Functional-Mixed Effects Models for Candidate Genetic Mapping in Imaging Genetic Studies

Ja-An Lin,[1] Hongtu Zhu,[1,2]* Ahn Mihye,[1] Wei Sun,[1,3] Joseph G. Ibrahim,[1] and for the Alzheimer's Neuroimaging Initiative[†]

[1]Department of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America; [2]Biomedical Research Imaging Center, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America; [3]Department of Genetics, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

ABSTRACT: The aim of this paper is to develop a functional-mixed effects modeling (FMEM) framework for the joint analysis of high-dimensional imaging data in a large number of locations (called voxels) of a three-dimensional volume with a set of genetic markers and clinical covariates. Our FMEM is extremely useful for efficiently carrying out the candidate gene approaches in imaging genetic studies. FMEM consists of two novel components including a mixed effects model for modeling nonlinear genetic effects on imaging phenotypes by introducing the genetic random effects at each voxel and a jumping surface model for modeling the variance components of the genetic random effects and fixed effects as piecewise smooth functions of the voxels. Moreover, FMEM naturally accommodates the correlation structure of the genetic markers at each voxel, while the jumping surface model explicitly incorporates the intrinsically spatial smoothness of the imaging data. We propose a novel two-stage adaptive smoothing procedure to spatially estimate the piecewise smooth functions, particularly the irregular functional genetic variance components, while preserving their edges among different piecewise-smooth regions. We develop weighted likelihood ratio tests and derive their exact approximations to test the effect of the genetic markers across voxels. Simulation studies show that FMEM significantly outperforms voxel-wise approaches in terms of higher sensitivity and specificity to identify regions of interest for carrying out candidate genetic mapping in imaging genetic studies. Finally, FMEM is used to identify brain regions affected by three candidate genes including CR1, CD2AP, and PICALM, thereby hoping to shed light on the pathological interactions between these candidate genes and brain structure and function.
Genet Epidemiol 00:1–12, 2014. © 2014 Wiley Periodicals, Inc.

KEY WORDS: adaptive smoothing; candidate genetic mapping; functional-mixed effects models; jumping surface model; likelihood ratio statistic; variance components

Introduction

Common mental and neurological disorders, such as autism and schizophrenia, are highly heritable and strongly associated with brain structure and function, but it has been difficult to unravel the genetic factors of these complex illnesses in that many genetic factors may contribute to the susceptibility of the disease while the contribution of each factor is small. Since imaging data provide the most effective measures of brain function and structure, such data may serve as an important intermediate phenotype that ultimately can lead to discoveries of genes for these complex disorders. Imaging genetic studies, which collect both imaging and genetic data,

have recently attracted extensive research interest for dissecting the genetic basis of neurological disorders [Gilmore et al., 2010; Loth et al., 2011; Savitz and Drevets, 2009]. The common and important themes of both the imaging and genetic data include ultra-high dimensionality and complex correlation structures determined by the physical location. However, most of the existing methods for genetic association studies focus on low dimensional phenotypes (e.g., case-control status) and thus cannot account for high-dimensional imaging phenotypes, while accounting for spatial smoothness of the imaging measurements. On the other hand, most existing association methods in the neuroimaging literature do not study the joint effects of multiple genetic markers, while accommodating their correlations due to linkage disequilibrium (LD).

There are two main genetic association approaches including (i) the candidate gene approaches and (ii) the genome-wide association studies (GWAS) for correlating imaging phenotypes with genotypes at one or more polymorphic markers in order to uncover genetic predispositions to disease [Amos et al., 2011; Zhu and Zhao, 2007]. Both approaches enjoy a combination of advantages and disadvantages. Candidate genes are commonly selected for study based on either

a priori knowledge of their biological functional impact on the phenotype or disease in question or previous GWAS studies, such as the NHGRI GWAS catalog [Hindorff et al., 2009]. The candidate gene approach tends to have rather high statistical power, but is incapable of discovering new genes or gene combinations. A standard statistical method for the candidate approach in imaging genetics is the voxel-wise analysis (VWA) framework. The voxel-wise analysis consists of Gaussian smoothing the imaging data and subsequently fitting a statistical model at each voxel. However, the voxel-wise analysis is generally not optimal in power and the use of Gaussian smoothing may introduce a substantial bias in the statistical results [Ball et al., 2012; Jones et al., 2005; Li et al., 2011, 2013].

Most imaging genetic studies have used the candidate gene approach, although more recently, more studies are beginning to scan the entire genome for common genetic variation [Thompson et al., 2013; Zhang et al., 2014]. A standard statistical method for GWAS in imaging genetics is the massive univariate linear modeling (MULM) framework [Hibar et al., 2011; Shen et al., 2010]. This approach repeatedly fits a linear regression model for each pair of imaging voxels and genetic markers. MULM entails a large number of comparisons, and thus MULM can only detect extremely significant imaging-marker pairs. Moreover, MULM ignores both the spatial information of the imaging data and the correlation among genetic markers. See more detailed discussions in Ge et al. [2012], Thompson et al. [2013], Vounou et al. [2010]. Recently, in Ge et al. [2012], a *cluster-wide, marker set association* framework was proposed by integrating cluster size inferences based on random field theory in order to utilize the spatial smoothness of the imaging data [Worsley et al., 2004] and a marker set analysis based on least-squares kernel machines in order to assess the joint association of potentially correlated and interacting loci [Liu et al., 2007].

The aim of this paper is to develop a functional-mixed effects modeling (FMEM) framework for the joint analysis of high-dimensional imaging data with a set of genetic markers and clinical covariates. Our FMEM is extremely useful for efficiently carrying out the candidate gene approaches in imaging genetic studies. FMEM consists of two novel components including a mixed effects model and a jumping surface model. Specifically, at each voxel, we use the mixed effects model with genetic random effects to assess the nonlinear association of potentially correlated and interacting loci with imaging phenotypes and the variance component (VC) of genetic random effects to detect the nonlinear effects of a marker set on imaging measures across voxels [Liu et al., 2007; Tzeng and Zhang, 2007; Wang and Chen, 2012]. To account for the spatial smoothness of the imaging data, we use the jumping surface model to explicitly model both the genetic variance component and fixed effects as piecewise smooth functions of the voxels with unknown edges and possible jumps. We develop a novel two-stage adaptive smoothing procedure to spatially estimate the genetic variance component function, while preserving its edges among different piecewise-smooth regions. We also develop weighted likelihood ratio tests and derive their exact approximations to test the effect of the

genetic markers across the brain. Our numerical examples show that FMEM significantly outperforms voxel-wide approaches in terms of detection of meaningful effect regions.

The rest of the paper is organized as follows. In Section 2, we describe the proposed FMEM and its adaptive estimation procedure. Then, we develop our hypothesis testing procedure to assess the genetic effects as well as the effect of clinical variables on the imaging phenotypes. In Section 3, we evaluate the finite-sample performance of FMEM by using simulation studies and analyzing a real dataset from the Alzheimer's Disease Neuroimage Initiative (ADNI). A few concluding remarks are given in the Discussion section.

Methods

Functional-Mixed Effects Model

Suppose that we observe imaging measures, clinical variables, and genetic markers from n unrelated subjects. Let $\mathcal{V}$ be the whole brain and v be a voxel in $\mathcal{V}$. For each individual i ($i = 1, \ldots, n$), an $N_V \times 1$ vector of imaging measures is observed and denoted by $Y_i = \{y_i(v) : v \in \mathcal{V}\}$. For notational simplicity, we only consider univariate imaging measures and thus, N_V equals the number of voxels in $\mathcal{V}$. Moreover, a $K \times 1$ vector of clinical covariates $\mathbf{x}_i = (x_{i1}, \ldots, x_{iK})^T$ and a $G \times 1$ vector $\mathbf{g}_i = (g_{i1}, \ldots, g_{iG})^T$ for genetic markers are also collected for each individual. For instance, imaging measures can be brain structural and functional data at each location [Ashburner and Friston, 2000; Friston, 2009], and genetic markers can be various polymorphism types, such as single nucleotide polymorphisms (SNPs), block substitutions, and copy number variants [Liu et al., 2007; Tzeng and Zhang, 2007; Wang and Chen, 2012]. The objective of this paper is to develop FMEM to quantify genetic contributions to high-dimensional imaging measures.

