

# Spatiotemporal mixed modeling of multi-subject task fMRI via method of moments



Benjamin B. Risk<sup>a,b,\*</sup>, David S. Matteson<sup>c</sup>, R. Nathan Spreng<sup>d</sup>, David Ruppert<sup>c</sup>

<sup>a</sup> Statistical and Applied Mathematical Sciences Institute, 19 T.W. Alexander Drive, P.O. Box 14006, Research Triangle Park, NC 27709, USA

<sup>b</sup> Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

<sup>c</sup> Department of Statistical Sciences, Cornell University, Ithaca, NY 14853, USA

<sup>d</sup> Department of Human Development, Cornell University, Ithaca, NY 14853, USA

## ARTICLE INFO

### Article history:

Received 12 February 2016

Revised 20 April 2016

Accepted 13 May 2016

Available online 19 May 2016

### Keywords:

Best linear unbiased prediction

Covariogram

General linear model

Human Connectome Project

Smoothing

## ABSTRACT

Estimating spatiotemporal models for multi-subject fMRI is computationally challenging. We propose a mixed model for localization studies with spatial random effects and time-series errors. We develop method-of-moment estimators that leverage population and spatial information and are scalable to massive datasets. In simulations, subject-specific estimates of activation are considerably more accurate than the standard voxel-wise general linear model. Our mixed model also allows for valid population inference. We apply our model to cortical data from motor and theory of mind tasks from the Human Connectome Project (HCP). The proposed method results in subject-specific predictions that appear smoother and less noisy than those from the popular single-subject univariate approach. In particular, the regions of motor cortex associated with a left-hand finger-tapping task appear to be more clearly delineated. Subject-specific maps of activation from task fMRI are increasingly used in pre-surgical planning for tumor removal and in locating targets for transcranial magnetic stimulation. Our findings suggest that using spatial and population information is a promising avenue for improving clinical neuroimaging.

© 2016 Elsevier Inc. All rights reserved.

## Introduction

The aim of this paper is to develop a spatiotemporal mixed model for activation in task fMRI for improved prediction of subject-specific activation maps. Prediction of individual patterns of activation is often a goal in neuroscience studies (e.g., Mitchell et al., 2008). Mapping brain regions in individuals is important in clinical neuroimaging, and localization studies with task fMRI are increasingly used for pre-surgical mapping of motor and language areas in tumor and epileptic patients (Pillai, 2010). Task fMRI is a non-invasive technique that can be used in neuronavigational planning to allow the surgeon to minimize damage to brain tissue in the path of the tumor (Holodny et al., 2011). Subject-specific activation maps from fMRI can also be used to identify targets for transcranial magnetic stimulation (tMS) (Neggers et al., 2004; Wig et al., 2005) and for multifocal transcranial direct current stimulation (Ruffini et al., 2014).

In this paper, we will address two shortcomings of a typical analysis of task fMRI: (1) population information is not utilized in subject-level maps of activation; and (2) spatial information is ignored during estimation. The most common approach to task fMRI fits a GLM to the time series for each voxel and each subject (Friston et al., 1995) and

then a second GLM is fit to first-level estimates. This two-stage approach can be formulated as a mixed model (Worsley et al., 2002), so called because it contains fixed population effects and random subject effects. We call this approach the massive univariate mixed model (MUMM). Notably, the subject-level estimates of activation are never revisited. Shrinking predictions towards population estimates improved subject-specific estimates of functional connectivity in resting-state fMRI (Shou et al., 2014) and in neurological pain signatures (Lindquist et al., 2015), but a model-based approach for task-fMRI is lacking. The MUMM contrasts with the conventional approach to mixed modeling wherein predictions utilize empirical best linear unbiased estimators (eBLUPs), which shrink subject-specific effects when the between subject variability is small (e.g., Searle et al., 1992). Secondly, by fitting a separate model to each location, the MUMM does not exploit the spatial structure. This contrasts with spatial statistics in which information from nearby locations is leveraged for improved prediction (e.g., Cressie, 1993).

It is useful to categorize spatial information into two types: 1) spatial dependence in the correlation structure and 2) spatial smoothness in the mean structure (Zhu et al., 2014). Most studies in neuroimaging focus on modeling the spatial dependence in the correlation structure using random effects, where the random effects are usually generated from Gaussian random fields (GRFs). The dependence structure of GRFs can be defined using a covariogram (Kang et al., 2012; Castruccio

\* Corresponding author at: Statistical and Applied Mathematical Sciences Institute, 19 T.W. Alexander Drive, P.O. Box 14006, Research Triangle Park, NC 27709, USA.

et al., 2016; Bernal-Rusiel et al., 2013) or a spatial autoregressive model (Hyun et al., 2014). Bowman (2007) used the exponential covariogram with functionally based distances in a region-of-interest (ROI) analysis comprising 239 voxels from twelve subjects with four time points each. This model was applied to a surface-based analysis of cortical thickness in which 149,000 vertices (locations on the FreeSurfer cortical surface) were classified into 12,000 independent parcels (contiguous collection of vertices), which was more powerful than the MUMM (Bernal-Rusiel et al., 2013). In single-subject analysis of fMRI, long-range correlations have been modeled by estimating a correlation structure between regions of interest (ROIs) (Kang et al., 2012; Castruccio et al., 2016). Hyun et al. (2014) improved prediction in a surface model of the left lateral ventricle (approximately 1000 vertices) by modeling local correlations using a simultaneous autoregressive (AR) model and long-range correlations using functional PCA (fPCA), which does not require ROIs. Perhaps the simplest spatial dependence in correlation assumes an exchangeable covariance structure within a parcel or ROI, such that the correlation between any two locations is equal (Derado et al., 2010; Bowman, 2005). Regarding 2), spatial smoothness in the mean structure can be modeled using local polynomial regression. The multiscale adaptive regression model (MARM) is a more general approach that allows jump discontinuities, which is useful when analyzing multiple tissue types (Li et al., 2011). Zhu et al. (2014) modeled both the mean using MARM and the correlation structure using fPCA in structural MRI, which increased statistical power. Note the distinction between the correlation structure and smoothness in the mean is less relevant in Bayesian approaches in which population effects are also random variables.

Spatial models for neuroimaging data have improved prediction and/or statistical power, but their application to multi-subject fMRI has been limited primarily by computational constraints. In particular, previous models are either for other imaging modalities with few timepoints (e.g., structural MRI, PET, or DTI); for a single-subject; or ignore time series information in the group-level analysis by using the test statistics from subject-specific analyses. Derado et al. (2013) improved prediction in longitudinal PET data by fitting a conditional autoregressive (CAR) model in each brain parcel and modeling the correlation between parcels. Harrison and Green (2010) developed a variational Bayes approach for whole-brain analysis of a single subject. Zhang et al. (2014) developed a single-subject model that clustered voxels with similar time courses using a Dirichlet process while modeling long-range temporal and local spatial dependence. Brown et al. (2014) applied a Bayesian CAR model to the t-statistics from a single subject to address the multiple testing problem. Bowman et al. (2008) developed a Bayesian model that flexibly modeled long-range correlations between regions and assumed correlations within a region were exchangeable. For a review of Bayesian approaches, see Zhang et al. (2015). A potential drawback of the spatial AR models is that the marginal properties can be undesirable. For instance, the variance may depend on the number of neighbors and boundary effects can be problematic. Xu et al. (2009) develop an alternative approach in which activation follows a spatial Poisson process using the t-statistics from a first-level analysis.

We propose a spatiotemporal mixed model (STMM) that contains the following features: fixed effects characterizing task activation in the population for each vertex; regional subject random effects characterizing an individual's activation common to all vertices in a region; subject-vertex spatial random effects that capture an individual's activation unique to each vertex; and time-series errors following an AR(p) model. In the marginal model (integrating out all random effects), population activation is defined by a separate parameter for each location, which does not explicitly incorporate spatial smoothness. The regional subject random effects capture an exchangeable covariance structure within regions, similar to Bowman et al. (2008) and Derado et al. (2010). We use the functional parcellation defined in (Gordon et al. (2016), which is based on resting-state correlations and

parcels vertices into groups that are more homogeneous than alternative parcellations such as Brodmann areas. The subject-vertex random effects characterize dependence as a function of distance, similar to Bowman (2007). To overcome issues with volume-based Euclidean distances, we focus on the cortical surface and use geodesic distances between vertices within the same parcel. The time-series errors allow timepoints closer in time to be more similar and can capture periodic behavior, similar to Worsley et al. (2002). If one conditions on the regional subject and subject-vertex random effects (integrating out the time-series errors only), then the regional subject effects allow for jump discontinuities between parcels and the subject-vertex random effects capture the spatial smoothness in the conditional mean for each vertex in the parcel. In this respect, “what is one person's covariance structure is another person's mean structure (p. 25, Cressie, 1993).”

Our contributions are the following:

- We build upon previous spatial models for single-subject fMRI or other imaging modalities to develop a model for multi-subject fMRI. A major obstacle to fitting our proposed model is the large covariance matrix. In our application, the likelihood includes a 500,000 by 500,000 dense covariance matrix (corresponding to the largest parcel with two runs of fMRI data). In order to fit this model to large datasets, we project the spatiotemporal data to the span of the covariates. For each region, the projected data then follow a two-factor crossed design with a fixed factor (vertex) and random factor (subject). This allows us to derive method-of-moments (ANOVA) estimators of the variance components.
  - Unlike previous methods, we incorporate eBLUPs into estimates of subject-specific activation that leverage both population and spatial information. In simulations, subject-specific estimates have much lower mean squared error (MSE) than the univariate approach.
  - We apply our model to the motor task from thirty subjects (a number representative of fMRI studies) and to the theory of mind task from ninety-eight subjects (the number of subjects included in the 2014 data sampler) from the Human Connectome Project (HCP). Whereas the subject-specific maps based on the univariate approach appear noisy, our approach includes model-based smoothing. In particular, the regions of motor cortex associated with a left-hand finger-tapping task appear to be more clearly delineated.
- The remainder of this paper is organized as follows. We review the MUMM in the next section. We then propose our model in the *Spatiotemporal Mixed Model (STMM)* section. In the *Simulations* section, we conduct simulations examining the accuracy of subject-level activation maps, type 1 error rates, and statistical power. In the *Analysis of HCP task data*, we analyze the HCP motor and theory of mind tasks for the right cerebral cortex. In the *Discussion* section, we conclude that the STMM improves subject-level activation maps and allows for population-level inference. Formulas for our estimators are in the Appendix. Derivations and additional tables and figures are available in the Web Supplement. Matlab code implementing our method is available at <http://www.benjaminrisk.com/software>.

## Massive univariate mixed model (MUMM)

The great majority of localization studies in fMRI use the MUMM, and although software may differ in some of the modeling details, they generally fit a separate univariate model with fixed population-level and random subject-level vertex effects to each location.

