log\left(\frac{(y + 0.01)}{1 - (y + 0.01)}\right)$ and $\mathrm{logit}^{-1}(\alpha) = \frac{e^\alpha}{e^\alpha + 1} - 0.01$.

When using the logit option in the implementation, the response data is transformed before fitting the model. Then predictions are transformed back to the probabilities using the logit^{-1} function. All analysis is conducted using R version 2.15.1.

For the *PW* method, the basis set $\{B_{s1,k}(s)\}_{k=1}^K$ and $\{B_{s2,k}(s)\}_{k=1}^K$ are taken to be the set of $K = 10$ B-spline basis functions of degree 4 with 6 interior knots over [0, 26]. The value of K was selected empirically. Similar results were observed for $K = 20$. We use one tuning parameter, λ, that is selected using GCV over [0.1,100]. To obtain the smooth covariance and cross-covariance functions, $\hat{G}_{X_1X_1}(s_1, s_2)$, $\hat{G}_{X_2X_2}(s_1, s_2)$, $\hat{G}_{X_1X_2}(s_1, s_2)$, $\hat{G}_{X_1Y}(s_1, s_2)$, and $\hat{G}_{X_2Y}(s_1, s_2)$, we use bivariate smoothing, via tensor products of univariate bases. The positive semidefiniteness of the covariance and cross-covariance functions is ensured by truncating the eigenvalues to be positive. The results are based on using the tensor product of 20 univariate cubic regression splines with 15 equally spaced interior knots.

For the *GB* method, the functions $B_{p,k}(s, t)$ are chosen to be 30 thin plate spline basis functions with 30 knots equally spaced. The penalty for equation (2.12) is taken to be $P_l(b_l) = b_l^\top S_l b_l$, where S_l is obtained by using the thin plate spline penalty [Wood (2006)]. The reconstructed trajectories are estimated for $t_{ij} \in [0, 190]$ using FPCA with 20 basis functions to smooth the covariance and cross-covariance functions and the percent variance explained set to 99%. To obtain the necessary smooth estimates, we use the *fpca.sc* function of the package *refund*. To obtain predictions, we use the *bam* function of the *mgcv* package [Wood (2006, 2011)].

Using this smoothing procedure, we are able to capture the natural variation in the predictor functions. To see this, we display a subset of voxels in the brain of a patient in this study; Figure 2(a) and (c) display the observed voxel-level predictor trajectories of the T1w and FLAIR images, respectively, and Figure 2(b) and (d) show the respective smoothed trajectories. The voxel-level trajectories in these figures correspond to the voxels in the lesion displayed on the observed images in Figure 1.

Since the true underlying MTR functions are unknown, we perform a 10-fold cross-validation of the data based on subject ID. The subject IDs are first randomly sorted, and then the corresponding voxel observations are split into groups. The first three groups consist of six subjects, and the other seven groups consist of five subjects. The validation was designed in this manner in order to produce a more realistic scenario in which imaging information is observed from entire subjects and prediction is also done for all voxels in the scan of a subject. Histograms of the number of voxels and observations per subject are included in the Supplementary Material [Pomann et al. (2016)].

We evaluate prediction by comparing the integrated mean squared error (*MSE*) of the models. We also compare this prediction error with the empirical measurement error of the MTR voxel-level trajectories. To do this, for each cross-validation step, we compute MSE/s_0^2 such that $s_0^2 = \left[\sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ Y_i(t_{ij}) - \tilde{Y}_i(t_{ij}) \right\}^2 \right] / \left[\sum_{i=1}^{n} m_i \right]$, where m_i is the number of time samples per curve, and n is the total number of observed voxels, and $\tilde{Y}_i(t_{ij})$ are the smoothed response trajectories without noise. A ratio of 1 indicates that the prediction is as good as smoothing the observed data. The smooth trajectories are obtained by using the *fpca.sc* function in the *refund* package in R with 20 basis functions to smooth the covariance and 99% variance explained to obtain the number of eigenfunctions used to reconstruct the data.

Figure 3(a) and (b) display the observed and smoothed MTR voxel-level trajectories, respectively, for the voxels in the lesion of Figure 1. For these same voxel-level trajectories, Figure 3(c) and (d) present the predicted response trajectories from using the *PW* and *GB* methods. Due to the lag window being fixed at six months, the estimation is restricted to $t > 26$ weeks. We see that the predictions obtained using our model are similar to the smoothed MTR trajectories.