Our FMEM framework consists of two novel components including a mixed effects model (MEM) at each voxel and a jumping surface model (JSM) for varying coefficient functions across the brain. First, at each voxel v in $\mathcal{V}$, a mixed effects model is introduced as

$$y_i(v) = \mathbf{x}_i^T \boldsymbol{\beta}(v) + \mathbf{z}_i^T \boldsymbol{\gamma}(v) + e_i(v), \qquad (1)$$

where $\boldsymbol{\beta}(v) = (\beta_1(v), \ldots, \beta_K(v))^T$ is a $K \times 1$ vector, $\boldsymbol{\gamma}(v) \sim N(0, \sigma_\gamma^2(v)\boldsymbol{I}_L)$, $e_i(v) \sim N(0, \sigma_e^2(v))$, and $\{e_i(v) : v \in \mathcal{V}\}$ are independent across subjects i and independent of $\boldsymbol{\gamma}(v)$ for all $v \in \mathcal{V}$, in which $\boldsymbol{I}_L$ is an identity matrix. Let $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ be an $L \times n$ matrix and $\mathbf{Y}(v) = (y_1(v), \ldots, y_n(v))^T$ be an $n \times 1$ vector. Thus, under model (1), we have

$$\mathrm{Var}(\mathbf{Y}(v)) = \Sigma_Y(v) = \sigma_\gamma^2(v)\mathbf{Z}^T\mathbf{Z} + \sigma_e^2(v)\mathbf{I}_n.$$

Model (1) can be regarded as an alternative representation of the variance component models used in the literature [Kang et al., 2010; Liu et al., 2007; Tzeng and Zhang, 2007; Wang and Chen, 2012]. For instance, when $\mathbf{Z}^T\mathbf{Z}$ equals a kernel matrix $\mathbf{K} = (K(\mathbf{g}_i, \mathbf{g}_{i'}))$, model (1) reduces to the linear-mixed model considered in [Ge et al., 2012; Liu et al., 2007], where $K(\cdot, \cdot)$ is a kernel function, such as the polynomial kernel,

identity-by-state (IBS) kernel, or the Gaussian kernel. For instance, the IBS kernel is used in [Ge et al., 2012].

Model (1) has a strong connection with a nonparametric-fixed effects model given by

$$y_i(v) = \mathbf{x}_i^T \boldsymbol{\beta}(v) + h(\mathbf{g}_i; v) + e_i(v) \text{ for } i = 1, \dots, n, \quad (2)$$

where $h(\cdot; v)$ is an unknown centered function corresponding to the genetic effects at voxel v. To estimate the unknown functions $h(\cdot; v)$ in model (2), a common approach is to express them as linear combinations of some prespecified basis functions (e.g., splines or kernels), that is $h(\mathbf{g}_i; v) = \mathbf{z}_i^T \boldsymbol{\gamma}(v)$, where $\boldsymbol{\gamma}(v) = (\gamma_1(v), \dots, \gamma_L(v))^T$ is an $L \times 1$ vector for genetic random effects and $\mathbf{z}_i$ is a prespecified $L \times 1$ vector of functions of $\mathbf{g}_i$. The variation of $h(\mathbf{g}_i; v)$ is then controlled by the variation of their basis coefficients in $\boldsymbol{\gamma}(v)$. Moreover, a penalty function (e.g., L_2 or L_1) coupled with a tuning parameter is usually introduced to impose certain constraints on $\boldsymbol{\gamma}(v)$ and such a penalty function can be regarded as a prior distribution of the random effects. Thus, this connection provides a simple way to connect the nonparametric-fixed effects model (2) with model (1) [Kang et al., 2010; Liu et al., 2007; Tzeng and Zhang, 2007; Wang and Chen, 2012]. We focus on model (1) from now on.

We also assume spatial correlations as follows. For v and v' in $\mathcal{V}$, we assume

$$\text{Cov}(\boldsymbol{\gamma}(v), \boldsymbol{\gamma}(v')) = \sigma_\gamma(v)\sigma_\gamma(v')1(v = v')\boldsymbol{I}_L$$

and

$$\text{Cov}(e_i(v), e_i(v')) = \sigma_e(v)\sigma_e(v')\rho_e(v, v'),$$

where $1(\cdot)$ is an indicator function and $\rho_e(v, v')$ characterizes the spatial correlation between the measurement errors. Therefore, the covariance structure of $y_i(v)$ is given by

$$\Sigma_{y,i}(v, v') = \text{Cov}(y_i(v), y_i(v'))$$
$$= \sigma_\gamma(v)\sigma_\gamma(v')1(v = v')\mathbf{z}_i^T \mathbf{z}_i + \sigma_e(v)\sigma_e(v')\rho_e(v, v'). \quad (3)$$

Following Zhu et al. [2014], we propose a JSM for the genetic varying coefficient function $\sigma_\gamma^2(\cdot) = \{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$ and the fixed effect varying coefficient functions $\beta_j(\cdot) = \{\beta_j(v) : v \in \mathcal{V}\}$ for $j = 1, \dots, K$. For notational simplicity, we only introduce JSM for $\sigma_\gamma^2(\cdot)$ as follows:

(i) (Disjoint Partition) There is a finite and disjoint partition $\{\mathcal{V}_l : l = 1, \dots, L_\gamma\}$ of $\mathcal{V}$ such that $\mathcal{V} = \cup_{l=1}^{L_\gamma} \mathcal{V}_l$.
(ii) (Piecewise Smoothness) $\sigma_\gamma^2(v)$ is smooth within each $\mathcal{V}_l$ for $l = 1, \dots, L_\gamma$, but $\sigma_\gamma^2(v)$ is discontinuous on $\cup_{l=1}^{L_\gamma} \partial\mathcal{V}_l$, where $\partial\mathcal{V}_l$ is the boundary of $\mathcal{V}_l$ and the jumping surface of $\sigma_\gamma^2(v)$.

A similar JSM can be defined for each $\beta_j(\cdot)$ for $j = 1, \dots, K$.

We use JSM to explicitly delineate the fact that imaging data $\{y_i(v) : v \in \mathcal{V}\}$ can be regarded as a noisy version of a piecewise-smooth function of $v \in \mathcal{V}$ with the possible existence of jumps or edges. In many neuroimaging datasets, those jumps or edges often reflect the functional and/or

structural changes, such as white matter (WM) and grey matter (GM), across the brain. Therefore, the varying coefficient functions in model (1) may inherit the piecewise-smooth features from the imaging data. Furthermore, it is more reasonable to assume that different varying coefficient functions have different jumps or edges, since they may play different roles in characterizing the piecewise-smooth pattern of the imaging data.

Our FMEM consisting of model (1) and JSM can be regarded as a novel extension of the existing FMEMs and varying coefficient models in the literature [Greven et al., 2010; Guo, 2002; Morris and Carroll, 2006; Yuan et al., 2014; Zhu et al., 2011, 2012, 2014], even though all of them are developed to model functional responses (of time or voxel) measured either cross-sectionally or longitudinally. However, at each voxel, the jumping surface model is introduced to spatially characterize the piecewise smoothness of the imaging data. In contrast, most existing FMEMs and varying coefficient models reduce to a parametric model at each voxel, while most FMEMs do not explicitly model the piecewise smoothness of the imaging data. Moreover, the primary interest of our FMEM is to estimate the genetic varying coefficient function $\sigma_\gamma^2(\cdot)$, whereas that of other FMEMs and varying coefficient models is to estimate the fixed effects varying coefficient functions $\beta_j(\cdot)$. Estimating the function $\sigma_\gamma^2(\cdot)$ is computationally and theoretically much harder than estimating the varying coefficient functions, since $\sigma_\gamma^2(\cdot)$ satisfies a nonnegative constraint.

Two-Stage Estimation Procedure

We propose a two-stage estimation procedure to estimate all varying coefficient functions and test their effects on imaging phenotypes. The key ideas of each stage are given as follows:

Stage I. Estimate $\sigma_e^2(\cdot)$, develop an adaptive smoothing method to estimate $\sigma_\gamma^2(\cdot)$, and test the null hypothesis $\sigma_\gamma^2(v) = 0$ across all voxels.
Stage II. Develop an adaptive smoothing method to spatially and adaptively estimate $\boldsymbol{\beta}(\cdot) = \{\boldsymbol{\beta}(v) : v \in \mathcal{V}\}$ and then test associated hypotheses.

Moreover, after calculating $\hat{\sigma}_e^2(\cdot)$, $\hat{\sigma}_\gamma^2(\cdot)$ and $\hat{\boldsymbol{\beta}}(\cdot)$, we can estimate $\hat{\boldsymbol{\gamma}}(v) = \mathbf{Z}\hat{\Sigma}_Y(v)^{-1}\{\mathbf{Y}(v) - \mathbf{X}^T\hat{\boldsymbol{\beta}}(v)\}$, where $\hat{\Sigma}_Y(v) = \hat{\sigma}_\gamma^2(v)\mathbf{Z}^T\mathbf{Z} + \hat{\sigma}_e^2(v)\mathbf{I}_n$ and $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a $p \times n$ matrix. To approximate $\rho_e(v, v')$, we calculate the empirical correlation between $\{\hat{e}_i(v) : i = 1, \dots, n\}$ and $\{\hat{e}_i(v') : i = 1, \dots, n\}$, where $\hat{e}_i(v) = y_i(v) - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}(v) + \mathbf{z}_i^T\hat{\boldsymbol{\gamma}}(v)$. Since $\boldsymbol{\gamma}(v)$ and $\rho_e(v, v')$ are nuisance parameters, we do not focus on them throughout. Since our primary interest lies in the genetic effect, we focus on Stage I and only briefly discuss Stage II for the sake of space.