### First level (subject effects)

A summary of the notation used in this paper appears in the Web Supplement Table S.1. Let  $i \in \{1, \dots, N\}$  denote subject,  $v \in \{1, \dots, V\}$  denote the vertex, and  $t \in \{1, \dots, T\}$  index time. Let  $y_{ivt}$  denote the

BOLD signal. Let  $q \in \{1, \dots, Q\}$  index a task. Let  $x_{itq}$  be the covariate created from the canonical hemodynamic response function (HRF) (difference of two gamma functions) convolved with task onsets and durations (Penny et al., 2007). Let  $\mathbf{x}_{it} = [x_{it1}, \dots, x_{itQ}]'$ , where  $'$  denotes the matrix transpose. Let  $a_{ivq}$  denote the magnitude of activation attributed to the  $q$ th task, and let  $\mathbf{a}_{iv} = [a_{iv1}, \dots, a_{ivQ}]'$ . In the simplest case, the same HRF is assumed for all locations. This assumption can be relaxed by including partial derivatives of parameters of the HRF evaluated across time, such that coefficients are estimated to allow the HRF to vary while the covariates remain the same at every location (Penny et al., 2007). We incorporate time-delay and dispersion derivatives as nuisance covariates. Let  $m \in \{1, \dots, M\}$  index the nuisance covariates, which additionally include a spline basis to capture scanner drift and the parameters used in the affine motion correction.

Let  $\mathbf{z}_{it} \in \mathbb{R}^M$  denote all nuisance variables. Let  $\gamma_{ivm}$  denote the coefficient for the  $m$ th nuisance covariate, and define  $\boldsymbol{\gamma}_{iv} = [\gamma_{iv1}, \dots, \gamma_{ivM}]'$ . Note that the covariates do not change with space. Let  $e_{ivt}$  be the error and let  $\mathbf{e}_{iv} = [e_{iv1}, \dots, e_{ivT}]'$ . Define the first-level model as

$$y_{ivt} = \mathbf{x}_{it}'\mathbf{a}_{iv} + \mathbf{z}_{it}'\boldsymbol{\gamma}_{iv} + e_{ivt} \quad (1)$$

with

$$\mathbf{e}_{iv} \sim \mathcal{N}(\mathbf{0}_T, \xi_{iv}^2 \boldsymbol{\Psi}_{iv}) \quad (2)$$

where  $\mathbf{0}_T$  is a vector of zeros of length  $T$ ,  $\boldsymbol{\Psi}_{iv}$  is a positive-definite Toeplitz matrix that captures the correlation between serial observations, and  $\mathbf{e}_{iv}$  are mutually independent for all  $i = 1, \dots, N$  and  $v = 1, \dots, V$ . We will refer to  $\boldsymbol{\Psi}_{iv}$  as the error correlation.

There is empirical support for the use of a causal stationary autoregressive (AR) model for the errors in (1) (Worsley et al., 2002). In our exploratory analysis of task HCP data, we found that an AR(3) model was preferred for many locations, and thus we will use an AR(3) model for all vertices (see also Fig. S.2 for the coefficients of the third AR parameter for a randomly chosen subject). Note that Lindquist (2008) suggests an AR(2) model. For the HCP dataset, the time between scans is less than in other fMRI studies (0.72 s versus approximately 2.5 s), which may be a factor contributing to the increased support for the AR(3) model.

When there exist multiple sessions for the same subject, we assume time-series errors are independent between sessions but with the same AR parameters. Let  $B$  denote the back-shift operator:  $By_{ivt} = y_{i,v,t-1}$ . Consider  $y_{ivt}$  such that  $t-3, \dots, t-1$  are in the same session. Then the first-level model for the  $i$ th subject is.

$$(1 - \phi_{iv1}B - \phi_{iv2}B^2 - \phi_{iv3}B^3)(y_{ivt} - \mathbf{x}_{it}'\mathbf{a}_{iv} - \mathbf{z}_{it}'\boldsymbol{\gamma}_{iv}) = \epsilon_{ivt}, \quad (3)$$

where  $\epsilon_{ivt} \stackrel{iid}{\sim} \mathcal{N}(0, \tau_{iv}^2)$ . Here,  $\epsilon_{ivt}$  are independent and identically distributed errors with innovation variance  $\tau_{iv}^2$  whereas in (2),  $e_{ivt}$  are the correlated errors with unconditional variance  $\xi_{iv}^2$ .

### Second level (population effects)

In the second-level, the subject-specific effects are generated from a fixed population effect plus a random effect. For each  $q$ , we define task activation.

$$a_{ivq} = \beta_{vq} + b_{ivq}, \quad (4)$$

where  $\beta_{vq}$  are arbitrary but fixed,  $b_{ivq} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{b_{vq}}^2)$  for  $i = 1, \dots, N$  and  $v = 1, \dots, V$ . We will also assume  $b_{ivq}$  and  $b_{ivq'}$  are mutually independent for all  $q \neq q' \in \{1, \dots, Q\}$ .

The two-level model can be formalized in a single mixed model by substituting (4) into (3):

$$(1 - \phi_{iv1}B - \phi_{iv2}B^2 - \phi_{iv3}B^3)(y_{ivt} - \mathbf{x}_{it}'(\beta_v + \mathbf{b}_{iv}) - \mathbf{z}_{it}'\boldsymbol{\gamma}_{iv}) = \epsilon_{ivt} \quad (5)$$

where  $\beta_v = [\beta_{v1}, \dots, \beta_{vQ}]'$  and  $\mathbf{b}_{iv} = [b_{iv1}, \dots, b_{ivQ}]'$ . Additionally, we assume all  $b_{ivq}$  and  $\epsilon_{ivt}$  are mutually independent.

In this paper, we define subject-specific activation according to  $a_{ivq}$  in (4), and we will focus on estimates  $\hat{a}_{ivq}$ . Other studies use t-statistics from a fixed-effects treatment of the first-level analysis to define statistical parametric maps (a scaled version of  $\hat{a}_{ivq}$ ) (Friston et al., 1995). Note that for any tasks  $q$  and  $q'$ ,  $a_{ivq}$  and  $a_{ivq'}$  are on the same scale because the covariates of interest in (5) are defined by convolving the same HRF with functions taking only zero and one as values.

### Estimating the MUMM

Different estimators include the “summary statistics” option in SPM (Penny et al., 2007), REML using the EM algorithm (Worsley et al., 2002), and a “hybrid” generalized least squares (GLS) and Bayesian approach used in FSL (Beckmann et al., 2003; Woolrich et al., 2004). In a survey of ninety fMRI papers, Mumford and Nichols (2009) found that 92% used the summary statistics approach. The summary statistics approach uses OLS (ordinary least squares) with temporally pre-whitened  $y_{ivt}$  to estimate subject-specific coefficients and then averages the subject-specific coefficients to obtain estimates of the population parameters  $\beta_v$ . Such estimates are the default in Statistical Parametric Mapping (SPM) software (Penny et al., 2007). When subject data are pre-whitened using the covariance matrix of the errors, it is equivalent to GLS (Worsley et al., 2002).

Other approaches incorporate information on the subject-specific error variance,  $\xi_{iv}^2$ , into second-level estimates. Then the GLS estimator of the fixed effects becomes a linear combination of the subject-specific coefficients, with weights determined by an estimate of the subject-specific error variance (Beckmann et al., 2003). Studies have found the summary statistics estimators are nearly as powerful as estimators from a single-stage REML (Friston et al., 2005) or estimators from the weighted approach when one group of subjects is being analyzed (e.g., healthy subjects) (Mumford and Nichols, 2009). Our data application contains healthy young adults. Thus we use the summary statistics approach with first-level GLS estimators as in (a.1) and second-level t-statistics from (a.4), see Appendix A. For the first-level, we calculate reduced-bias estimators of the AR parameters as developed in Worsley et al. (2002), which we have found are very accurate and fast to calculate. We provide a detailed account of the biasedness of the OLS estimator of the error variance in the Web Supplement Section 2 and a derivation of the reduced biased estimator in Web Supplement Section 3.

### A spatiotemporal mixed effects model (STMM)

#### Defining the STMM

We will assume there exists a parcellation such that the random effects between regions are independent. Then we can treat each region as a separate estimation problem, which makes model estimation computationally feasible. In this framework, information from one region does not provide any information on another region. Note this is also a convenient method to allow for different strengths of spatial dependence in different regions (local stationarity).

Let  $\mathbf{Y}_{ir} = [y_{ir11}, \dots, y_{ir21}, \dots, y_{irV_rT}]'$  denote observations from the  $r$ th region (parcel) containing vertices  $1, \dots, V_r$  for  $r = 1, \dots, R$ .



Let  $\mathbf{X}_i \in \mathbb{R}^{T \times Q}$  (covariates of interest for all timepoints). Let  $\beta_r = [\beta_{r1}, \dots, \beta_{rQ}]'$  denote the region-level fixed effects. Define the vertex-level fixed effects  $\beta_{rv} = [\beta_{rv1}, \dots, \beta_{rvQ}]'$  and  $\beta_r = [\beta_{r1}, \dots, \beta_{rV}]'$ . Let  $\mathbf{s}_{ir} = [s_{ir1}, \dots, s_{irQ}]'$  denote the subject-specific regional random slopes of the covariates of interest (tasks convolved with HRF). Define the interaction between the fixed vertex and random regional subject effects:  $\mathbf{b}_{irv} = [b_{irv1}, \dots, b_{irvQ}]'$  and  $\mathbf{b}_{ir} = [b_{ir1}, \dots, b_{irV}]'$ . Let  $\mathbf{b}_{ir}^q = [b_{ir1q}, \dots, b_{irVq}]'$  denote an ordering of the interaction effects that will be convenient for defining the covariance structure. Let  $\mathbf{Z}_i \in \mathbb{R}^{T \times M}$  (nuisance covariates) and  $\gamma_{irv} = [\gamma_{irv1}, \dots, \gamma_{irvM}]'$  denote the fixed effects associated with the nuisance terms. Finally define the errors  $\mathbf{e}_{irv} = [e_{irv1}, \dots, e_{irvV}]'$  and  $\mathbf{e}_{ir} = [e_{ir1}, \dots, e_{irV}]'$ .