Table 3 displays the prediction error results for the different approaches. The HFLM model consistently has higher predictive accuracy than the corresponding cross-sectional models. Both estimation approaches for the HFLM model perform comparably. This model yields a 61% improvement over the *CS* model and a 48% improvement over the *CSNL* model. These improvements indicate the importance of accounting for temporal dependence over the previous six months when predicting MTR. Similar results were observed for the average in-sample error.

This analysis was repeated for a lag window of 3 and 9 months, and it was observed that the error between the different window sizes was the same up to the third significant digit for both methods. It may be of future interest to explore the model with interactions as well as other commonly acquired predictors such as proton density volumes or T2w images. This analysis can provide further insight regarding which commonly acquired images can be used to predict MTR and analyze tissue damage. Observe that using both predictors in the HFLM model yields a 9% improvement over the model with only one predictor (see the Supplementary Material [Pomann et al. (2016)]). However, when using the cross-sectional models, this relationship is not maintained; each of the two-predictor models does not perform as well as the corresponding model with only the T1w intensities used as a

predictor. The models with only FLAIR intensities used as a predictor consistently do worse than all of the corresponding cross-sectional two-predictor models. These results confirm the qualitative assessment that the T1w image on its own contains detailed information about tissue damage. Furthermore, adding the FLAIR intensities to any of the cross-sectional models does not improve predictive power.

5. Conclusion.

We propose using the HFLM to estimate the current value of MTR in lesions using only the previous six months of both the T1w and FLAIR voxel intensities. We provide two methods to estimate the regression parameter functions: the *PW* method, a pointwise least squares approach; and the *GB* method, a global penalized functional regression approach. The *PW* method directly accounts for the covariance of the predictor functions by using a pointwise least squares approach to estimate the regression parameters. Additionally, in the case of sparse data, this method does not require the estimation of smooth trajectories as the *GB* method does. However, the *GB* method has a few advantages over the *PW* method. First, the performance of the *PW* method depends on the choice of the range of λ values for cross-validation. It was observed experimentally that if this range were not properly selected, then high out-of-sample prediction errors would be observed. Second, the *PW* method does not guarantee smoothness across time since estimated coefficients and predictions are performed at each time value. Finally, the *PW* method assumes that the β functions can be modeled as a sum of tensor products of basis functions, whereas the *GB* method makes use of general bivariate basis functions.

Our findings from the study suggest that the MRI signal within lesions depends on the local recent history of the signal at that location. Thus, the proposed model can facilitate the study of tissue damage in lesions of patients with MS. Furthermore, estimation of MTR using the HFLM model with the recent six months of both the T1w and FLAIR voxel trajectories outperforms cross-sectional models that do not take temporal dependence of the voxel-level trajectories into account. Future work will explore the use of this HFLM model to predict other quantitative imaging modalities and its application to other tissue classes. Estimability should be investigated further when making inference about the β coefficients is of interest.

The estimation of MTR in lesions using the traditional clinical FLAIR and T1w sequences has implications for both research and clinical practice. Estimating MTR within lesions with these sequences will provide additional information about lesions without needing to acquire MTR. As MTR has not been a traditional imaging modality, most retrospective data will not have MTR measured. Estimating the MTR in lesions may allow us to answer important research questions in large historical databases of MRI images. The proposed methods facilitate the estimation of MTR within lesions, providing additional information to clinicians to make treatment decisions in patients with MS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments.

REFERENCES

Barkhof F (2002). The clinico-radiological paradox in multiple sclerosis revisited. Curr. Opin. Neurol 15 239–245. [PubMed: 12045719]

Besse P and Ramsay JO (1986). Principal components analysis of sampled functions. Psychometrika 51 285–311.

Bosq D (2000). Linear Processes in Function Spaces: Theory and Applications. Lecture Notes in Statistics 149. Springer, New York.

Brex PA, Ciccarelli O, O'Riordan JI, Sailer M, Thompson AJ and Miller DH (2002). A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. N. Engl. J. Med 346 158–164. [PubMed: 11796849]

Chen JT, Kuhlmann T, Jansen GH, Collins DL, Atkins HL, Freedman MS, O'Connor PW, Arnold DL, Group CMS et al. (2007). Voxel-based analysis of the evolution of magnetization transfer ratio to quantify remyelination and demyelination with histopathological validation in a multiple sclerosis lesion. NeuroImage 36 1152–1158. [PubMed: 17543541]

Chen JT, Collins DL, Atkins HL, Freedman MS and Arnold DL (2008). Magnetization transfer ratio evolution with demyelination and remyelination in multiple sclerosis lesions. Ann. Neurol 63 254–262. [PubMed: 18257039]

Crainiceanu CM, Staicu A-M and Di C-Z (2009). Generalized multilevel functional regression. J. Amer. Statist. Assoc 104 1550–1561.