The key novelty of our estimation procedure is the adaptive smoothing method in Stage I for smoothing $\sigma_\gamma^2(\cdot)$. Since the true variance components $\sigma_\gamma(v)$ can be zero in some regions of interest and their estimates are always nonnegative, directly applying standard smoothing methods, such as

splines or kernel smoothing, to these nonnegative variance component estimates can introduce substantial bias in the estimation of functional genetic variance components. Thus, most existing smoothing methods cannot be successfully used in such smoothing problems [Greven et al., 2010; Guo, 2002; Morris and Carroll, 2006; Polzehl and Spokoiny, 2000; Polzehl et al., 2010; Yuan et al., 2014; Yue et al., 2010; Zhu et al., 2011, 2012, 2014].

Stage I

The first stage consists of three major steps as follows:

Step I.1. Calculate the restricted maximum likelihood (REML) estimator of $\eta(v) = (\sigma_\gamma^2(v), \sigma_e^2(v))$ across all voxels $v \in \mathcal{V}$.

Step I.2. Spatially and adaptively reestimate $\{\sigma_\gamma^2(v) : v \in V\}$ by incorporating information from neighboring voxels.

Step I.3. Construct weighted likelihood ratio statistics and derive their approximate distributions to test the null hypothesis of $H_0(v) : \sigma_\gamma^2(v) = 0$ across all voxels.

In Step I.1, we calculate the REML estimator of $\eta(v)$ across voxels. There exists an $(n-p) \times n$ matrix K_x such that $K_x \mathbf{X}^T = \mathbf{0}$ and $\text{rank}(K_x) = n - p$. A mixed effects model for $\mathbf{Y}^*(v) = K_x \mathbf{Y}(v)$ is given by

$$\mathbf{Y}^*(v) = K_x \mathbf{Z}^T \boldsymbol{\gamma}(v) + K_x E(v), \qquad (4)$$

where $E(v) = (e_1(v), \ldots, e_n(v))^T$. Based on the distributional assumptions in (1), we have $\mathbf{Y}^*(v) \sim N(\mathbf{0}, \Sigma_{Y^*}(v))$, where $\Sigma_{Y^*}(v) = \sigma_\gamma^2(v) K_x \mathbf{Z}^T \mathbf{Z} K_x^T + \sigma_e^2(v) K_x K_x^T$. Thus, at each voxel v, the REML estimate of $\eta(v)$, denoted by $\hat{\eta}(v) = (\hat{\sigma}_\gamma^2(v), \hat{\sigma}_e^2(v))^T$, is to maximize the REML function given by

$$\ell_{REML}(\mathbf{Y}^*(v)|\mathbf{Z}, \eta(v))$$
$$= -0.5 \log |\Sigma_{Y^*}(v)| - 0.5 \mathbf{Y}^*(v)^T \Sigma_{Y^*}(v)^{-1} \mathbf{Y}^*(v). \quad (5)$$

Since our primary interest lies on $\sigma_\gamma^2(v)$, we fix $\sigma_e^2(v)$ as $\hat{\sigma}_e^2(v)$ from here on.

In Step I.2, we construct a weighted REML function to estimate $\sigma_\gamma^2(v)$ by incorporating the spatial information in a neighborhood $B(v, h)$ for each voxel v with a specific radius h as follows:

$$L_{REML}(\sigma_\gamma^2(v)|\mathbf{Y}^*, B(v, h))$$
$$= \sum_{v' \in B(v,h)} \omega_\gamma(v, v'; h) \ell_{REML}(\mathbf{Y}^*(v')|\mathbf{Z}, \sigma_\gamma^2(v), \hat{\sigma}_e^2(v')),$$

$$(6)$$

where $\omega_\gamma(v, v'; h)$ is a weight function of voxels v, v', and the radius h. We maximize $L_{REML}(\sigma_\gamma^2(v)|\mathbf{Y}^*, B(v, h))$ in order to calculate the weighted REML estimator of $\sigma_\gamma^2(v)$, denoted by $\hat{\sigma}_\gamma^2(v, h)$. The weight function $\omega_\gamma(v, v'; h)$ measures the data similarity between the two voxels v and v' such that $\sum_{v' \in B(v,h)} \omega_\gamma(v, v'; h) = 1$ and $\omega_\gamma(v, v'; h) \geq 0$. A large value of $\omega_\gamma(v, v'; h)$ means that the information contained in the

voxels v and v' is very similar, whereas $\omega_\gamma(v, v'; h) \approx 0$ indicates that the data in voxel v' do not have too much information for $\sigma_\gamma(v)$. The adaptive weight $\omega_\gamma(v, v'; h)$ plays a critical role in preventing over-smoothed estimation of $\sigma_\gamma^2(v)$ and in preserving the edges of significant regions of $\{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$.

In Step I.3, to assess the genetic effects on imaging phenotypes across all voxels, we formulate it as testing the following null and alternative hypotheses:

$$H_{0,\gamma}(v) : \sigma_\gamma^2(v) = 0 \quad \text{vs.} \quad H_{1,\gamma}(v) : \sigma_\gamma^2(v) > 0. \quad (7)$$

We test (7) by using the weighted REML ratio statistic defined by

$$\text{RLRT}_{\sigma_\gamma^2}(v) = 2\big\{ L_{REML}\big(\hat{\sigma}_\gamma^2(v)|\mathbf{Y}^*, B(v, h)\big)$$
$$- L_{REML}(0|\mathbf{Y}^*, B(v, h))\big\}. \quad (8)$$

Since all the subjects share the same random effect $\gamma(v)$, the standard asymptotic results in Stram and Lee [1994] are invalid and can perform very poorly even for the unweighted REML ratio statistics for testing random effects in model (1). However, we provide an exact null distribution of $\text{RLRT}_{\sigma_\gamma^2}(v)$ below.

Step I.2: Adaptive Estimation of $\sigma_\gamma^2(v)$

Following the adaptive estimation (AET) procedure proposed in Li et al. [2011], we adaptively determine $\{\omega_\gamma(v, v'; h) : v, v' \in \mathcal{V}\}$ and then calculate $\hat{\sigma}_\gamma(v; h)$ as h increases from $h_0 = 0$ to a predetermined value $h_S = r_0$. The key novelty of AET is to build a sequence of $L_{REML}(\sigma_\gamma^2(v)|\mathbf{Y}^*, B(v, h))$ for $h_0 = 0 < h_1 < \cdots < h_S = r_0$ at each voxel $v \in \mathcal{V}$ and then sequentially determine $\omega_\gamma(v, v'; h_s)$ for all $v' \in B(v, h_s)$ based on $\{\hat{\sigma}_\gamma^2(v', h_{s-1}) : v' \in B(v, h_s)\}$ for all $v \in \mathcal{V}$ and $s = 1, \ldots, S$. However, one cannot apply many existing smoothing methods, such as local kernel or the propagation-seperation method [Fan and Gijbels, 1996; Polzehl and Spokoiny, 2000; Polzehl et al., 2010; Zhu et al., 2014], to directly smooth $\{\hat{\sigma_\gamma^2}(v) : v \in \mathcal{V}\}$ in the nonactivation region $\mathcal{V}_{NA} = \{v : \sigma_\gamma(v) = 0, \ v \in \mathcal{V}\}$. Specifically, since $\hat{\sigma_\gamma^2}(v)$ is always nonnegative even in the voxels of $\mathcal{V}_{NA}$, directly calculating the weighted means of $\hat{\sigma_\gamma^2}(v)$ does not lead to the bias reduction in $\mathcal{V}_{NA}$. A path diagram of AET is given as follows:

$$
\begin{array}{ccccc}
h_0 = 0 & < & h_1 & < \cdots < & h_S = r_0 \\
B(v, h_0) = \{v\} & \subset & B(v, h_1) & \subset \cdots \subset & B(v, h_S) \\
\Downarrow & & \Downarrow & \nearrow \cdots \nearrow & \Downarrow \\
\big\{\hat{\sigma}_\gamma^2(v) : v \in \mathcal{V}\big\} \Rightarrow & & \omega_\gamma(v, v'; h_1) & \cdots & \omega_\gamma(v, v'; h_S = r_0) \\
& & \Downarrow & \nearrow \cdots \nearrow & \Downarrow \\
& & \big\{\hat{\sigma}_\gamma^2(v; h_1) : v \in \mathcal{V}\big\} & \cdots & \big\{\hat{\sigma}_\gamma^2(v; h_S) : v \in \mathcal{V}\big\}.
\end{array}
$$

The three key steps of AET, including weight adaptation, estimation, and termination checking, are presented as follows.