We define the spatiotemporal mixed model (STMM) as.

$$\mathbf{Y}_{ir} = \mathbf{1}_{V_r} \otimes \mathbf{X}_i \beta_r + (\mathbf{I}_{V_r} \otimes \mathbf{X}_i) \beta_r + \mathbf{1}_{V_r} \otimes \mathbf{X}_i \mathbf{s}_{ir} + (\mathbf{I}_{V_r} \otimes \mathbf{X}_i) \mathbf{b}_{ir} + (\mathbf{I}_{V_r} \otimes \mathbf{Z}_i) \gamma_{ir} + \mathbf{e}_{ir}. \quad (6)$$

where

$$\mathbf{s}_{ir} \stackrel{iid}{\sim} \mathcal{N}(0_Q, \mathbf{S}_r) \quad (7)$$

with  $\mathbf{S}_r = \text{diag}(\sigma_{s_{r1}}^2, \dots, \sigma_{s_{rQ}}^2)$ ;

$$\mathbf{b}_{ir}^q \stackrel{iid}{\sim} \mathcal{N}(0_{V_r}, \sigma_{b_{ir}^q}^2 \mathbf{\Omega}_{rq}) \quad (8)$$

and define  $\mathbf{B}_r = \text{diag}(\sigma_{b_{r1}}^2, \dots, \sigma_{b_{rQ}}^2)$ ; and

$$\mathbf{e}_r \sim \mathcal{N}\left\{0_{NV_rT}, \text{diag}\left(\xi_{1r1}^2 \mathbf{\Psi}_{1r1}, \dots, \xi_{NrV}^2 \mathbf{\Psi}_{NrV}\right)\right\}, \quad (9)$$

where  $\text{diag}(\xi_{1r1}^2 \mathbf{\Psi}_{1r1}, \dots, \xi_{NrV}^2 \mathbf{\Psi}_{NrV})$  is the block diagonal matrix formed from the error covariance matrices from all subjects and all vertices. The spatial correlation matrix  $\mathbf{\Omega}_{rq}$  corresponds to a stationary and isotropic process from the exponential covariogram, discussed below. We allow this correlation matrix to vary across parcels, creating locally stationary processes. We use the geodesic distance between vertices on the FreeSurfer 32 K spherical template. As in the MUMM and (2),  $\mathbf{\Psi}_{irv}$  is the autocorrelation matrix of an AR(3) process in which time-series errors from different sessions of the same subject are independent but have the same AR parameters. We assume  $\mathbf{b}_{ir}^q$ ,  $\mathbf{s}_{ir}$ , and  $\mathbf{e}_{ir}$  for  $i = 1, \dots, N$  and  $q = 1, \dots, Q$  are mutually independent.

For an observation  $y_{irvt}$  such that  $t = 3, \dots, T$  are in the same session, the model is

$$\left(1 - \phi_{irv1}B - \phi_{irv2}B^2 - \phi_{irv3}B^3\right)y_{irvt} = \mathbf{x}_{it}'(\beta_r + \beta_{rv} + \mathbf{s}_{ir} + \mathbf{b}_{irv}) + \mathbf{z}_{it}'\gamma_{irv} + \epsilon_{irvt}.$$

The subject-level activation for the  $q$ th task is

$$a_{irvq} = \beta_{r,q} + \beta_{rvq} + s_{irq} + b_{irvq}. \quad (10)$$

Insight into this model can be gained by considering the marginal and conditional formulations. The marginal model characterizes the BOLD signal due to population task activation and subject-specific nuisance terms:  $E(y_{irvt} | \mathbf{s}_{ir}, \mathbf{b}_{irv}) = \mathbf{x}_{it}'(\beta_r + \beta_{rv}) + \mathbf{z}_{it}'\gamma_{irv}$ . The regional subject random effects  $\mathbf{s}_{ir}$  capture a baseline correlation in the BOLD signal between vertices within the same region; this is equivalent to the exchangeable covariance structure used for local correlations in Bowman et al. (2008) and Derado et al. (2010). The subject-vertex random effects  $\mathbf{b}_{ir}$  capture dependence that is a function of distance; this is similar to Bowman (2007). Additionally, the AR(3) errors capture correlations that are a function of distance in time. Let  $\psi_{t,t'}^{(irv)}$  be equal to the

corresponding element of  $\mathbf{\Psi}_{irv}$ . Let  $\mathbf{\Omega}_{r,v,v'} = \text{diag}(\mathbf{\Omega}_{r1:v,v'}, \dots, \mathbf{\Omega}_{rQ:v,v'})$ . In the marginal model, we have.

$$\text{Cov } y_{irvt}, y_{i'r'v't'} = \begin{cases} \mathbf{x}_{it}'(\mathbf{S}_r + \mathbf{B}_r)\mathbf{x}_{i't'} + \xi_{irv}^2 \psi_{t,t'}^{(irv)} & i' = i; r' = r; v' = v; \text{ any } t, t', \\ \mathbf{x}_{it}'(\mathbf{S}_r + \mathbf{B}_r \mathbf{\Omega}_{r,v,v'})\mathbf{x}_{i't'} & i' = i; r' = r; v' \neq v; \text{ any } t, t', \\ 0 & i' \neq i \text{ or } r' \neq r. \end{cases}$$

The conditional model characterizes a subject's BOLD signal:  $E(y_{irvt} | \mathbf{s}_{ir}, \mathbf{b}_{irv}) = \mathbf{x}_{it}'(\beta_r + \beta_{rv} + \mathbf{s}_{ir} + \mathbf{b}_{irv}) + \mathbf{z}_{it}'\gamma_{irv}$ . In the conditional model,  $\mathbf{s}_{ir}$  represents the deviation of the  $i$ th subject from the population regional activation  $\beta_r$ , while  $\mathbf{b}_{irv}$  represents the deviation of the  $i$ th subject from the vertex-specific activation  $\beta_{rv}$ . Given the regional subject and subject-vertex random effects, the BOLD signal for two locations,  $y_{irvt}$  and  $y_{i'r'v't'}$ , are conditionally independent.

Bowman (2007) examined a variety of variograms for volume-based fMRI data and found that the exponential variogram was most supported. Our exploration of the HCP data on the cortical surface for the Gordon networks also indicate this dependence is appropriate (see also Analysis of HCP task data section). For some random variables  $z_v$ ,  $z_{v'}$ , the exponential covariogram is defined as

$$\text{Cov } z_v, z_{v'} = \begin{cases} \lambda_0 + \lambda_1 & v = v' \\ \lambda_1 e^{-\theta_{rq}\|v-v'\|} & v \neq v'. \end{cases}$$

where  $\lambda_0$  is the nugget effect, which is equal to the micro-scale variance plus the variance due to measurement error. Our hierarchical model in fact includes a spatially varying micro-scale and measurement error variance component via  $\xi_{irv}^2$  (see (2)), so we assume the nugget effect is equal to zero. Then.

$$\mathbf{\Omega}_{rq,v,v'} = e^{-\theta_{rq}\|v-v'\|}, \quad (11)$$

where  $\theta_{rq} > 0$ .

### Fitting the STMM

Fitting the model using maximum likelihood or restricted maximum likelihood is difficult at best. For the largest parcel with two sessions, the covariance matrix is greater than  $500,000 \times 500,000$ . For computational feasibility, we first project the data to the span of the covariates. We then make the observation that the projected data follow a two-factorial crossed design. This allows us to define expected mean squares, which in turn allows us to create method of moment estimators for the variance components. The steps are summarized below and in Fig. 1. Equations and additional details are provided in Appendix B.

1. Project each subject's data onto the space spanned by the covariates, (a.6), and retain the covariates of interest. This converts the data from an  $N \times V \times T$  array to  $N \times V \times Q$ . This is equivalent to retaining the coefficients from a first-level analysis using OLS and in that respect is similar in spirit to approaches that use the output from a first-level analysis. However, in contrast to other spatial studies, we also retain the AR parameters, which will play a role in the subject-specific predictions. For each region, the projected data,  $\mathbf{d}_{irv} = [d_{irv1}, \dots, d_{irvQ}]$ , then follow a two-factor crossed design with a fixed factor (vertex) and random factor (subject).
2. For each subject and vertex, estimate the AR(p) parameters from the OLS residuals using reduced-bias estimators in Worsley et al. (2002) as in the MUMM.
3. For each subject, regularize these estimates using a biweight kernel with bandwidth determined using generalized cross validation (GCV), (a.5). This step utilizes the geodesic distance between vertices in the FreeSurfer 32 k spherical template. The use of the spherical template is discussed below. Calculate the mean square residual for the projected data from the AR parameter estimates as defined in (a.9).

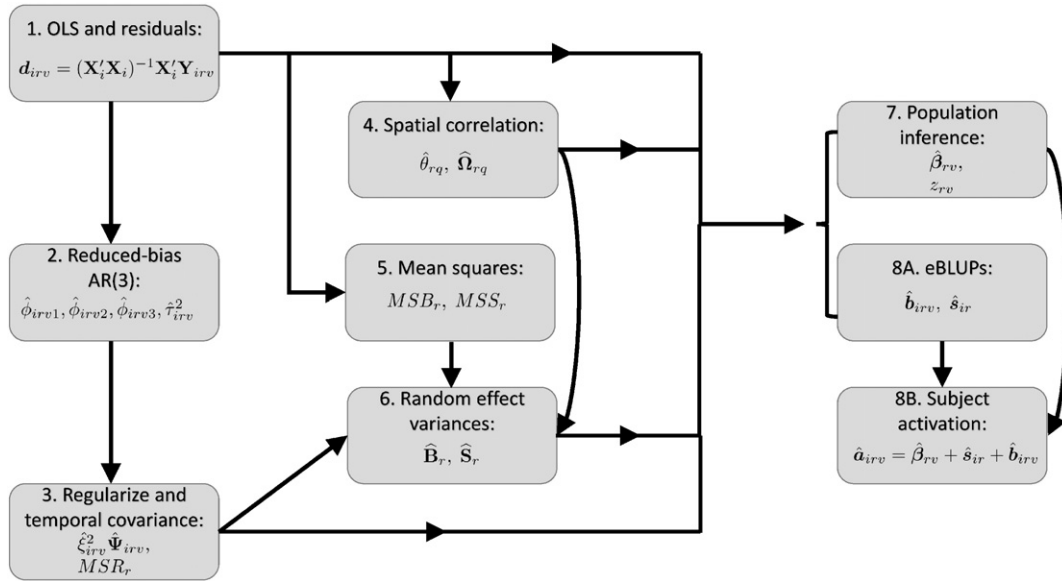


Fig. 1. Schematic overview of the model fitting steps in the STMM.

4. For each region and pooling all subjects, estimate the spatial dependence in the projected data. We create empirical covariograms from the covariance between projected data at different locations, (a.8), and estimate the spatial correlation from the empirical covariogram. We used the geodesic distance on the FreeSurfer 32 k spherical template because it represents a common coordinate system for all subjects (Fischl et al., 1999), and it was also used in Bernal-Rusiel et al. (2013). This results in a single covariance matrix for the subject-vertex random effects (rather than subject-specific). Practically, this assumes that the correlation between two nodes in one subject is the same as the correlation between two nodes in another subject. We provide evidence of appropriateness in Analysis of HCP task data section.
5. For each region and pooling all subjects, calculate the mean square from the subject-vertex random effect, (a.10), and the regional subject random effect, (a.13).
6. By equating the expectations of the mean squares to their empirical estimates, we derive estimators of the variance components. For each region, calculate estimates of the variance due to subject-vertex, (a.12), and the variance due to subject, (a.14).
7. Calculate population coefficients using generalized least squares. Note the GLS estimator in (a.16) includes the variances from the time-series errors, i.e., the within subject variance, allowing for heteroscedasticity across subjects. If the time-series errors of a subject are highly variable, the subject will have a smaller effect on the estimated population activation than subjects with comparably small variances, which can be seen by inspecting (a.16). If the regional subject and subject-vertex variances are equal to zero, this aspect is similar to Beckmann et al. (2003). Finally, use the estimated covariance to generate approximate z-statistics, (a.18).
8. Calculate eBLUPs for the regional subject random effect, (a.19), and subject-vertex random effect, (a.20), and then calculate subject-level predictions of activation. An inspection of (a.22) reveals that when the within-subject variation from  $\hat{\Psi}_{ir}$  is large relative to the between-subject variation from the subject and subject-vertex random effects, the eBLUPs are “shrunk” towards zero. Consequently, the population information dominates the subject-level estimate of activation, (a.21). Intuitively, the model leverages more population information when the measurement error is large and/or there is a small contribution of the subject and subject-vertex effects to the overall variance in the BOLD signal. Additionally, the subject-vertex random effects for a given vertex can be thought of as a weighted

average of nearby vertices with more vertices contributing to a focal vertex under high spatial dependence, (a.20).