Di C-Z, Crainiceanu CM, Caffo BS and Punjabi NM (2009). Multilevel functional principal component analysis. Ann. Appl. Stat 3 458–488. [PubMed: 20221415]

Ferraty F, Vieu P and Viguier-Pla S (2007). Factor-based comparison of groups of curves. Comput. Statist. Data Anal 51 4903–4910.

Goldsmith J, Greven S and Crainiceanu, C. I. P. R. I. A. N. (2012). Corrected confidence bands for functional data using principal components. Biometrics

Hall P, Müller H-G and Wang J-L (2006). Properties of principal component methods for functional and longitudinal data analysis. Ann. Statist 34 1493–1517.

Harezlak J, Coull BA, Laird NM, Magari SR and Christiani DC (2007). Penalized solutions to functional regression problems. Comput. Statist. Data Anal 51 4911–4925.

Hawkins CP, Munro PMG, Mackenzie F, Kesselring J, Tofts PS, Du Boulay EPGH, Landon DN and McDonald WI (1990). Duration and selectivity of blood-brain barrier breakdown in chronic relapsing experimental allergic encephalomyelitis studied by gadolinium-DTPA and protein markers. Brain 113 365–378. [PubMed: 2328409]

He G, Müller H-G, Wang J-L and Yang W (2010). Functional linear regression via canonical analysis. Bernoulli 16 705–729.

Horváth L and Kokoszka P (2012). Inference for Functional Data with Applications Springer, New York.

Ivanescu AE, Staicu A-M, Scheipl F and Greven S (2015). Penalized function-on-function regression. Comput. Statist 30 539–568.

Jog A, Roy S, Carass A and Prince JL (2013a). Pulse sequence based multi-acquisition MR intensity normalization. In SPIE Medical Imaging 86692H–86692H. International Society for Optics and Photonics

Jog A, Roy S, Carass A and Prince JL (2013b). Magnetic resonance image synthesis through patch regression. In IEEE 10th International Symposium on Biomedical Imaging (ISBI) 350–353. IEEE, New York.

Kim K, entürk D and Li R (2011). Recent history functional linear models for sparse longitudinal data. J. Statist. Plann. Inference 141 1554–1566.

Krivobokova T and Kauermann G (2007). A note on penalized spline smoothing with correlated errors. J. Amer. Statist. Assoc 102 1328–1337.

Malfait N and Ramsay JO (2003). The historical functional linear model. Canad. J. Statist 31 115–128.

McDonald WI, Compston A, Edan G, Goodkin D, Hartung H-P, Lublin FD, Mcfarland HF, Paty DW, Polman CH, Reingold SC et al. (2001). Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. Ann. Neurol 50 121–127. [PubMed: 11456302]

Mclean MW, Hooker G, Staicu A-M, Scheipl F and Ruppert D (2014). Functional generalized additive models. J. Comput. Graph. Statist 23 249–269.

Meier DS, Weiner HL and Guttmann CRG (2007). Time-series modeling of multiple sclerosis disease activity: A promising window on disease progression and repair potential? Neurotherapeutics 4 485–498. [PubMed: 17599713]

Mejia A, Sweeney EM, Dewey B, Nair G, Sati P, Shea C, Reich DS and Shinohara RT (2016). Statistical estimation of T1 relaxation times using conventional magnetic resonance imaging. NeuroImage 133 176–188. [PubMed: 26732403]

Meyer MJ, Coull BA, Versace F, Cinciripini P and Morris JS (2015). Bayesian function-on-function regression for multilevel functional data. Biometrics 71 563–574. [PubMed: 25787146]

Morris JS (2015). Functional regression. Annual Reviews of Statistics and Its Applications 2 321–359.

Polman CH, Reingold SC, Edan G, Filippi M, Hartung H-P, Kappos L, Lublin FD, Metz LM, Mcfarland HF, O'Connor PW et al. (2005). Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald criteria". Ann. Neurol 58 840–846. [PubMed: 16283615]

Pomann G-M, Sweeney EM, Reich DS, Staicu A-M and Shinohara RT (2015). Scan-stratified case-control sampling for modeling blood-brain barrier integrity in multiple sclerosis. Stat. Med 34 2872–2880. [PubMed: 25939401]

Pomann G-M, Staicu A, Lobaton EJ, Mejia AF, Dewey BE, Reich DS, Sweeney EMEM and Shinohara RTRT (2016). Supplement to "A lag functional linear model for prediction of magnetization transfer ratio in multiple sclerosis lesions." DOI:10.1214/16-AOAS981SUPP.