- In the weight *adaption* step (i), we select a series $\{h_s = c_h^s : s = 1, \ldots, S\}$ of radii with $c_h \in (1, 2)$, say $c_h = 1.125$. We use a relatively small c_h in order to increase estimation

robustness and prevent oversmoothing. We then set $s = 1$ and $h_1 = c_h$. The adaptive weights in (6) are given by

$$\omega_\gamma(v, v'; h_s) = K_{loc}(||v - v'||_2/h_s)K_{st}(D_\gamma(v, v'; h_{s-1})/C_n),$$

(9)

where $K_{loc}(u) = (1 - u)_+$ and $K_{st}(u) = \min(1, 2(1 - u^2))_+$, and $|| \cdot ||_2$ denotes the Euclidean norm of a vector (or a matrix). Moreover, $D_\gamma(v, v'; h_{s-1})$ is set as $\{\hat{\sigma}_\gamma^2(v; h_{s-1}) - \hat{\sigma}_\gamma^2(v'; h_{s-1})\}^2/\text{var}[\widehat{\sigma_\gamma^2(v)}]$ so that the difference between consecutive $\hat{\sigma}_\gamma^2(v; h_{s-1})$s is within the precision of $\sqrt{\text{var}[\widehat{\sigma_\gamma^2(v)}]}$, where $\text{var}[\widehat{\sigma_\gamma^2(v)}]$ is estimated by using the inverse of the Fisher information matrix of $(\sigma_\gamma^2(v), \sigma_e^2(v))$ from the likelihood function (5) with $h = h_0$. We choose $C_n = n^{1/3}\chi^2(1)^{0.5}$ for $D_\gamma(v, v'; h_{s-1})$ defined in (9), where $\chi^2(1)^{0.5}$ is the 0.5-percentile of the $\chi^2(1)$ distribution. These quantities are fixed for subsequent updates of h_s.

The rationale for choosing different tuning parameters given above is given as follows. The weight $K_{st}(D_\gamma(v, v'; h_{s-1})/C_n)$ downweights the role of a voxel $v' \in B(v, h_s)$ in the REML functions if $D_\gamma(v, v'; h_{s-1})$ is large. The weight $K_{loc}(||v - v'||_2/h_s)$ gives less weight to the voxel $v' \in B(v, h_s)$, whose location is far from the voxel v. The scale C_n is used to penalize the similarity between any two voxels v and v' in a similar manner to bandwidth, and an appropriate choice of C_n is crucial for the behavior of the adaptive smoothing method in Stage I. As discussed in Zhu et al. [2014] and Li et al. [2011], C_n should satisfy $C_n/n = o(1)$ and $C_n^{-1} \log(N_V) = o(1)$.

- In the *estimation* step (ii), for each $v \in \mathcal{V}$ and for the radius h_s given $\omega_\gamma(v, v'; h_s)$, we calculate $\hat{\sigma}_\gamma(v; h_s)$ by maximizing $L_{REML}(\sigma_\gamma^2(v)|\mathbf{Y}^*, B(v, h))$ defined in equation (6).
- In the *termination checking* step (iii), after the S_0–th iteration, we calculate a stopping criterion based on a distance between $\hat{\sigma}_\gamma^2(v; h_{S_0})$ and $\hat{\sigma}_\gamma^2(v; h_s)$ given by

$$D\left(\hat{\sigma}_\gamma^2(v; h_{S_0}), \hat{\sigma}_\gamma^2(v; h_s)\right)$$

$$= \left\{\hat{\sigma}_\gamma^2(v; h_{S_0}) - \hat{\sigma}_\gamma^2(v; h_s)\right\}^2 var[\widehat{\sigma_\gamma^2(v)}]^{-1} \quad (10)$$

for $s > S_0$. Then, we compare $D(\hat{\sigma}_\gamma^2(v; h_{S_0}), \hat{\sigma}_\gamma^2(v; h_s))$ with a benchmark, denoted by $\tilde{C}(s)$, for $s > S_0$. If $D(\hat{\sigma}_\gamma^2(v; h_{S_0}), \hat{\sigma}_\gamma^2(v; h_s)) > \tilde{C}(s)$, then we set $\hat{\sigma}_\gamma^2(v) = \hat{\sigma}_\gamma^2(v, h_{s-1})$ and the estimation for this voxel v is terminated. If $s = S$ and $D(\hat{\sigma}_\gamma^2(v; h_{S_0}), \hat{\sigma}_\gamma^2(v; h_s)) \leq \tilde{C}(s)$, $\hat{\sigma}_\gamma^2(v)$ is set as $\hat{\sigma}_\gamma^2(v, h_S)$ and the estimation process terminates. The algorithm stops when the estimation is finished for all v in $\mathcal{V}$. If $s \leq S_0$ or $D(\hat{\sigma}_\gamma^2(v; h_{S_0}), \hat{\sigma}_\gamma^2(v; h_s)) \leq \tilde{C}(s)$ for $s < S_0 \leq S - 1$, then we go back to the weight adaptation step (i) with an increased radius $h = h_{s+1} = c_h^{s+1}$. Throughout the paper, we set $S_0 = 2$, $\tilde{C}(s) = \chi^2(p)^{0.7/(s-1)}$, and $S = 10$. Note that $\tilde{C}(s)$ is a decreasing function in s that makes the stopping criteria more and more stringent when the radius increases in order to prevent oversmoothing.

Step I.3: Testing $H_{0,\gamma}(v) : \sigma_\gamma^2(v) = 0$

We perform hypothesis testing in (7) by using the test statistics $\text{RLRT}_{\sigma_\gamma^2}(v)$ and their corresponding P-values. Let $\Omega = K_x\mathbf{Z}^T\mathbf{Z}K_x^T = UD_0U^T$ be the spectral decomposition of Ω such that $D_0 = \text{diag}(d_1, \ldots, d_{n-p})$ is the diagonal matrix of eigenvalues d_k and U is an $(n-p) \times (n-p)$ orthonormal matrix. Without loss of generality, we choose K_x such that $K_xK_x^T = \mathbf{I}_{n-p}$. We obtain the following theorem, whose proof is included in the supplementary document.

Theorem 1. *Under model* (1), *$RLRT_n(v)$ can be written as*

$$RLRT_n(v) = \sup_{\lambda(v) \geq 0} \left\{ \sum_{v' \in B(v,h)} \omega(v, v'; h)D(v'; \lambda(v)/\sigma_e^2(v')) \right\},$$

(11)

where $\lambda(v) = \sigma_\gamma^2(v)$ and $D(v'; t)$ is given by

$$\sigma_e^{-2}(v')\mathbf{Y}^{*T}(v') U diag\left(\frac{td_1}{1 + td_1}, \ldots, \frac{td_{n-p}}{1 + td_{n-p}}\right) U^T\mathbf{Y}^*(v')$$

$$- \sum_{l=1}^{n-p} \log(1 + td_l).$$

(12)

Moreover, under the null hypothesis $H_{0,\gamma}(v)$, we have

$$D(v'; t) \overset{\mathcal{D}}{=} \sum_{l=1}^{n-p} \frac{\delta_l^2(v')td_l}{1 + td_l} - \sum_{l=1}^{n-p} \log(1 + td_l), \quad (13)$$

where $\overset{\mathcal{D}}{=}$ means equality in distribution and the $\delta_l(v)$'s are i.i.d $N(0, 1)$ random variables.

Although Theorem 1 provides an efficient way of approximating the null distribution of $\text{RLRT}_n(v)$, a complex issue arises from the complex spatial correlations among the $\delta_l^2(v')$ across voxels $v' \in B(v, h)$. One approach for dealing with such an issue is to estimate the spatial correlation for any pair of voxels, which can be computationally intensive. To avoid calculating spatial correlations, we develop a wild bootstrap method to efficiently simulate the null finite sample distribution of $\text{RLRT}_n(v)$. The detailed steps of this bootstrap method are presented in the supplementary document. After the P-values for all voxels $v \in \mathcal{V}$ are computed, either a false discovery rate (FDR) method or random field theory (RFT) is applied to correct for multiple comparisons [Ge et al., 2012].

Stage II

The second stage is to estimate $\boldsymbol{\beta}(v)$ and carry out statistical inference on $\boldsymbol{\beta}(v)$. At each voxel v, we consider model (1) given by

$$\mathbf{Y}(v) = \mathbf{X}\boldsymbol{\beta}(v) + \mathbf{Z}^T\boldsymbol{\gamma}(v) + E(v). \quad (14)$$

After calculating $\{(\hat{\sigma}_\gamma(v; h_S), \hat{\sigma}_e(v)) : v \in \mathcal{V}\}$, we can calculate $\hat{\Sigma}_Y(v) = \hat{\sigma}_\gamma^2(v; h_S)\mathbf{Z}^T\mathbf{Z} + \hat{\sigma}_e^2(v)\mathbf{I}_n$. Since all components of $\boldsymbol{\beta}(v)$ and $\Sigma_Y(v)$ are statistically orthogonal to each other, we fix $\Sigma_Y(v)$ at $\hat{\Sigma}_Y(v)$ from here on. Since the true $\boldsymbol{\beta}(v)$ is not on the boundary of the parameter space, different adaptive smoothing methods can be used here [Polzehl and Spokoiny,
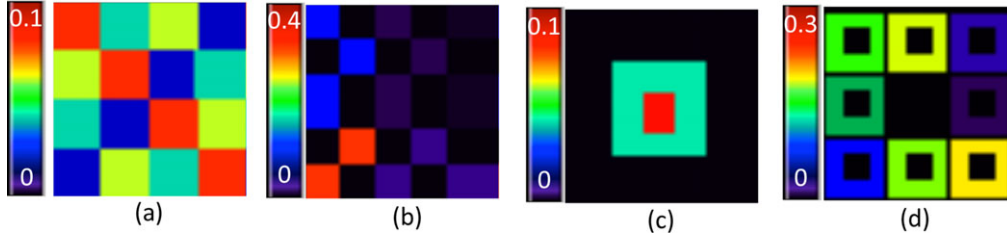
Figure 1. Simulation setting: (a) True image of β_0; (b) true image of β_1, in which the colors represent the values of $\beta_1(v) \times 10^4$; (c) true image of β_2; and (d) true image of β_3.