## Simulations

We performed a simulation study to compare the performance of the MUMM and STMM in terms of subject-level prediction accuracy, power, and type 1 error rates. We created a  $2 \times 2 \times 2$  full factorial design with low and high values for  $\mathbf{S}_r$ ,  $\mathbf{B}_r$ , and  $\theta_r$  for thirty subjects with 250 simulations each. We based the simulation study design on the theory of mind task from the HCP with parameter values based on a parcel with a representative number of vertices (parcel 82 with 215 vertices; see Analysis of HCP task data section) and two runs. We used the theory of mind experiment because it involved two tasks, which was convenient for simulations examining both power and type 1 errors.

The values for parameters that varied across scenarios were  $\sigma_{s_1}^2 = \sigma_{s_2}^2 \in \{423, 1700\}$  (as defined in (7)),  $\sigma_{b_1}^2 = \sigma_{b_2}^2 \in \{9, 2346\}$  (see (8)), and  $\theta_1 = \theta_2 \in \{0.75, 0.23\}$  (see (11)). The low and high values were equal to the 0.10 and 0.90 quantiles, respectively, from the estimated variance components pooled across tasks and regions from the theory of mind experiment. We set  $\beta_{v1} = 31$  and  $\beta_{v2} = 0$  for all vertices because their difference equals the 0.95 quantile of the contrast in the theory of mind analysis (see Appendix A for the definition of a contrast). The AR parameters and innovation variance were set equal to their smoothed estimates and thus were vertex- and subject-specific. Their mean values were 0.14, 0.08, 0.07, and 29,376, respectively. The subject design matrices were equivalent to those used in the theory of mind analysis and included the tasks convolved with the canonical HRF from SPM12 plus twenty-six nuisance covariates (time-delay and dispersion derivatives for each task, piecewise linear spline with five basis functions for each session, six motion parameters from the affine registration and the interaction between each motion parameter and session; see Table S.2). Parameters for nuisance covariates were set equal to their empirical estimates. We generated the task covariates and their derivatives with respect to the temporal delay parameter and the dispersion parameter by convolving task onsets and durations with the canonical HRF in SPM12. We also included the affine registration parameters to correct for motion-induced activation. This resulted in a total of forty covariates.

Estimates of subject-level activation were much more accurate (lower MSE) in the STMM than the MUMM for all scenarios, while the

**Table 1**Prediction accuracy, power, and type 1 error rates from the STMM and MUMM for the  $2 \times 2 \times 2$  design from 250 simulations for each scenario.

Scenario			Model	MSE			Rejection rate		
$\sigma_{\sigma_q}^2$	$\sigma_{b_{rq}}^2$	$\theta_{rq}$		$a_{irv1}$	$a_{irv2}$	$a_{irv1} - a_{irv2}$	$\beta_{rv1} = 31$ (Power)	$\beta_{rv2} = 0$ (Type 1)	$\beta_{rv1} - \beta_{rv2}$ (Power)
lo	lo	lo	STMM	85.6	87.0	116.8	0.880	0.041	0.685
			MUMM	2093.9	2232.9	3445.3	0.903	0.050	0.715
hi	lo	lo	STMM	115.2	115.6	125.6	0.756	0.057	0.482
			MUMM	2092.0	2233.1	3432.3	0.759	0.051	0.503
lo	hi	lo	STMM	1026.1	945.5	1598.0	0.709	0.055	0.456
			MUMM	2091.7	2227.3	3427.0	0.663	0.054	0.421
lo	lo	hi	STMM	96.3	86.0	133.8	0.898	0.043	0.691
			MUMM	2094.6	2226.2	3438.1	0.904	0.049	0.708
hi	hi	lo	STMM	1031.4	980.3	1613.4	0.602	0.053	0.355
			MUMM	2092.1	2233.9	3439.7	0.560	0.053	0.326
hi	lo	hi	STMM	109.4	105.3	133.5	0.734	0.052	0.438
			MUMM	2094.9	2232.1	3435.7	0.738	0.054	0.459
lo	hi	hi	STMM	706.5	617.0	855.8	0.726	0.047	0.474
			MUMM	2089.1	2229.6	3435.5	0.673	0.048	0.434
hi	hi	hi	STMM	691.1	619.1	849.2	0.603	0.042	0.365
			MUMM	2095.3	2234.2	3450.2	0.557	0.042	0.334
All scenarios			STMM	482.7	444.5	678.3	0.738	0.049	0.493
			MUMM	2093.0	2231.1	3438.0	0.720	0.050	0.487

power for testing population-level hypotheses was similar and type 1 error rates near nominal  $\alpha = 0.05$  for both models (Table 1). The MSE for MUMM estimates were approximately equivalent across scenarios. The STMM estimates were least accurate when the subject and subject-vertex random effect variances were high and the spatial dependence low (hi hi lo), although still considerably more accurate than the MUMM. Intuitively, the mixed model subject-specific estimates from the STMM become more similar to the fixed effect estimates from the MUMM as the subject-vertex variance increases. When the subject-vertex random effects are small, the mixed modeling approach gains predictive power by shrinking  $\hat{b}_{irvq}$  towards zero, resulting in  $\hat{a}_{irvq}$  shrunk towards  $\hat{\beta}_{irvq}$ . When the subject-vertex variance was high, prediction from the simulations with smaller values of  $\theta_q$  (higher spatial correlation) were more accurate than those with high  $\theta_q$  (lower spatial correlation) (lo hi lo versus lo hi hi; hi hi lo versus hi hi hi). When there exists sufficient variation in the subject-random effects (e.g.,  $\sigma_{b_q}^2 = 2346$  but not  $\sigma_{b_q}^2 = 9$ ), spatial dependence can be leveraged for improved prediction by incorporating information from nearby vertices into the eBLUPs. Overall, the type-1 error rates are approximately equivalent to their nominal levels for both the STMM and MUMM and the models' powers are similar.

### Analysis of HCP task data

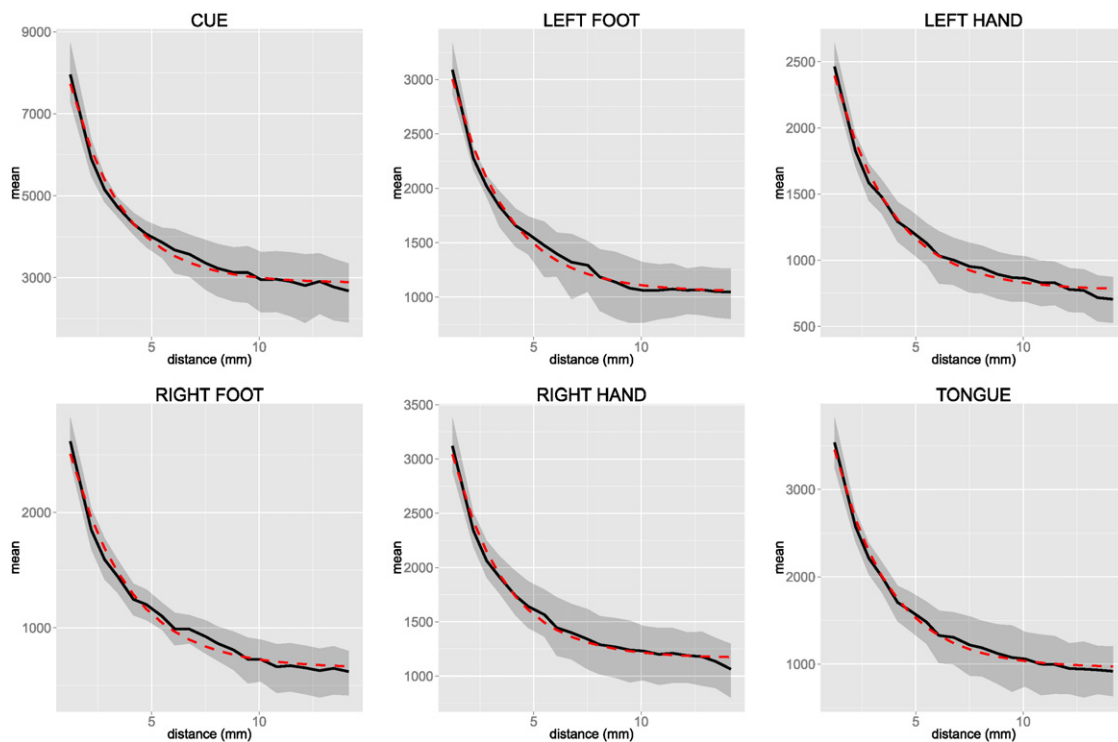
We applied the MUMM and STMM to data from a motor task involving thirty subjects in the HCP. The aim of the HCP is to characterize brain function, connectivity, and variability in healthy adults. To further this goal, the HCP is making large amounts of their unprocessed and preprocessed data publicly available (Van Essen et al., 2012). We used the minimally preprocessed data from the 100 unrelated (i.e., non-siblings) data sampler released August 5, 2014. In our analysis of the motor task, we used thirty subjects because it is a typical sample size in an fMRI experiment. This task was chosen because motor tasks are used in pre-surgical mapping (Holodny et al., 2011). Participants were presented visual cues to move their tongue, squeeze their right or left toes, tap their left or right fingers, or a generic cue. For details see Barch et al. (2013). Whole-brain data were acquired from two sessions with 284 volumes,  $2 \times 2 \times 2$  mm voxels, and 0.72 s repetition time. The minimally preprocessed data include fMRI data registered to the FreeSurfer 32 k spherical template with 2 mm full width at half maximum (FWHM) Gaussian smoothing on the surface, the end result of which is a set of approximately 30,000 time series for each cortical hemisphere and each session and each subject on a standard mesh

where the vertex indices correspond to spatially matched locations (Glasser et al., 2013). We modeled the right cerebral cortex. We concatenated the two sessions and assume the AR errors between the two sessions are independent but generated from processes with the same AR parameters. We built design matrices with the same nuisance covariates described in Simulations section. The complete motor task analysis for thirty subjects took approximately 2.5 h on a quad-core 3.60 GHz computer.