Ramsay JO and Dalzell CJ (1991). Some tools for functional data analysis. J. R. Stat. Soc. Ser. B. Stat. Methodol 53 539–572.

Ramsay JO and Silverman BW (2005). Functional Data Analysis, 2nd ed. Springer, New York.

Reich DS, White R, Cortese IC, Vuolo O, Shea CD, Collins TL and Petkau J (2015). Sample-size calculations for short-term proof-of-concept studies of tissue protection and repair in multiple sclerosis lesions via conventional clinical imaging. Mult. Scler 21 1693–1704. [PubMed: 25662351]

Reiss PT and Ogden RT (2009). Smoothing parameter selection for a class of semiparametric linear models. J. R. Stat. Soc. Ser. B. Stat. Methodol 71 505–523.

RICE JA and Silverman BW (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. J. R. Stat. Soc. Ser. B. Stat. Methodol 53 233–243.

Roy S, Carass A and Prince J (2011). A compressed sensing approach for MR tissue contrast synthesis. In Information Processing in Medical Imaging 371–383. Springer, Berlin.

Roy S, Carass A and Prince JL (2013). Magnetic resonance image example-based contrast synthesis. IEEE Trans. Med. Imag 32 2348–2363.

Ruppert D, Wand MP and Carroll RJ (2003). Semiparametric Regression. Cambridge Series in Statistical and Probabilistic Mathematics 12. Cambridge Univ. Press, Cambridge.

Scheipl F and Greven S (2015). Identifiability in penalized function-on-function regression models Technical report, Univ. of Munich.

Scheipl F, Staicu A-M and Greven S (2015). Functional additive mixed models. J. Comput. Graph. Statist 24 477–501.

Schmierer K, Scaravilli F, Altmann DR, Barker GJ and Miller DH (2004). Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. Ann. Neurol 56 407–415. [PubMed: 15349868]

Staniswalis JG and Lee JJ (1998). Nonparametric regression analysis of longitudinal data. J. Amer. Statist. Assoc 93 1403–1418.

Suttner L, Mejia A, Dewey B, Sati P, Reich DS and Shinohara RT (2015). Statistical estimation of white matter microstructure from conventional MRI. UPenn Biostatistics Working Papers. Working Paper 44.

Sweeney EM, Shinohara RT, Shea CD, Reich DS and Crainiceanu CM (2013a). Automatic lesion incidence estimation and detection in multiple sclerosis using multisequence longitudinal MRI. Am. J. Neuroradiol 34 68–73. [PubMed: 22766673]

Sweeney EM, Shinohara RT, Shiee N, Mateen FJ, Chudgar AA, Cuzzocreo JL, Calabresi PA, Pham DL, Reich DS and Crainiceanu CM (2013b). OASIS is automated statistical inference for segmentation, with applications to multiple sclerosis lesion segmentation in MRI. NeuroImage: Clinical 2 402–413. [PubMed: 24179794]

Sweeney EM, Shinohara RT, Dewey BE, Schindler MK, Muschelli J, Reich DS, Crainiceanu CM and Eloyan A (2015). Relating multi-sequence longitudinal intensity profiles and clinical covariates in new multiple sclerosis lesions. Preprint. Available at arXiv:1509.08359

Van Den Elskamp IJ, Lembcke J, Dattola V, Beckmann K, Pohl C, Hong W, Sandbrink R, Wagner K, Knol DL, Uitdehaag B et al. (2008). Persistent T1 hypointensity as an MRI marker for treatment efficacy in multiple sclerosis. Mult. Scler 14 764–769. [PubMed: 18611989]

Wood SN (2006). Generalized Additive Models: An Introduction with R Chapman & Hall, Boca Raton, FL.

Wood SN (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. Ser. B. Stat. Methodol 73 3–36.

Wu Y, Fan J and Müller H-G (2010). Varying-coefficient functional linear regression. Bernoulli 16 730–758.

Yao F, Müller H-G and Wang J-L (2005a). Functional data analysis for sparse longitudinal data. J. Amer. Statist. Assoc 100 577–590.