2000; Polzehl et al., 2010; Zhu et al., 2014]. For simplicity, we put the detailed steps of Stage II in the supplementary document.
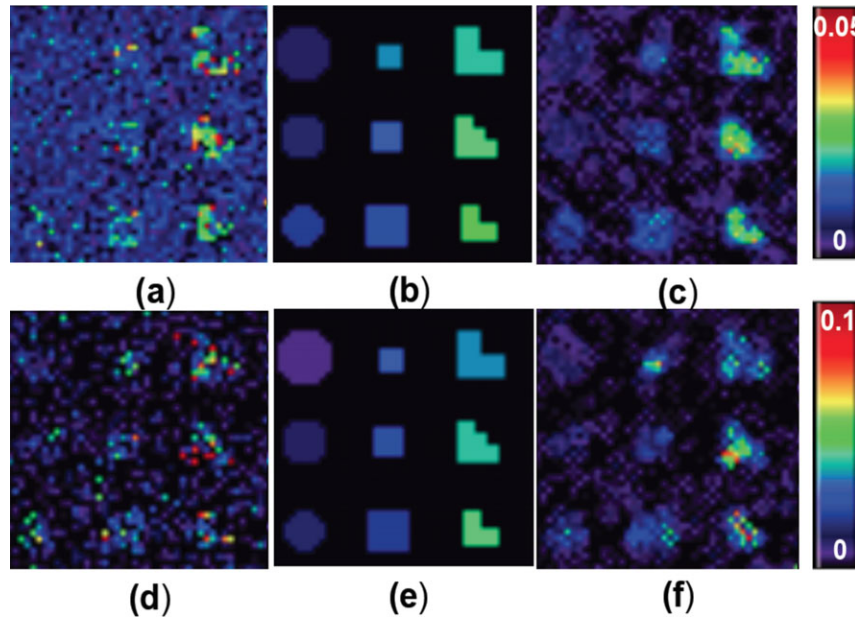
Results

Simulation Studies

We simulated data at all $N_V = 5,808$ voxels on a $44 \times 44 \times 3$ phantom image. Each z-slice contains the same effect regions. At each voxel, we simulated the univariate imaging measures according to model (1) with $\boldsymbol{\beta}(v) = (\beta_0(v), \beta_1(v), \beta_2(v), \beta_3(v))^T$ and $\boldsymbol{x}_i = (1, x_{i1}, x_{i2}, x_{i3})^T$. Moreover, the covariates x_{i1}, x_{i2}, and x_{i3} were generated from a Gaussian distribution with mean 40 and standard deviation 10, a Bernoulli distribution with success probability 0.5, and a Bernoulli distribution with success probability 0.3, respectively. These three covariates were de-signed to mimic the common clinical variables: age, gender, and disease status. For a slice of the phantom image, the effect areas for $\beta_0(v)$ were divided into 16 regions with four different values ranging from 0.02 to 0.08, increasing by 0.02 (Fig. 1a); for $\beta_1(v)$, the effect regions were divided into 25 regions ranging from $10^{-2.5}$ to $10^{-12.5}$, decreasing by a rate of $10^{-2.5}$ (Fig. 1b); for $\beta_2(v)$, the whole space was separated into three regions with values 0, 0.05, and 0.1 (Fig. 1c); the effect area of $\beta_3(v)$ was divided into nine regions with values ranging from 0 to 0.1, increasing by differences of 0.025 (Fig. 1d).

The genetic information was simulated according to the SNP data obtained from the public accessible data of the Alzheimer's Disease Neuroimage Initiative (ADNI). It is an ongoing longitudinal study with the primary purpose of exploring the genetic and neuroimaing information associated with late-onset Alzheimer's disease (LOAD). The study recruited elderly subjects older than 65 years of age consisting about 400 subjects with mild cognitive impairment

Figure 2. Simulation results for estimation accuracy: Scenario I: (a) estimated $\sigma_\gamma^2(v)$ by using voxel-wise approach; (b) true $\sigma_\gamma^2(v)$ image; and (c) estimated $\sigma_\gamma^2(v)$ by using FMEM. Scenario II: (d) estimated $\sigma_\gamma^2(v)$ by using voxel-wise approach; (e) true $\sigma_\gamma^2(v)$ image; (f) estimated $\sigma_\gamma^2(v)$ by using FMEM.

(MCI), about 200 subjects with Alzheimer's disease (AD), and around 200 healthy controls. Each subject was followed for at least 3 years. During the study period, the subjects were assessed with magnetic resonance imaging (MRI) measures and psychiatric evaluation to determine the diagnosis status at each time point. Genetic information was also collected from each subject at baseline and is genotyped by the Illumina 610 Quad array with more than 620,000 single nucleotide polymorphisms (SNPs). More information on ADNI is provided in the real data analysis of Section 3.2. We simulated the genetic information based on the two following scenarios:

- Scenario I. To preserve the linkage disequilibrium among SNPs, we utilize all of the SNPs on chromosome 1 from 197 Caucasian controls to generate the genetic effects. After eliminating the SNPs with minor allele frequency (MAF) less than 5%, there were 31,554 out of 45,627 SNPs left. Then, we randomly chose 20 SNPs and 100 subjects among the 197 healthy controls as the simulated genetic data z_i in (1). In this case, $n = 100$. If any of these 20 SNPs have MAF less than 5%, then the genetic data were resampled until all of the 20 SNPs have MAF $\geq$ 5%.
- Scenario II. To evaluate the performance of FMEM in the case of high LD, we selected the SNPs from the same gene in the second scenario. Searching the SNPs on the gene PICALM, which has been found to be relevant for Alzheimer's disease in many studies [Harold et al., 2009] using the gene list "glist-hg18" provided by PLINK, there were 23 SNPs on PICALM with MAF larger than 5%. After eliminating the missing values, there are 176 healthy controls with complete genotype data at these 23 SNPs. We randomly selected 7 SNPs from 75 healthy controls to be z_i in (1). Although there is strong LD among these 7 SNPs, no SNP has perfect correlation (1 or -1) with any other SNPs in these 75 subjects. In this case, $n = 75$.

In both scenarios, the SNP effects were assumed to be additive. The $\gamma(v)$'s were generated from a multivariate normal distribution with mean zero and covariance matrix $\sigma_\gamma^2(v)\boldsymbol{I}_L$. Different $\sigma_\gamma^2(v)$ values, which represent different signal-to-noise ratios, were chosen to examine the finite sample performance of our method at different signal-to-noise ratios and also to test whether FMEM can perform well for different shapes. See Figures 3b and 3e for Scenarios I and II. Moreover, we overlay some of the effect areas of $\beta_3(v)$ and $\sigma_\gamma^2(v)$ in order to account for the fact that the brain phenotype is an intermediate expression of disease progression. The $\{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$ of the effect regions in Scenario I ranged from 0.005 to 0.025, increasing by 0.0025, whereas the $\{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$ of effect regions in Scenario II ranged from 0.005 to 0.045, increasing by 0.005. The random error $e_i(v)$ was independently distributed as a univariate normal distribution with mean 0 and standard deviation 3 for all voxels. We set the number of bootstrap samples M and the number of repetitions to be 200.

Tables 1 and 2 summarize the estimation results of $\sigma_\gamma^2(v)$ obtained from FMEM and the voxel-wise method for both scenarios. The tables include the average absolute value of the bias, the root mean square error (RMS), standard devia-

Table 1. Simulation results: the estimation results of $\sigma_\gamma^2(v)$ in *Scenario I* using FMEM and the voxel-wise method in terms of average absolute value of bias (BIAS), root mean square error (RMS), standard deviation (SD), and the ratio between RMS and SD (RE)
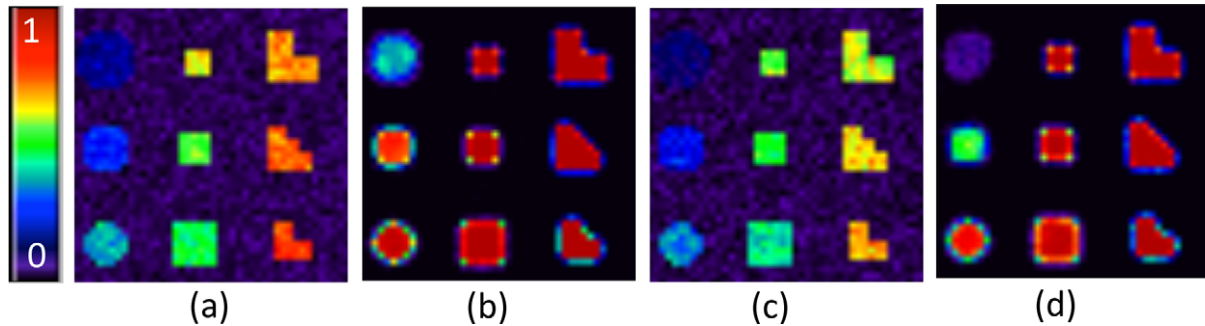
	FMEM				Voxel-wise			
$\sigma_\gamma^2(v)$	|BIAS|	RMS	SD	RE	|BIAS|	RMS	SD	RE
0	0.001	0.002	0.002	1	0.007	0.005	0.005	1
0.005	2.36e-06	0.003	0.003	1	0.005	0.005	0.005	1
0.0075	0.0005	0.003	0.003	1	0.005	0.006	0.006	1
0.01	0.001	0.004	0.004	1	0.006	0.008	0.008	1
0.0125	0.001	0.004	0.004	1	0.006	0.008	0.008	1
0.015	0.002	0.005	0.005	1	0.008	0.010	0.010	1
0.0175	0.003	0.005	0.005	1	0.008	0.010	0.010	1
0.020	0.002	0.006	0.006	1	0.010	0.012	0.012	1
0.0225	0.003	0.006	0.006	1	0.010	0.013	0.013	1
0.025	0.004	0.006	0.006	1	0.010	0.014	0.014	1