GCv resulted in a very modest amount of smoothing of the AR parameters. A bandwidth of 1.5 mm was selected, which resulted in a median of six neighbors with positive weights, see Fig. S.1, and the smoothed and unsmoothed parameters are very similar (an example is provided in Fig. S.2).

We defined spatially distinct regions according to the cortical parcellation in Gordon et al. (2016), which is based on correlations between the BOLD signal at each vertex in resting-state fMRI data. The Gordon parcellation for the right cerebral cortex comprises 172 spatially disjoint networks defined in the FreeSurfer 32 k template. A total of 8509 out of 29,716 vertices are unclassified. Each unclassified vertex was assigned to the parcel containing the closest classified vertex. When a vertex was equidistant from classified vertices in two different parcels, then the vertex was assigned based on which parcel contained the second closest vertex. This resulted in a unique classification of all vertices. The revised parcels range in size from 29 to 986 vertices. This expansion of the Gordon parcels was necessary to allow predictions at all locations. Fig. 2 depicts the empirical and fitted covariograms for a typical parcel (324 vertices, median size from parcels with non-zero subject-vertex random effect for all tasks). The exponential covariogram provides a good fit to the empirical covariograms.

In a randomly chosen subject, the STMM estimate of activation has an overall smoother appearance than the MUMM with more clearly delineated areas of activation (Fig. 3). Note that  $\hat{b}_{irvq}$  is smoothed by  $\hat{\Omega}_{rq}$  according to  $\hat{\theta}_{rq}$ , see (11) and (a.20). The MUMM map has a speckled appearance as small areas not associated with the motor cortex also appear to be active, whereas activated regions appear to be better delineated in the STMM. This is due to higher variability in the estimates from the MUMM, since the MUMM does not incorporate population or spatial information. In addition to portions of the motor cortex typically associated with the left hand, this subject also has activation in the retrosplenial cortex (Brodmann areas 29 and 30), which is absent from the population. Thus the STMM utilizes population information while also incorporating subject-specific features.



**Fig. 2.** Empirical covariogram (black line) and its 95% CI bootstrapped on subject (gray) and the fitted exponential covariogram (dashed red) for an example parcel (median size from parcels with  $\sigma_{bq}^2 > 0$  for all  $q = 1, \dots, 6$ ). Here,  $\theta_{rq} = 0.41, 0.39, 0.38, 0.36, 0.41$ , and  $0.39$  for the cue, left foot, left hand, right foot, right hand, and tongue tasks.

In contrast to the subject-specific estimates of activation, the test statistics of the population contrasts are very similar between the MUMM and the STMM (Fig. 4). Both models clearly delineate the motor cortex associated with the left-hand (e.g., Meier et al., 2008).

We also conducted an analysis of the theory of mind task from the HCP for the ninety-eight subjects included in the unrelated 100 data sampler released August 5, 2014. Theory of mind refers to the ability to intuit another person's actions or feelings. In the HCP experiment, subjects in an fMRI scanner viewed cartoons that either depicted shapes acting in human-like ways (e.g., a large triangle leading a smaller triangle out of a maze) or in random ways, which were the “mentalizing” (hereafter, xMental) and “random” (xRandom) tasks, respectively. For details of the experimental paradigm see Barch et al. (2013). Whole-brain data were acquired from two sessions with 274 volumes. The analysis was conducted in the same manner as the motor task, with the exception that we used all ninety-eight subjects contained in the sampler to demonstrate the scalability of the STMM. In a randomly chosen subject, the STMM estimate of activation again has a smoother appearance than the MUMM (Fig. S.3). Note that the population effect sizes are much smaller than the single-subject effect sizes, which is likely due to high variation in the locations of activation between subjects. In the population-level analysis, we found the surprising result that the standard errors for the MUMM tended to be smaller than the STMM, such that the MUMM had higher test statistics (Fig. S.4). Since our simulation studies indicate the MUMM preserves the nominal type-1 error rate, this likely reflects a situation in which the MUMM is more powerful than the STMM for group analysis.

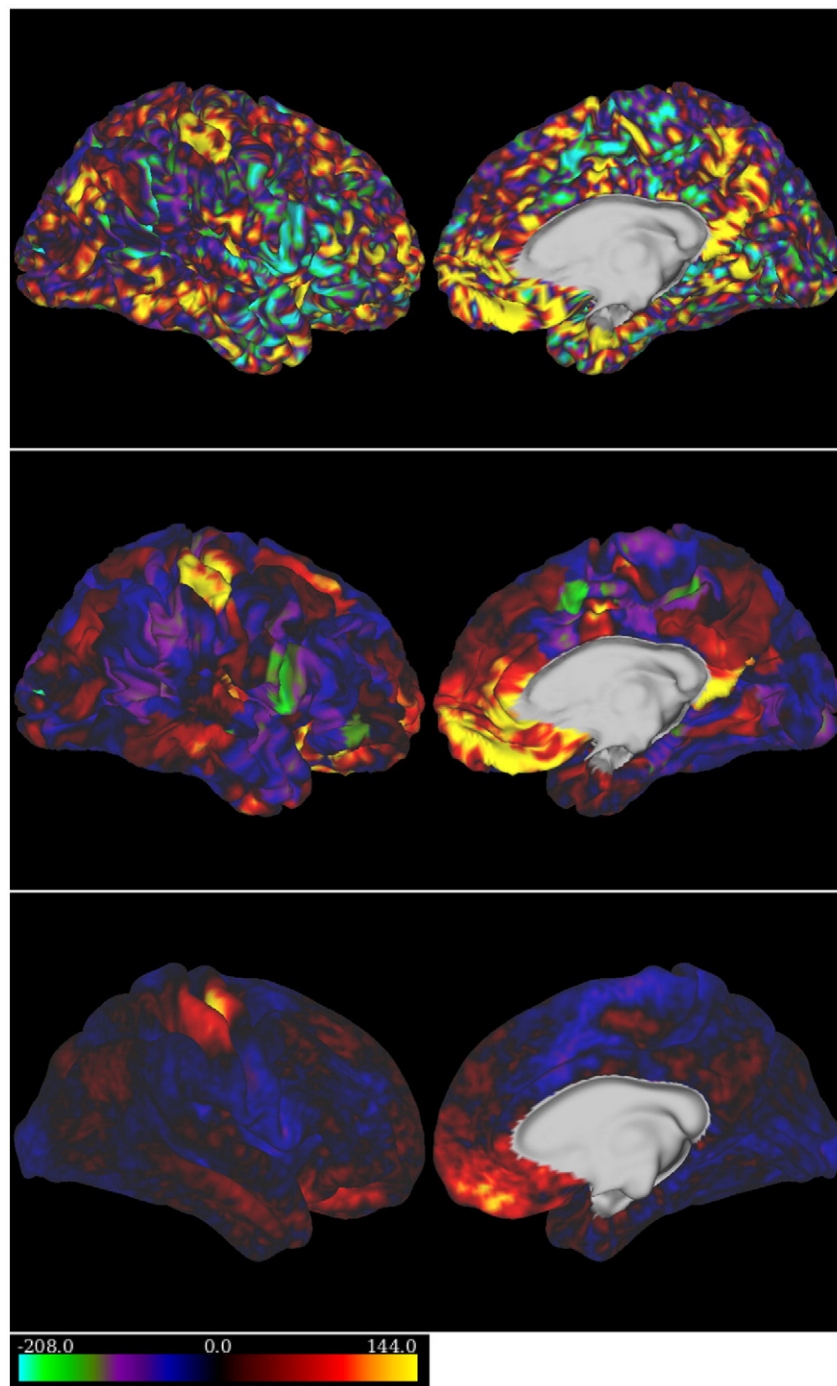
Even in cases in which there is high variability between subjects, the STMM may still improve estimation by leveraging the spatial information. An extreme example is provided in Fig. S.5, in which we chose a subject whose activation patterns differed markedly from the population. The STMM activation map appears to be a smoothed version of the MUMM. In Fig. S.5, one can see that the subject's nearby points are used to inform estimates of a given vertex.

## Discussion

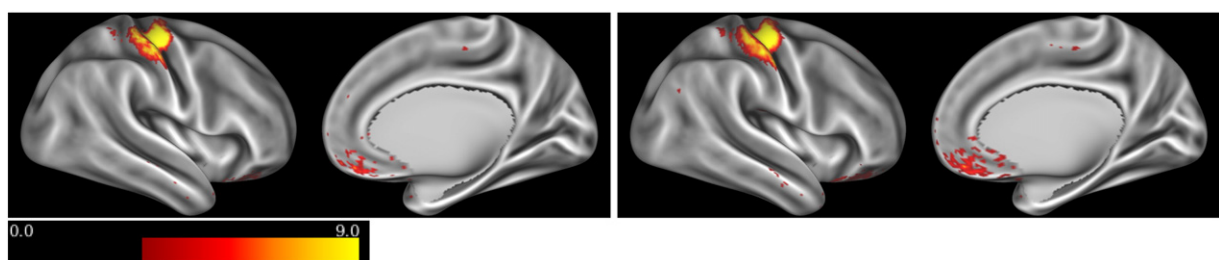
We present a unified model for multi-subject fMRI studies that utilizes spatial eBLUPs to improve prediction of subject-level activation. The model accounts for spatial dependence in two ways: with the regional subject random effect, which creates a baseline correlation between all vertices in a parcel, and the subject-vertex random effect, which includes a covariance function based on geodesic distance. The model also leverages population information as the subject-specific activation is shrunk towards the population coefficients when the variance between subjects decreases and/or the variance within subject increases. The STMM is scalable to multi-subject fMRI because we project the data to the span of the covariates, develop method-of-moments estimators that account for both spatial and temporal dependence, and model parcels independently. Compared with the MUMM, subject activation maps from the STMM have lower MSE in simulations and appear smoother and less noisy in the HCP tasks.

Our modeling approach may allow for higher spatial precision than typical in group studies because we utilize cortical sheet registration with a minimal amount of smoothing (2 mm), and then our predictions contain spatial smoothing in which the smoothing parameter is allowed to adapt to each parcel. In volume-based analyses, one rule-of-thumb recommends smoothing with a Gaussian kernel with FWHM equal to at least three times the voxel size (Pettersson et al., 1999). Volume-based group studies often use a large amount of smoothing to increase the overlap between subjects. This could result in too much smoothing and render individual subject maps less useful. In general, cortical surface analyses can be conducted with higher resolution than volume-based analyses (Fischl et al., 1999) and less smoothing is required (Glasser et al., 2013). The Freesurfer 32k template offers a convenient space in which to conduct group analysis, and then the vertices can be mapped back to a subject's coordinate space (Fig. 3). In our model, the subject-vertex random effects are equivalent to kriging predictors (e.g., (4.28) in Cressie and Wikle, 2011). Thus the subject-vertex random effect predictions are spatially smoothed according to the exponential covariogram estimated from the data rather than a





**Fig. 3.** MUMM (top) and STMM (middle) estimates of the contrast between the left-hand finger tap versus other tasks for a randomly selected subject (123,925) (overlaid on the subject's midthickness cortical surface) and the population contrast of the STMM from thirty subjects (bottom) (overlaid on the midthickness group average surface) in the right cerebral cortex.