Yao F, Müller H-G and Wang J-L (2005b). Functional linear regression analysis for longitudinal data. Ann. Statist 33 2873–2903.

Zhang J-T and Chen J (2007). Statistical inferences for functional data. Ann. Statist 35 1052–1079.
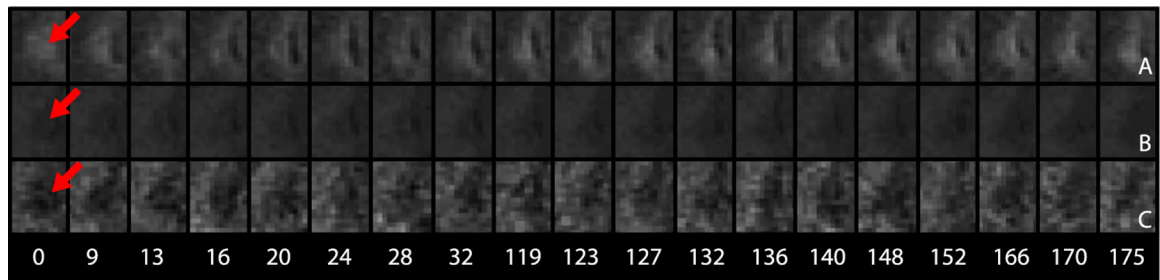
Fig. 1.
From top to bottom, axial slices of an MS lesion in the FLAIR (A), T1w (B) and MTR (C) sequences. This MS patient is imaged over time in weeks from left to right. Time zero corresponds to lesion incidence, and the lesion on each sequence is denoted at time zero with a red arrow. On the MTR and T1w images the lesion is the hypointense region; on the FLAIR the lesion is hyperintense.
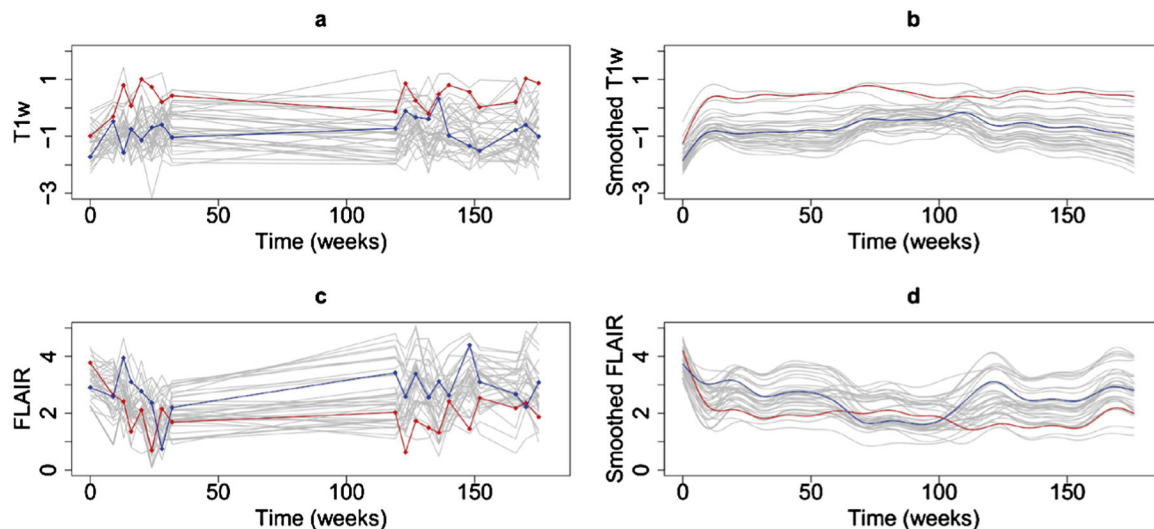
Fig. 2.
(a) Normalized T1w voxel intensities for a subset of the lesion voxels in the same patient as in Figure 1; (b) corresponding smoothed T1w voxel intensities using 20 basis functions to smooth all of the data; (c) corresponding FLAIR voxel intensities; (d) corresponding smoothed FLAIR voxel intensities using 20 basis functions to smooth all of the data. The red and blue curves are two particular curves.
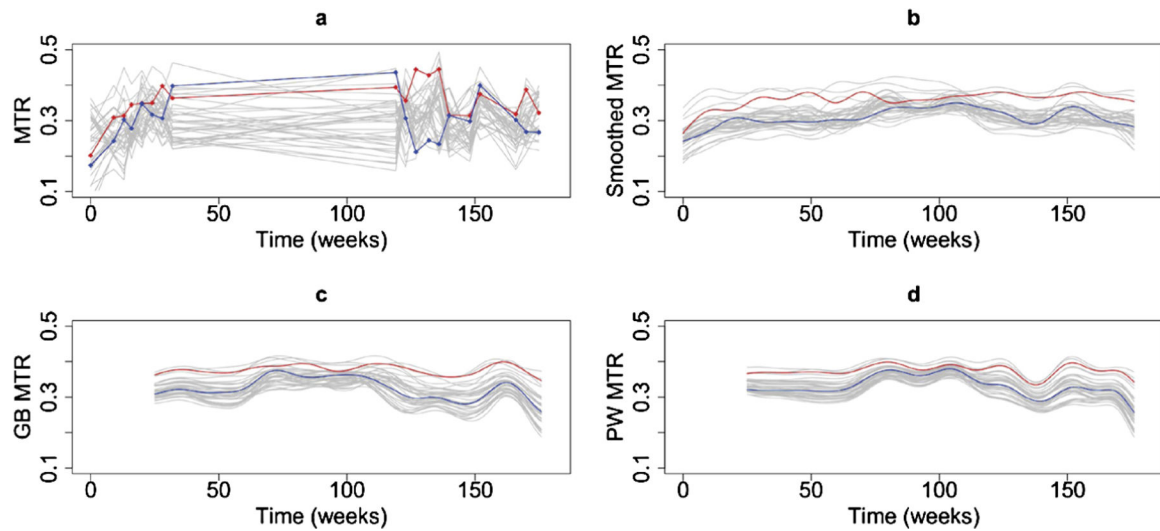
Fig. 3.
(a) MTR voxel intensities for a subset of the lesion voxels in the same patient as in Figure 1; (b) smoothed MTR voxel intensities using 20 basis functions to smooth all of the data; (c) predicted MTR voxel intensities using the GB method; (d) predicted MTR voxel intensities using the PW method. The values for the predicted MTR are not available for the first 26 weeks since this is the window used for prediction. The red and blue curves are two particular curves.