Table 2. Simulation results: the estimation results of $\sigma_\gamma^2(v)$ in *Scenario II* by using FMEM and the voxel-wise method in terms of average absolute value of bias (BIAS), root mean square error (RMS), standard deviation (SD), and the ratio between RMS and SD (RE)

	FMEM				Voxel-wise			
$\sigma_\gamma^2(v)$	|BIAS|	RMS	SD	RE	|BIAS|	RMS	SD	RE
0	0.002	0.003	0.003	1	0.003	0.008	0.008	1
0.005	0.0001	0.004	0.004	1	0.007	0.126	0.126	1
0.010	0.001	0.006	0.006	1	0.011	0.016	0.016	1
0.015	0.002	0.008	0.008	1	0.014	0.020	0.02	1
0.020	0.002	0.010	0.010	1	0.024	0.017	0.017	1
0.025	0.003	0.020	0.020	1	0.020	0.029	0.029	1
0.030	0.005	0.013	0.013	1	0.023	0.032	0.032	1
0.035	0.004	0.014	0.014	1	0.026	0.035	0.035	1
0.040	0.015	0.006	0.006	1	0.028	0.040	0.040	1
0.045	0.007	0.016	0.016	1	0.031	0.040	0.040	1

tion (SD), and the ratio of RMS to SD. RMS is based on the empirical mean and the SD is based on the theoretical mean. As shown in both tables, FMEM produces smaller estimation bias, RMS, and SD compared with the voxel-wise method, which indicates that FMEM yields much more accurate estimation. Figure 2 presents the true $\{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$ image and the estimated $\{\sigma_\gamma^2(v) : v \in \mathcal{V}\}$ images by using the voxel-wise approach and FMEM. It is clear that FMEM outperforms the voxel-wise approach.

We tested the hypotheses $H_0 : \sigma_\gamma^2(v) = 0$ and $H_1 : \sigma_\gamma^2(v) \neq 0$ for all voxels in $\mathcal{V}$ based on FMEM and its corresponding voxel-wise method, and the score test based on the IBS kernel used in Ge et al. [2012]. Moreover, we evaluated their finite-sample performance in cluster-based thresholding [Silver et al., 2011]. Specifically, we first thresholded the P-values for all voxels in $\mathcal{V}$ by using an initial P-value of 0.01 as suggested by Silver et al. [2011] in order to identify clusters of contiguous supra-threshold voxels. Then, the thresholded clusters were matched with the nine true effect areas in Figure 3b or 3e. If a specific thresholded cluster overlaps with at least one voxel in any of the nine true effect regions, we

Figure 3. Simulation results for testing the genetic effects: Scenario I: the rejection rate image at a selected slice by using (a) voxel-wise approach and (b) FMEM; Scenario II: the rejection rate image by using (c) voxel-wise approach and (d) FMEM.

Table 3. Simulation results: comparisons of FMEM, voxel-wise method, and the score test based on the IBS kernel (Score) for *Scenario I*: the dice overlap ratio (DOR), average number of false-positive clusters, and average size of false-positive clusters with different cluster size thresholds

		Scenario I					
		FMEM		Voxel-wise		Score	
Threshold		Mean	SD	Mean	SD	Mean	SD
Voxel size = 1	DOR	0.94	0.05	0.99	0.02	0	0
	False-positive cluster number	1.88	6.12	21.30	12.29	0	0
	False-positive cluster size	1.03	0.04	1.06	0.06	NA	NA
Voxel size = 10	DOR	0.91	0.04	0.83	0.10	0	0
	False-positive cluster number	0	0	0	0	0	0
	False-positive cluster size	NA	NA	NA	NA	NA	NA

Table 4. Simulation results: comparisons of FMEM, voxel-wise method, and the score test based on the IBS kernel (Score) for *Scenario II*: the dice overlap ratio (DOR), average number of false-positive clusters, and average size of false-positive clusters for *Scenario II* with different cluster size thresholds

		Scenario II					
		FMEM		Voxel-wise		Score	
Threshold		Mean	SD	Mean	SD	Mean	SD
Voxel size = 1	DOR	0.86	0.06	0.996	0.02	0.49	0.21
	False-positive cluster number	1.35	4.85	15.45	12.92	0	0
	False-positive cluster size	1.07	0.08	1.05	0.07	NA	NA
Voxel size = 10	DOR	0.85	0.07	0.78	0.11	0.01	0.03
	False-positive cluster number	0	0	0	0	0	0
	False-positive cluster size	NA	NA	NA	NA	NA	NA

call such a cluster a "true positive." In contrast, if a specific thresholded cluster does not overlap with any voxels of the nine effect regions, we call such a cluster a "false positive." We summarized the hypothesis testing results by the average dice overlap ratio (DOR), the average number of false-positive clusters, and the average size in the number of voxels of false-positive clusters. DOR is the ratio between the number of true positive clusters over the true number of effect areas, which is nine in this simulation setting. Thus, the higher DOR means the higher the probability of detecting the true effect regions.

As shown in Tables 3 and 4, if we set the cluster size threshold as one voxel, FMEM has smaller DOR and a smaller number of false-positive clusters compared with the voxel-wise method. When the cluster size threshold increases to 10 voxels, FMEM has a similar DOR value as that of the no threshold case, whereas the DOR of the voxel-wise approach reduces by about 20%. The score test based on the IBS kernel has little power to detect the nine effect regions with subtle effects. It may be caused by both the relatively low sensitivity of the score test itself and the misspecified IBS kernel.

Table 5 summarizes the number of significant voxels identified by the three methods in each effect region of Scenarios I and II. In Table 5, FMEM identifies less voxels in the noneffect regions, while detecting more voxels in the effect regions in both scenarios. For FMEM, its computational time is around

30 min for the first scenario and 20 min for the second one. Table 5 also confirms that the score test based on the IBS kernel has little statistical power in detecting the nine effect regions. Therefore, FMEM significantly outperforms the voxel-wise method and the score test based on the IBS kernel [Ge et al., 2012] in terms of detecting the true effect regions and controlling the false-positive error rate.

ADNI Data Analysis

The ADNI study began in 2004 and has three phases thus far, including ADNI-1, ADNI GO, and ADNI-2. The overall objective of the ADNI study is to determine the relationships among the clinical, cognitive, imaging, genetic, and biochemical biomarker characteristics of the entire spectrum of AD as the pathology evolves from normal aging through early mild cognitive impairment, to mild cognitive impairment, to late mild cognitive impairment, to dementia or AD. The Principal Investigator of this initiative is Dr. Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. For up-to-date information, see www.adni-info.org for details.

The aim of this ADNI data analysis is to use FMEM to identify brain regions affected by candidate genes, thereby hoping to shed light on the pathological interactions between these candidate genes and brain structure and function. The

Table 5. ADNI data analysis: global power calculation for number of significant voxels detected by voxel-wise approach and FMEM in both scenarios

Number of voxels	4749	207	135	111	147	75	48	144	111	81
Scenario I										
$\sigma_\gamma^2(v)$	0	0.005	0.0075	0.010	0.0125	0.015	0.0175	0.020	0.0225	0.025
Score	0	0	0	0	0	0	0	0	0	0
Voxel-wise	123.91	36.3	38.25	43.08	72.245	42.98	31.535	104.32	86.1	65.97
FMEM	68.43	50.52	77.32	83.09	133.7	69.51	45.57	142.26	110.07	80.54
Scenario II										
$\sigma_\gamma^2(v)$	0	0.005	0.010	0.015	0.020	0.025	0.030	0.035	0.040	0.045
Score	0.165	0.905	2.475	5.11	11.065	7.91	6.715	24.135	21.655	17.81
Voxel-wise	110.93	27.08	33.105	40.05	65.92	39.06	27.51	90.69	73.42	56.59
FMEM	76.57	8.11	46.05	71.82	124.42	66.15	43.182	139.51	108.35	79.49

data we employed evaluates the performance of FMEM from ADNI-1. About 800 subjects with age older than 65 were recruited and followed at least 3 years. The 800 subjects included 200 healthy controls, 400 subjects with different levels of mild cognitive impairment (MCI), and 200 subjects with Alzheimer's disease (AD). Besides the SNPs and the T1 weighted MRI imaging measurements, the subjects were assessed with demographic information and psychiatric examination scores to determine the diagnosis status at each scheduled visit.