**Fig. 4.** Point-wise t-statistics ( $df = 29$ ) (a.4) and approximate z-statistics (a.18) of the population contrast for the MUMM (left) and STMM (right), respectively, between the left-hand versus other motor tasks in the right cerebral cortex from thirty subjects (thresholded at  $p < 0.001$  uncorrected for multiple comparisons) (overlaid on inflated group average surface).



rule-of-thumb. Additionally, the covariograms are estimated for each region, allowing the smoothing to adapt to regional differences. A recent study explored optimizing pre-processing for pre-surgical mapping in fMRI, including spatial smoothing (Stevens et al., 2010). The STMM offers an alternative approach that includes data-based smoothing for each region and each task, where information across subjects is pooled under the assumption that the spatial dependence for a region and task is equivalent for all subjects.

Our findings suggest that the STMM may be useful in clinical neuroimaging, although for some applications additional complications would need to be addressed. The proposed method could be applied without modification when predicting activation maps for subjects from a relatively homogeneous population, which may be the case for identifying targets for tMS or transcranial direct current stimulation. In studies involving a mixture of control and treatment groups, the STMM could be extended by including an indicator variable for treatment, its interaction with the fixed effects, and its interaction with the random effects. In applications to pre-surgical mapping for tumor patients, applying the STMM to a mixture of controls and patients may be problematic because the underlying population model may not apply (namely, that the regional subject random effects as well as the subject-vertex random effects are equal in distribution across subjects). One direction to investigate is combining a single patient's data with baseline control subjects for a given site and scanner. This information could be used to decrease damage to healthy tissue in the path of the abnormal region. Even for an individual's abnormal region, Fig. 3 as well as Figs. S.3 and S.5 indicate that the STMM captures unique individual features. Moreover, the Gordon parcellation provides a framework in which population information can be leveraged for regions that are homogeneous across control and patient populations, whereas subject variation is generally preserved for regions with high variability across subjects.

Three potential issues are our choice of parcellation including assigning unclassified vertices (nearly 1/3 of all vertices), accuracy of registration across subjects, and the assumption that random effects are independent between parcels. A poor parcellation would result in greater heterogeneity in the activation of vertices within each parcel, which would tend to increase  $\hat{\mathbf{B}}_r$ . Poor registration would tend to increase the variance between subjects, resulting in increases in  $\hat{\mathbf{S}}_r$  and/or  $\hat{\mathbf{B}}_r$ . As  $\hat{\mathbf{S}}_r$  and  $\hat{\mathbf{B}}_r$  increase, the STMM estimate becomes more similar to the OLS estimate of subject-specific activation, as seen in (a.22). Thus the accuracy of the parcellation and registration affects how much we gain from the STMM relative to first-level OLS. Overall, a poor parcellation and/or poor registration would result in estimates more similar to the MUMM, since the MUMM estimators in (a.1) are generally similar to OLS. Note population inference is robust to the parcellation, since  $\hat{\beta}_{vr}$  and the vertex-wise z-statistics from the STMM are very similar to the MUMM, which does not use the parcellation. Regarding the independence of random effects across parcels, it should be noted that this assumption does not affect the type-1 error rate of vertex-wise tests (Table 1). However, incorporating correlations between regions could result in additional improvements in the accuracy of subject-level predictions. Future research should look at incorporating a model for regional or long-range correlations, although this would involve new computational challenges.

We did not find an increase in the statistical power of tests of population fixed effects at the vertex-level of analysis, and our results suggest that there is little reason to use the STMM if a researcher is exclusively interested in vertex-level group inference. This was initially surprising to us because it appears to be contrary to the spatial models in Bernal-Rusiel et al. (2013) and Bowman (2007). However, Hyun et al. (2014) noted that when a separate activation parameter is estimated for every location and the covariates are constant across space, spatial modeling may not improve estimation of population effects. This is easy to see for the special case of a block design where all subjects have the same design matrix and the data are temporally pre-whitened. In this case, the GLS estimator of the population fixed effects is equal to the OLS estimator

that is naive to spatial dependence; hence the test statistics are equal, see (a.23). Surprisingly, we actually found the STMM group analysis resulted in smaller test statistics than the MUMM for the theory of mind study (Fig. S.4). The theory of mind experiment involves smaller population effect sizes than the motor task or simulation studies. Note the subject maps from the theory of mind experiment incorporate spatial smoothing (Figs. S.3 and S.4) which should generally result in a lower MSE. This interpretation is supported by the results of our simulations in which the subject-level predictions from the STMM were still considerably more accurate than the MUMM when the between subject variance was high, and additionally, high spatial dependence decreased the MSE in the STMM (e.g., “hi hi lo” versus “hi hi hi” scenarios, Table 1). With respect to the vertex-level tests of group activation, future research could examine whether eliminating the dimension reduction step, see (a.6), and using maximum likelihood estimation could substantively improve the power of vertex-level tests. However, as previously noted, ML estimators are problematic for our large data application because the likelihood includes a  $500,000 \times 500,000$  dense covariance matrix. A more promising avenue may be to develop estimation techniques that utilize spatial information in the mean structure. In structural MRI, the weighted likelihood approach in the MARM led to increases in statistical power, since the spatial information from nearby points is explicitly part of estimating the effect at a location (Zhu et al., 2014).

Although power was not increased at the vertex level, a separate question is whether the STMM may improve power in a group analysis of a region of interest. A common approach to ROI analysis is to average the signal within the region for each subject and then conduct the traditional two-level analysis (Brett et al., 2002). Averaging the signal within an ROI allows one to test the hypothesis that the activation across the ROI is non-zero. This may result in a loss of information when the signal varies across the ROI. In the STMM, one can estimate activation at a Gordon parcel using a different parameterization of the STMM (e.g., estimating  $\beta_r$  and using sum-to-zero contrasts on  $\beta_{vr}$ , then testing whether  $\beta_r = 0$ ), or more generally, one can test the significance of any region by approximating the distribution of a linear combination of the fixed effects using the covariance matrix constructed in the STMM. Since activation is allowed to vary by vertex, this may be more powerful than averaging when the signal is heterogeneous across the ROI.

## Conclusion

We present a spatiotemporal model for multisubject fMRI that improves subject-specific prediction by leveraging population information (utilizing similarities across subjects) and leveraging spatial information (utilizing similarities between a subject's nearby locations). In general, the predictive gains from the STMM should be greater when activation patterns are similar across subjects and/or when there is high spatial dependence between vertices. Matlab code implementing the model for the HCP motor task and simulated data is available at <http://www.benjaminrisk.com/software>.

## Acknowledgments

Data were provided (in part) by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. BBR was partially supported by the NSF grant DMS-1127914 to the Statistical and Applied Mathematical Science Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. DSM was supported by a Xerox PARC Faculty Research Award and NSF grant DMS-1455172.

## Appendix A. Additional information on estimating the MUMM

Let  $\hat{\xi}_{iv}^2$  and  $\hat{\Psi}_{iv}$  denote estimates of the unconditional variance and correlation matrix from (2) (in this section these are the non-smoothed estimators). Let  $\mathbf{X}_i = [\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}]'$ ,  $\mathbf{Z}_i = [\mathbf{z}_{i1}, \dots, \mathbf{z}_{iT}]'$ , and define  $\mathbf{X}_i^* = [\mathbf{X}_i, \mathbf{Z}_i]$ . Let  $\hat{\mathbf{a}}_{iv}$  be an estimate of  $\mathbf{a}_{iv} = [a_{iv1}, \dots, a_{ivq}]'$ . Similarly define  $\hat{\gamma}_{iv}$ . Then define the MUMM estimator of subject-level activation and nuisance coefficients:

$$\begin{bmatrix} \hat{\mathbf{a}}_{iv}^{MUMM} \\ \hat{\gamma}_{iv}^{MUMM} \end{bmatrix} = (\mathbf{X}_i' \hat{\Psi}_{iv}^{-1} \mathbf{X}_i^*)^{-1} \mathbf{X}_i' \hat{\Psi}_{iv}^{-1} \mathbf{Y}_{iv}. \quad (\text{a.1})$$

Next, define the estimator of the population coefficients as

$$\hat{\beta}_v^{MUMM} = \frac{1}{N} \sum_{i=1}^N \hat{\mathbf{a}}_{iv}^{MUMM}. \quad (\text{a.2})$$

An estimator of the variance of (a.2) is

$$\widehat{\text{Cov}} \hat{\beta}_v^{MUMM} = \frac{1}{N(N-1)} \sum_{i=1}^N (\hat{\mathbf{a}}_{iv}^{MUMM} - \hat{\beta}_v^{MUMM}) (\hat{\mathbf{a}}_{iv}^{MUMM} - \hat{\beta}_v^{MUMM})'. \quad (\text{a.3})$$

We are typically interested in whether vertices are differentially activated in one task versus others. Contrasts are used to test hypotheses of these types. Let  $\hat{\beta}_v \in \mathbb{R}^Q$ , and let  $\mathbf{c} = [c_1, \dots, c_Q]$ . Suppose there are  $Q_j$  tasks for which we are interested in positive activation and  $Q_k$  tasks that are being used to control for unwanted activation. Then we might consider defining a contrast as  $c_q = 1/Q_j$  for  $q$  corresponding to tasks of interest and  $c_q = -1/Q_k$  for  $q$  corresponding to control tasks. For example, in analyzing activation by a left-hand finger-tapping task in the HCP motor task experiment, we have  $c_q = 1$  for the left finger-tapping task and  $c_q = -1/5$  for the cue, left foot, right foot, right hand, and tongue tasks. Then define

$$t_v = \frac{\mathbf{c}' \hat{\beta}_v}{\sqrt{\mathbf{c}' \widehat{\text{Cov}}(\hat{\beta}_v) \mathbf{c}}}. \quad (\text{a.4})$$

Under the null hypothesis that the contrast is equal to zero and normality assumptions, this statistic is  $t$ -distributed with  $N-1$  degrees of freedom.