Table 1

Prediction Error—Setting (I), Average Error Variance = 0.012

n		F	reMSE(Y) (SE)—PW Method	reMSE(Y) (SE)—GB Method
20	0.4	0.4	5.1e–3 (5.9e–3)	5.2e–4 (1.6e–4)
20	0.4	0.6	6.5e–3 (1.1e–2)	6.5e–4 (2.3e–4)
50	0.4	0.4	2.9e–3 (4.6e–3)	2.3e–4 (6.4e–5)
50	0.4	0.6	3.2e–3 (2.8e–3)	2.9e–4 (8.4e–5)
100	0.4	0.4	1.8e–3 (1.4e–3)	1.2e–4 (3.2e–5)
100	0.4	0.6	2.1e–3 (2.0e–3)	1.5e–4 (4.3e–5)

Table 2

Prediction Error–Setting (II), Average Error Variance = 0.012

n		F	reMSE(Y) (SE)—PW Method	reMSE(Y) (SE)—GB Method
20	0.4	0.4	1.6e–2 (2.4e–2)	7.1e–3 (1.9e–2)
20	0.4	0.6	2.2e–2 (7.7e–2)	9.0e–3 (6.1e–2)
50	0.4	0.4	9.6e–3 (4.3e–3)	4.2e–3 (3.2e–3)
50	0.4	0.6	1.0e–2 (4.6e–3)	4.0e–3 (2.4e–3)
100	0.4	0.4	6.7e–3 (4.3e–3)	3.6e–3 (5.7e–3)
100	0.4	0.6	7.1e–3 (3.1e–3)	3.2e–3 (3.9e–3)

Table 3

Prediction error from 10-fold cross-validation on MTR data with T1w and FLAIR predictors. Integrated mean squared error (MSE) along with the average MSE/s_0^2, where s_0^2 is the estimated measurement error in the observed MTR voxel trajectories for all models and estimation approaches. The last two columns show the improvement for each approach over the CS and CSNL models

Model	$MSE[SE] \cdot (10^3)$	MSE/s_0^2	Imp. over CS	Imp. over CSNL
CS	12.7 [3.4]	4.1	0%	−32%
CSNL	9.6 [2.7]	3.1	31%	0%
HFLM (*GB*)	5.0 [2.1]	1.6	61%	48%
HFLM (*PW*)	4.9 [2.1]	1.6	61%	49%