The raw MRI data were collected from a variety of 1.5 Tesla MRI scanners with protocols individualized for each scanner, including standard T1-weighted images obtained using volumetric 3-dimensional sagittal MPRAGE or equivalent protocols with varying resolutions. The typical protocol included: repetition time (TR) = 2,400 ms, inversion time (TI) = 1,000 ms, flip angle = 80, field of view (FOV) = 24 cm, with a $256 \times 256 \times 170$ acquisition matrix in the x- y-, and z-dimensions yielding a voxel size of $1.25 \times 1.26 \times 1.2$ mm^3.

The T1-weighted MRI images were preprocessed by standard image processing steps including AC (anterior commissure) and -PC (posterior commissure) correction, bias field correction, skull-stripping, intensity inhomogeneity correction, cerebellum removal, segmentation, and nonlinear registration [Wang et al., 2011]. After segmentation, the brain was segmented into four different tissues: gray matter (GM), white matter (WM), ventricle (VN), and cerebrospinal fluid (CSF). We quantified the local volumetric group differences by generating a RAVENS maps [Davatzikos et al., 2001] for the whole brain and each of the segmented tissue types (GM, WM, VN, and CSF) respectively, using the deformation field which we obtained during registration. RAVENS methodology is based on a volume-preserving spatial transformation, since this process changes an individual′s brain morphology to conform it to the morphology of the Jacob template.

We are interested in detecting meaningful brain regions of interest that are associated with several candidate genes. We included only the subjects whose diagnostic status was healthy control or Alzheimer's disease at baseline and had no status change during ADNI1. After screening, the total number of subjects we included was 372 (195 Healthy Controls (HCs) and 177 ADs). The clinical covariates of interest included gender, baseline age, square of baseline age, handedness, education, baseline intracranial volume, and the risk of Apolipoprotein E (ApoE). Specifically, handedness was treated as a binary variable, the education information was the self-reported years of education by the subjects, and the risk of APOE is assumed to be additive. Specifically, the risk of APOE for a subject was 3 if he/she carries $\varepsilon4$ at both alleles; it was 2 if he/she carries $\varepsilon3$ and $\varepsilon4$ in two alleles, the risk would be considered 0 if the two APOE alleles were the combination of $\varepsilon2$ and $\varepsilon3$, and other combinations of APOE alleles are assumed to have risk 1.

Many genes have been reported to be causal in the progression of Alzheimer's disease. We selected three candidate causal genes including CR1 on chromosome 1, CD2AP on chromosome 6, and PICALM on chromosome 11 due to their strong association with the progression of Alzheimer's disease [Harold et al., 2009; Lambert et al., 2009; Naj et al., 2011]. Specifically, PICALM encodes the protein phosphatidylinositol-binding clathrin assembly and is highly correlated with the emergence of late-on-set AD, which is possibly due to the perturbation at synapse triggering its function change [Harold et al., 2009]. The gene CD2AP encodes the CD2-asscociated protein and involves the cell membrane, including endocytosis, that plays critical roles in neurodegeneration and Aβ clearance from the brain [Naj et al., 2011]. The gene CR1 encodes the complement component (3b/4b) receptor 1 and the pathways involving CR1 are involved in the AD process, specifically in clearance of Aβ peptides, which is the primary composition of amyloid plaques [Lambert et al., 2009].

We first matched the SNPs in ADNI with the gene list "glist-hg18" provided by PLINK [Purcell et al., 2007] and were able to locate 16, 15, and 23 SNPs on the selected CR1, CD2AP, and PICALM genes, respectively. All these SNPs pass the quality control procedure with MAF > 5% and the Hardy Weinberg Equilibrium (HWE) test P-value > 0.01. The MAFs of the SNPs of the selected genes vary from 0.1 to 0.5. After deleting missing values, there are 335, 299, and 328 subjects corresponding to the CR1, CD2AP, and PICALM genes, respectively. The MAFs of all selected SNPs and demographic information are included in the supplementary document.

For each selected gene, we fitted FMEM (1) with **z** coded as the number of minor alleles in order to detect its associated
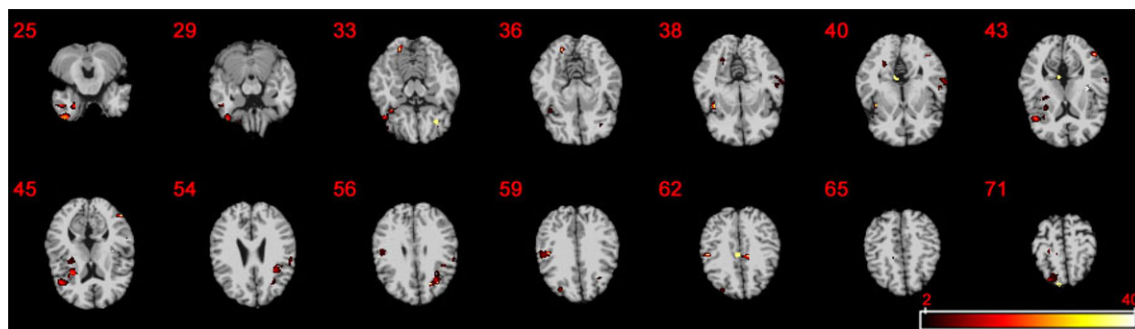
Figure 4. ADNI data analysis: the $-\log_{10}(p)$ maps for testing the genetic effect of CD2AP on RAVEN images by using FMEM from 14 selected slices.

brain regions of interest (ROIs). For comparisons, we fit the same model by using the classical voxel-wise method and Ge's method to the same data sets. To formally detect significant ROIs, by following Ge et al. [2012], we used a cluster-form of threshold of 0.1% with a minimum voxel clustering value of 50 voxels. The names of the brain regions were included in Tables 6–8 of the supplementary document. FMEM is able to to detect 45, 45, and 27 significant clusters for CR1, CD2AP, and PICALM, respectively, whereas the standard voxel-wise method can only identify 6, 14, and 2 significant clusters, and none from Ge's method for CR1, CD2AP, and PICALM, respectively. We also fit FMEM on the same data but only with the HC and AD samples only to investigate white noise signal. For HC samples only, FMEM detected 15, 8, and 31 significant clusters for CR1, CD2AP, and PICALM, respectively. For AD samples only, FMEM detected 9, 8, and 41 significant clusters for CR1, CD2AP, and PICALM, respectively. Although there are some discrepancies between the results based on the HC and AD samples only and those based on the combined sample, the results are highly similar to each other. The results obtained from the combined sample are generally more significant due to a larger sample size.

Finally, we overlapped these significant clusters with the 96 predefined ROIs in the Jacob template and were able to detect several predefined ROIs for CR1, CD2AP, and PICALM. For CD2AP, based on the combined sample, FMEM identified relatively large clusters with more than 150 voxels of right superior temporal gyrus, left and right inferior temporal gryus, left and right precentral gyrus, left and right middle frontal gyrus, right postcentral gyrus, right fusiform, left angular, left inferior frontal gyrus, left inferior occipital gyrus, left and right postcentral gyrus, left and right superior frontal gyrus, left anterior cingulate and paracingulate gyri, left median cingulate and paracingulate gyri, right calcarine fissure and surrounding cortex, right cuneus, right superior occipital gyrus, right middle occipital gyrus, right caudate, and right middle temporal gyrus.

For CR1, based on the combined sample, FMEM identified relatively large clusters with more than 150 voxels of right superior temporal gyrus, left and right putamen, left inferior temporal gyrus, left angular, left inferior occipital gyrus, right postcentral gyrus, right superior frontal gyrus, left anterior cingulate and paracingulate gyri, left median cingulate and paracingulate gyri, left cuneus, left middle occipital gyrus, and right caudate.

For PICALM, based on the combined sample, FMEM identified relatively large clusters with more than 150 voxels of right inferior frontal gyrus-triangular and orbital parts, right insula, right fusiform, right superior temporal gyrus, right temporal pole, right middle temporal gyrus, right inferior temporal gyrus, right precentral gyrus, right supramarginal gyrus, right middle occipital gyrus, right angular, right middle frontal gyrus, and right middle frontal gyrus.

As shown in the supplementary document, we were able to detect several major ROIs, such as superior temporal gyrus, inferior temporal gyrus, middle frontal gyrus, angular, anterior cingulate and paracingulate gyri, hippocampus, putamen, and fusiform. Our finding of these ROIs are highly similar to previous reports on brain morphology in Alzheimer's disease in the AD literature [Convit et al., 2000; Fennema-Notestine et al., 2009; Jones et al., 2006; Ohnishi et al., 2001]. The superior temporal gyrus is an essential structure involved in auditory processing, in social cognition processes, as well as in the function of language. The inferior temporal gyrus is one of the higher levels of the ventral stream of visual processing. The middle frontal gyrus plays a role in sustaining attention and working memory. The angular gyrus is involved in a number of processes related to language, number processing and spatial cognition, memory retrieval, attention, and theory of mind. The anterior cingulate and paracingulate gyri are involved in rational cognitive functions, such as reward anticipation, decision-making, empathy, impulse control, and emotion. The hippocampus is known to be associated with memory and cognition. The fusiform is associated with color recognition, word, and body recognition and the putamen is associated with motor skills. Figure 4 shows the $-\log_{10}(p)$ map of selected slices with significant clusters for testing the genetic effects of CD2AP on RAVEN images identified by FMEM.