## Appendix B. Estimating the STMM

### B.1. Smoothing the AR(p) parameters

Estimation of the subject- and vertex-specific variance and AR parameters results in  $4NV$  parameters. We regularize these parameters using a single bandwidth chosen using a modified GCV in which the estimates are treated as data. GCV approximates leave-one-out cross-validation using effective degrees of freedom (Hastie et al., 2009). We use the biweight kernel:  $K_{ih}(x) = \frac{15}{16h_i} \{1 - (x/h_i)^2\}^2 1_{|x| < h_i}$ . Let  $\hat{\xi}_{irv}^{*2}$  be standardized estimates (mean zero and unit variance across voxels) and similarly define  $\hat{\phi}_{irv\ell}^*$ ,  $\ell = 1, \dots, 3$ . Let  $\|\mathbf{v} - \mathbf{v}'\|$  denote the geodesic distance between the two vertices on the FreeSurfer 32 k spherical template. This is equivalent to the great circle distance using spherical coordinates. For each subject, we find the bandwidth that minimizes

$$\text{MSE}_i(h) = \frac{\sum_{v=1}^V \left\{ \left( \hat{\xi}_{irv}^{*2} - \hat{\xi}_{irv}^{\dagger 2} \right)^2 + \sum_{\ell=1}^3 \left( \hat{\phi}_{irv\ell}^* - \hat{\phi}_{irv\ell}^{\dagger} \right)^2 \right\}}{\left( 1 - \frac{1}{V} \sum_{v=1}^V K_{ih}(0)/w_{ih}(v) \right)^2} \quad (\text{a.5})$$

where  $\hat{\xi}_{irv}^{\dagger 2} = \sum_{v'=1}^V K_{ih}(\|\mathbf{v} - \mathbf{v}'\|) \hat{\xi}_{irv'}^{*2} / w_{ih}(v)$  with  $w_{ih}(v) = \sum_{v'=1}^V K_{ih}(\|\mathbf{v} - \mathbf{v}'\|)$  and similarly define  $\hat{\phi}_{irv\ell}^{\dagger}$ .

### B.2. Hierarchical formulation of the STMM

For computational reasons, we project the data  $\mathbf{Y} \in \mathbb{R}^{NVT}$  onto the space spanned by the covariates. Define  $\mathbf{X}_i^* = [\mathbf{X}_i, \mathbf{Z}_i]$  and  $\mathbf{K}_i^* = \mathbf{X}_i^* (\mathbf{X}_i^{*'} \mathbf{X}_i^*)^{-1}$ . Then let  $\mathbf{K}_i$  comprise the first  $Q$  columns of  $\mathbf{K}_i^*$ . Then define

$$\mathbf{d}_{irv} = \mathbf{K}_i' \mathbf{Y}_{irv}. \quad (\text{a.6})$$

Let  $\mathbf{d}_{ir} = [\mathbf{d}_{ir1}', \dots, \mathbf{d}_{irV'}']'$ ; let  $\mathbf{d}_r = [\mathbf{d}_{r1}', \dots, \mathbf{d}_{rN'}']'$ ; let  $\mathbf{s}_r = [\mathbf{s}_{r1}', \dots, \mathbf{s}_{rN'}']'$ ; let  $\mathbf{b}_r = [\mathbf{b}_{r1}', \dots, \mathbf{b}_{rN'}']'$ ; let  $\mathbf{k}_{irv} = \mathbf{K}_i \mathbf{e}_{irv}$ . Let  $\{c \mathbf{k}_{irv}\}_{v=1}^{V_r}$  indicate stacked  $\mathbf{k}_{ir1}, \dots, \mathbf{k}_{irV_r}$ . Then

$$\begin{aligned} \mathbf{d}_r &= (\mathbf{I}_N \otimes \mathbf{1}_{V_r} \otimes \mathbf{I}_Q) \beta_r + (\mathbf{I}_N \otimes \mathbf{I}_{V_r} \otimes \mathbf{I}_Q) \beta_r \\ &\quad + (\mathbf{I}_N \otimes \mathbf{1}_{V_r} \otimes \mathbf{I}_Q) \mathbf{s}_r + (\mathbf{I}_N \otimes \mathbf{I}_{V_r} \otimes \mathbf{I}_Q) \mathbf{b}_r + \left\{ c \{ \mathbf{k}_{irv} \}_{v=1}^{V_r} \right\}_{i=1}^N. \end{aligned}$$

Note that  $\mathbf{d}_{irv}$  is equivalent to the OLS estimate of subject-level activation. Here, we are treating  $\mathbf{d}_{irv}$  as a dimension-reducing transformation of  $\mathbf{Y}_{irv}$ , and we are keeping track of the transformed error  $\mathbf{e}_{irv}$ . Put another way, consider the hierarchical formulation of the STMM:

$$y_{irvt} = \mathbf{x}_{it}' \mathbf{a}_{irv} + \mathbf{z}_{it}' \gamma_{irv} + e_{irvt}$$

and

$$a_{irvq} = \beta_{r \cdot q} + \beta_{rvq} + s_{irvq} + b_{irvq}, q = 1, \dots, Q. \quad (\text{a.7})$$

Then  $a_{irvq}$  is different from  $d_{irvq} = \beta_{r \cdot q} + \beta_{rvq} + s_{irvq} + b_{irvq} + k_{irvq}$  where  $k_{irvq}$  is the  $q$ th element of the transformed error,  $\mathbf{k}_{irv}$ .

### B.3. Estimating the spatial dependence

Now we consider estimating the spatial dependence of  $\mathbf{b}_{ir}$ . Let  $z_v$ ,  $v = 1, \dots, V$  be arbitrary, spatially indexed random variables. Define the population variogram

$$\nu(z_v, z_{v'}) = E \{ (z_v - E z_v) - (z_{v'} - E z_{v'}) \}^2.$$

Under stationarity,  $\nu(z_v, z_{v'}) = \nu(z_{v_k}, z_{v_l})$  for  $\|\mathbf{v}_i - \mathbf{v}_j\| = \|\mathbf{v}_k - \mathbf{v}_l\|$ . Assume for the moment that  $E z_v$  is known. Then the empirical variogram is

$$\hat{\nu}(h) = \frac{1}{N_h} \sum_{\{v, v'\}: \|\mathbf{v} - \mathbf{v}'\| \in (h - \eta, h + \eta]} \{ (z_v - E z_v) - (z_{v'} - E z_{v'}) \}^2,$$

where  $2\eta$  represents the bin width and  $N_h$  is the number of pairs with distance in  $(h - \eta, h + \eta]$ .

In our model, the variance of  $d_{irvq}$  is not stationary due to  $\hat{\xi}_{irv}^2 \mathbf{K}_i' \Psi_{irv} \mathbf{K}_i$  and hence  $\|\mathbf{v}_i - \mathbf{v}_j\| = \|\mathbf{v}_k - \mathbf{v}_l\|$  does not imply  $\nu(d_{irvq}, d_{irv'q}) = \nu(d_{irv_kq}, d_{irv_lq})$ . We propose

$$\hat{\delta}_r(h) = \frac{1}{(N-1)N_h} \sum_{i=1}^N \sum_{\{v, v'\}: \|\mathbf{v} - \mathbf{v}'\| \in (h - \eta, h + \eta]} \{ d_{irvq} d_{irv'q} - \bar{d}_{rvq} \bar{d}_{rv'q} \} \quad (\text{a.8})$$

where  $\bar{d}_{rvq} = \sum_{i=1}^N d_{irvq}/N$ . Then note

$$\begin{aligned} E\hat{\delta}_r(h) &= \frac{1}{(N-1)N_h} \sum_{i=1}^N \sum_{\{v,v'\}: \|v-v'\| \in (h-\eta, h+\eta]} \left\{ \sigma_{sq}^2 + \sigma_{brq}^2 \mathbf{\Omega}_{rq;vv'} \right. \\ &\quad \left. + (\beta_{r,q} + \beta_{rvq}) (\beta_{r,q} + \beta_{rvq}) \right. \\ &\quad \left. - \left( \frac{1}{N} \sigma_{sq}^2 + \frac{1}{N} \sigma_{brq}^2 \mathbf{\Omega}_{rq;vv'} + (\beta_{r,q} + \beta_{rvq}) (\beta_{r,q} + \beta_{rvq}) \right) \right\} \\ &= \sigma_{sq}^2 + \sigma_{brq}^2 \frac{1}{N_h} \sum_{\{v,v'\}: \|v-v'\| \in (h-\eta, h+\eta]} e^{-\theta_{rq} \|v-v'\|}. \end{aligned}$$

We evaluate  $\hat{\delta}_r(h)$  over a suitable range of  $h$  and find the parameters that minimize the sum of squared errors with the function  $\lambda_{0q} + \lambda_{1q} \exp(-\theta_{rq}h)$  using `fmincon` in Matlab where only  $\hat{\theta}_{rq}$  is retained for additional analysis (estimates of variance components described below). We use the geodesic distance between vertices in the FreeSurfer 32 k spherical template in our simulations and data analyses.

#### B.4. Estimating the variance of the subject-vertex random effect

Let  $\hat{\xi}_{irv}^2$  and  $\hat{\Psi}_{irv}$  correspond to the smoothed estimates of the AR parameters. To develop estimators for the variance components, it is convenient to define the  $Q \times Q$  matrix

$$MSR_r = \frac{1}{NV_r} \sum_{i=1}^N \sum_{v=1}^{V_r} \hat{\xi}_{irv}^2 \mathbf{K}_i' \hat{\Psi}_{irv} \mathbf{K}_i, \quad (\text{a.9})$$

which will serve a role similar to the mean square residual in ANOVA decompositions. Then define

$$\bar{\mathbf{d}}_{rv} = \frac{1}{N} \sum_{i=1}^N \mathbf{d}_{irv}, \quad \bar{\mathbf{d}}_{ir} = \frac{1}{V_r} \sum_{v=1}^{V_r} \mathbf{d}_{irv}, \quad \bar{\mathbf{d}}_{r.} = \frac{1}{NV_r} \sum_{i=1}^N \sum_{v=1}^{V_r} \mathbf{d}_{irv}.$$

Towards estimating  $\mathbf{B}_r$ , consider a measure of the mean square due to the interaction between subject and vertex:

$$MSB_r = \frac{1}{(N-1)(V_r-1)} \sum_{i=1}^N \sum_{v=1}^{V_r} (\mathbf{d}_{irv} - \bar{\mathbf{d}}_{ir} - \bar{\mathbf{d}}_{rv} + \bar{\mathbf{d}}_{r.}) (\mathbf{d}_{irv} - \bar{\mathbf{d}}_{ir} - \bar{\mathbf{d}}_{rv} + \bar{\mathbf{d}}_{r.})'. \quad (\text{a.10})$$

Here,  $MSB_r$  is a  $Q \times Q$  matrix. We assume the random effects associated with each task are independent and thus only calculate the diagonal elements.

Let

$$w_{rq} = \sum_{v=1}^{V_r} \sum_{v'=1}^{V_r} \mathbf{\Omega}_{rq;vv'}$$

and  $\mathbf{W}_r = \text{diag}(w_{r1}, \dots, w_{rQ})$ . It can be shown (Web Supplement Section 4) that

$$EMSB_r = \left( \frac{V_r}{V_r-1} \mathbf{I}_Q - \frac{1}{V_r(V_r-1)} \mathbf{W}_r \right) \mathbf{B}_r + \frac{1}{NV_r} \sum_{i=1}^N \sum_{v=1}^{V_r} \hat{\xi}_{irv}^2 \mathbf{K}_i' \hat{\Psi}_{irv} \mathbf{K}_i. \quad (\text{a.11})$$

In the univariate case under zero spatial and temporal dependence with  $\mathbf{Z}_i = \mathbf{0}$ ,  $\mathbf{X}_i = \mathbf{1}_T$  for all  $i$ , and  $\xi_{irv}^2 = \sigma_a^2$  for all  $i, v$ , we have  $EMSB_r = \sigma_b^2 + \frac{1}{T} \sigma_a^2$ . Then the model resembles the classic two-way crossed mixed effects model (e.g., p.123, Searle et al., 1992).