Discussion

We have developed FMEM to carry out an association analysis between neuroimaging phenotypes and a group of genetic markers, while adjusting for the clinical variables of interest. We have proposed a multiscale adaptive procedure with three features: spatial, hierarchical, and adaptive. Our simulation results have shown substantial gains in parameter estimation precision and statistical power in detecting the true effect of ROIs compared to the voxel-wise method. More research is needed for optimizing the choice of the tuning parameters in FMEM. We will borrow some key ideas of FMEM to develop a fast procedure to carry out GWAS for imaging genetic studies. We will also develop fast FMEM for the joint analysis of neuroimaging and genetic data with rare or common variants [Fan et al., 2012, 2013].

References

Amos W, Driscoll E, Hoffman JI. 2011. Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proc Biol Sci* 278:1183–1188.

Ashburner J, Friston KJ. 2000. Voxel-based morphometry: the methods. *Neuroimage* 11:805–821.

Ball T, Breckel TP, Mutschler I, Aertsen A, Schulze-Bonhage A, Hennig J, Speck O. 2012. Variability of FMRI-response patterns at different spatial observation scales. *Hum Brain Mapp* 33:1155–1171.

Convit A, de Asis J, de MJ Leon, Tarshish CY, De Santi S, Rusinek H. 2000. Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. *Neurobiol Aging* 21:19–26.

Davatzikos C, Genc A, Xu D, Resnick SM. 2001. Voxel-based morphometry using the ravens maps: methods and validation using simulated longitudinal atrophy. *Neuroimage* 14:1361–1369.

Fan J, Gijbels I. 1996. *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Fan R, Zhang Y, Albert PS, Liu A, Wang Y, Xiong M. 2012. Longitudinal association analysis of quantitative traits. *Genet Epidemiol* 36(8):856–869.

Fan R, Wang Y, Mills JL, Wilson AF, Bailey-Wilson JE, Xiong M. 2013. Functional linear models for association analysis of quantitative traits. *Genet Epidemiol* 37(7):726–742.

Fennema-Notestine C, Hagler DJ Jr, McEvoy LK, Fleisher AS, Wu EH, Karow DS, Dale AM, Alzheimer's Disease Neuroimaging Initiative. 2009. Structural MRI biomarkers for preclinical and mild Alzheimer's disease. *Hum Brain Mapp* 30:3238–3253.

Friston KJ. 2009. Modalities, modes, models in functional neuroimaging. *Science* 326:399–403.

Ge T, Feng J, Hibar DP, Thompson PM, Nichols TE. 2012. Increasing power for voxel-wise genome-wide association studies: the random field theory, least square kernel machines and fast permutation procedures. *Neuroimage* 63:858–873.

Gilmore JH, Schmitt JE, Knickmeyer RC, Smith JK, Lin W, Styner M, Gerig G, Neale MC. 2010. Genetic and environmental contributions to neonatal brain structure: a twin study. *Hum Brain Mapp* 31:1174–1182.

Greven S, Crainiceanu C, Caffo B, Reich D. 2010. Longitudinal functional principal component analysis. *Electron J Statist* 4:1022–1054.

Guo W. 2002. Functional mixed effects models. *Biometrics* 58:121–128.

Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Williams A and others. 2009. Genome-wide association study identifies variants at clu and picalm associated with Alzheimer's disease. *Nat Genet* 41:1088–1093.

Hibar DP, Stein JL, Kohannim O, Jahanshad N, Saykin AJ, Shen L, Kim S, Pankratz N, Foroud T, Huentelman MJ and others. 2011. Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. *Neuroimage* 56:1875–1891.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 106:9362–9367.

Jones BF, Barnes J, Uylings HB, Fox NC, Frost C, Witter MP, Scheltens P. 2006. Differential regional atrophy of the cingulate gyrus in Alzheimer disease: a volumetric MRI study. *Cereb Cortex* 16:1701–1708.

Jones DK, Symms MR, Cercignani M, Howard RJ. 2005. The effect of filter size on VBM analyses of DT-MRI data. *Neuroimage* 26:546–554.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42:348–354.

Lambert J, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B and others. 2009. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 41:1094–1099.

Li Y, Zhu H, Shen D, Lin W, Gilmore JH, Ibrahim JG. 2011. Multiscale adaptive regression models for neuroimaging data. *J Royal Stat Soc Series B* 73:559–578.

Li Y, Gilmore JH, Shen D, Styner M, Lin W, Zhu H. 2013. Multiscale adaptive generalized estimating equations for longitudinal neuroimaging data. *Neuroimage* 72:91–105.

Liu D, Lin X, Ghosh D. 2007. Semiparametric regression of multidimensional genetic pathway data: least-squares kernel machines and linear mixed models. *Biometrics* 63:1079–1088.

Loth E, Carvalho F, Schumann G. 2011. The contribution of imaging genetics to the development of predictive markers for addictions. *Cell* 15:436–446.

Morris JS, Carroll RJ. 2006. Wavelet-based functional mixed models. *Stat JR Soc Ser B Stat Methodol* 68:179–199.

Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK and others. 2011. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 43:436–441.

Ohnishi T, Matsuda H, Tabira T, Asada T, Uno M. 2001. Changes in brain morphology in Alzheimer disease and normal aging: is Alzheimer disease an exaggerated aging process? *AJNR Am J Neuroradiol* 22:1680–1685.

Polzehl J, Spokoiny VG. 2000. Adaptive weights smoothing with applications to image restoration. *Statist JR Soc B* 62:335–354.

Polzehl J, Voss HU, Tabelow K. 2010. Structural adaptive segmentation for statistical parametric mapping. *Neuroimage* 52:515–523.

Purcell S, Neale B, Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P and others. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.

Savitz JB, Drevets WC. 2009. Imaging phenotypes of major depressive disorder: genetic correlates. *Neuroscience* 164:300–330.

Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, Foroud TM, Pankratz ND, Moore JH, Sloan SD and others. 2010. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort. *Neuroimage* 53:1051–1063.

Silver M, Montana G, Nichols TE, Alzheimer's Disease Neuroimaging Initiative. 2011. False positives in neuroimaging genetics using voxel-based morphometry data. *Neuroimage* 54:992–1000.

Stram DO, Lee JW. 1994. Variance components testing in the longitudinal mixed effects model. *Biometrics* 50:1171–1177.

Thompson PM, Ge T, Glahn DC, Jahanshad N, Nichols TE. 2013. Genetics of the connectome. *Neuroimage* 80:475–488.

Tzeng JY, Zhang D. 2007. Haplotype-based association analysis via variance component score test. *Am J Hum Genet* 81:927–938.

Vounou M, Nichols TE, Montana G, the Alzheimer's Disease Neuroimaging Initiative. 2010. Discovering genetic associations with high-dimensional neuroimaging phenotypes: a sparse reduced-rank regression approach. *Neuroimage* 53:1147–1159.

Wang Y, Chen H. 2012. On testing an unspecified function through a linear mixed effects model with multiple variance components. *Biometrics* 68:1113–1125.

Wang Y, Nie J, Yap PT, Shi F, Guo L, Shen D. 2011. Robust deformable-surface-based skull-stripping for large-scale studies. In: Fichtinger G, Martel A, and Peters T, editors. *Medical Image Computing and Computer-Assisted Intervention*, Volume 6893. Toronto, Canada: Springer Berlin/Heidelberg, pp. 635–642.

Worsley KJ, Taylor JE, Tomaiuolo F, Lerch J. 2004. Unified univariate and multivariate random field theory. *Neuroimage* 23:189–195.

Yuan Y, Gilmore JH, Geng X, Styner M, Chen K, Wang JL, Zhu H. 2014. FMEM: functional mixed effects modeling for the analysis of longitudinal white matter Tract data. *Neuroimage* 84:753–764.

Yue Y, Loh JM, Lindquist MA. 2010. Adaptive spatial smoothing of fMRI images. *Stat Interface* 3:3–14.

Zhang Y, Xu Z, Shen X, Pan W. 2014. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *Neuroimage* 96:309–325.

Zhu H, Brown PJ, Morris JS. 2011. Robust, adaptive functional regression in functional mixed model framework. *J Am Stat Assoc* 106:1167–1179.

Zhu HT, Li RZ, Kong LL. 2012. Multivariate varying coefficient model for functional responses. *Annal Stat* 40:2634–2666.

Zhu HT, Fan JQ, Kong LL. 2014. Spatially varying coefficient model for neuroimaging data with jump discontinuities. *J Am Stat Assoc* 109:in press.

Zhu M, Zhao S. 2007. Candidate gene identification approach: progress and challenges. *Int J Biol Sci* 3:420–427.