We define the estimator

$$\hat{\mathbf{B}}_r = \left( \frac{V_r}{V_r-1} \mathbf{I}_Q - \frac{1}{V_r(V_r-1)} \mathbf{W}_r \right)^{-1} (MSB_r - MSR_r). \quad (\text{a.12})$$

As is the case for method of moments estimators in other mixed models,  $\hat{\sigma}_{brq}^2$  can be negative, in which case we replace it with 1e-06. This introduces a slight bias into our estimator but decreases its mean squared error (p.130, Searle et al., 1992).

Estimators of the subject variance components are described in Appendix B.5.

#### B.5. Estimating the variance of the regional subject random effect

Towards estimating  $\mathbf{S}_r$ , consider the mean square due to subject:

$$MSS_r = \frac{1}{N-1} \sum_{i=1}^N \sum_{v=1}^{V_r} (\bar{\mathbf{d}}_{ir} - \bar{\mathbf{d}}_{r.}) (\bar{\mathbf{d}}_{ir} - \bar{\mathbf{d}}_{r.})'. \quad (\text{a.13})$$

It can be shown

$$E \, MSS_r = V_r \mathbf{S}_r + \frac{1}{V_r} \mathbf{W}_r \mathbf{B}_r + \frac{1}{NV_r} \sum_{i=1}^N \sum_{v=1}^{V_r} \hat{\xi}_{irv}^2 \mathbf{K}_i' \hat{\Psi}_{irv} \mathbf{K}_i.$$

In the univariate case under no spatial or temporal dependence, the result again parallels the expected mean square of a mixed effects model for a two-factorial crossed design. Namely, if  $\mathbf{\Omega}_r = \mathbf{I}_{V_r}$  and  $\xi_{irv}^2 \mathbf{K}_i' \hat{\Psi}_{irv} \mathbf{K}_i = \frac{1}{T} \sigma_a^2$ , then  $EMSS_r = V_r \sigma_{sr}^2 + \sigma_{br}^2 + \frac{1}{T} \sigma_a^2$ .

We propose the estimator

$$\hat{\mathbf{S}}_r = \frac{1}{V_r} MSS_r - \frac{1}{V_r^2} \mathbf{W}_r \hat{\mathbf{B}}_r - \frac{1}{V_r} MSR_r. \quad (\text{a.14})$$

As in the case of  $\hat{\mathbf{B}}_r$ , if some  $\hat{\sigma}_{sq}^2 < 0$ , then we replace it with 1e-06.

#### B.6. Estimators of population activation

We parameterize the model with  $\beta_{r.} = \mathbf{0}_Q$  such that the estimates of the vertex effects incorporate the mean activation of the region. Define

$$\hat{\mathbf{\Sigma}}_{ir} = \mathbf{1}_{V_r}' \mathbf{1}_{V_r}' \hat{\mathbf{\Sigma}}_r + \hat{\mathbf{\Omega}}_r + \text{diag}(\mathbf{K}_i' \hat{\Psi}_{i1r} \mathbf{K}_i, \dots, \mathbf{K}_i' \hat{\Psi}_{iV_r} \mathbf{K}_i). \quad (\text{a.15})$$

Let  $\hat{\mathbf{\Sigma}}_r = \text{diag}(\hat{\mathbf{\Sigma}}_{1r}, \dots, \hat{\mathbf{\Sigma}}_{Nr})$ . Define the eBLUES (empirical best linear unbiased estimators) of the fixed effects given the transformed data  $\mathbf{d}_r$  as

$$\hat{\beta}_r = \left\{ (\mathbf{1}_N' \otimes \mathbf{I}_{V_r} \otimes \mathbf{I}_Q) \hat{\mathbf{\Sigma}}_r^{-1} (\mathbf{1}_N \otimes \mathbf{I}_{V_r} \otimes \mathbf{I}_Q) \right\}^{-1} (\mathbf{1}_N' \otimes \mathbf{I}_{V_r} \otimes \mathbf{I}_Q) \hat{\mathbf{\Sigma}}_r^{-1} \mathbf{d}_r$$

which is equivalent to

$$\hat{\beta}_r = \left\{ \sum_{i=1}^N \hat{\mathbf{\Sigma}}_{ir}^{-1} \right\}^{-1} \sum_{i=1}^N \hat{\mathbf{\Sigma}}_{ir}^{-1} \mathbf{d}_{ir} \quad (\text{a.16})$$

and we have

$$\widehat{\text{Cov}} \hat{\beta}_r = \left\{ \sum_{i=1}^N \hat{\mathbf{\Sigma}}_{ir}^{-1} \right\}^{-1}. \quad (\text{a.17})$$

For inference, we construct z-statistics from

$$\hat{\beta}_r \left( \text{diag} \widehat{\text{Cov}}(\hat{\beta}_r) \right)^{-1/2}, \quad (\text{a.18})$$

which is approximately multivariate normal with unit variances.

### B.7. Estimating activation

The best linear unbiased predictor (BLUP) for a random effect is equal to its conditional mean given the data (e.g., p.315 of McCulloch et al., 2008). Using the parameterization  $\hat{\beta}_{rq} = 0$  (Appendix B.6), define the eBLUPs for the transformed data  $\mathbf{d}_{ir}$  as

$$\hat{\mathbf{s}}_{ir} = \hat{\mathbf{S}}_r \left( \mathbf{1}'_{V_r} \otimes \mathbf{I}_Q \right) \hat{\mathbf{S}}_{ir}^{-1} \left( \mathbf{d}_{ir} - \hat{\beta}_r \right) \quad (\text{a.19})$$

and

$$\hat{\mathbf{b}}_{ir} = \hat{\mathbf{Q}}_r \hat{\mathbf{S}}_{ir}^{-1} \left( \mathbf{d}_{ir} - \hat{\beta}_r \right). \quad (\text{a.20})$$

Finally we define the predictors of subject-level activation as

$$\hat{\mathbf{a}}_{ir}^{STMM} = \hat{\beta}_{rv} + \hat{\mathbf{s}}_{ir} + \hat{\mathbf{b}}_{ir}. \quad (\text{a.21})$$

Intuition into the behavior of this estimator can be gained by rewriting (a.21) for the region as

$$\begin{aligned} \hat{\mathbf{a}}_{ir}^{STMM} &= \hat{\beta}_r + \left( \mathbf{1}'_{V_r} \mathbf{1}'_{V_r} \otimes \hat{\mathbf{S}}_r + \hat{\mathbf{Q}}_r \right) \\ &\quad \times \left\{ \mathbf{1}'_{V_r} \mathbf{1}'_{V_r} \otimes \hat{\mathbf{S}} + \hat{\mathbf{Q}}_r + \text{diag}(\mathbf{K}'_i \hat{\Psi}_{i1r} \mathbf{K}_i, \dots, \mathbf{K}'_i \hat{\Psi}_{iNr} \mathbf{K}_i) \right\}^{-1} \left( \mathbf{d}_{ir} - \hat{\beta}_r \right). \end{aligned} \quad (\text{a.22})$$

Holding the within-subject variance fixed and letting  $\mathbf{1}'_{V_r} \mathbf{1}'_{V_r} \otimes \hat{\mathbf{S}}_r + \hat{\mathbf{Q}}_r$  grow, note that this estimator approaches  $\mathbf{d}_{ir}$ , i.e., the OLS estimator.

## Appendix C. An example when the GLS and OLS estimators of population activation are equivalent

Consider the special case of a block design where all subjects have the same design matrix and the data are temporally pre-whitened. Let the spatial covariance,  $\mathbf{Q}$ , be given. The GLS estimator of  $\hat{\beta} \in \mathbb{R}^{VQ}$  is

$$\begin{aligned} \hat{\beta} &= \left\{ \left( \mathbf{1}'_N \otimes \mathbf{I}_Q \otimes \mathbf{X}'_T \right) \left( \mathbf{I}_Q \otimes \mathbf{Q} \otimes \mathbf{I} \right)^{-1} \left( \mathbf{1}_N \otimes \mathbf{I}_Q \otimes \mathbf{X}_T \right) \right\}^{-1} \left( \mathbf{1}'_N \otimes \mathbf{I}_Q \otimes \mathbf{X}'_T \right) \left( \mathbf{I}_Q \otimes \mathbf{Q} \otimes \mathbf{I} \right)^{-1} \mathbf{Y} \\ &= \left\{ \frac{1}{N} \mathbf{1}'_N \otimes \mathbf{I}_Q \otimes \left( \mathbf{X}'_T \mathbf{X}_T \right)^{-1} \mathbf{X}'_T \right\} \mathbf{Y}, \end{aligned} \quad (\text{a.23})$$

which is equivalent to the OLS estimator.

## Appendix D. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2016.05.038>.

## References

- Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., et al., 2013. Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage* 80, 169–189.
- Beckmann, C.F., Jenkinson, M., Smith, S.M., 2003. General multilevel linear modeling for group analysis in fMRI. *NeuroImage* 20 (2), 1052–1063.
- Bernal-Rusiel, J.L., Reuter, M., Greve, D.N., Fischl, B., Sabuncu, M.R., 2013. Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimaging data. *NeuroImage* 81, 358–370.
- Bowman, F.D., 2005. Spatio-temporal modeling of localized brain activity. *Biostatistics* 6 (4), 558–575.

- Bowman, F.D., 2007. Spatiotemporal models for region of interest analyses of functional neuroimaging data. *J. Am. Stat. Assoc.* 102 (478), 442–453.
- Bowman, F.D., Caffo, B., Bassett, S.S., Kilts, C., 2008. A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* 39 (1), 146–156.
- Brett, M., Anton, J.-L., Valabregue, R., Poline, J.-B., 2002. Region of interest analysis using the MarsBar toolbox for SPM 99. *NeuroImage* 16 (2), S497.
- Brown, D.A., Lazar, N.A., Datta, G.S., Jang, W., McDowell, J.E., 2014. Incorporating spatial dependence into Bayesian multiple testing of statistical parametric maps in functional neuroimaging. *NeuroImage* 84, 97–112.
- Castruccio, S., Ombao, H., Genton, M.G., 2016. A Scalable Multi-resolution Spatio-temporal Model for Brain Activation and Connectivity in fMRI Data (arXiv preprint arXiv:1602.02435).
- Cressie, N.A., 1993. *Statistics for Spatial Data*. Wiley New York.
- Cressie, N., Wikle, C.K., 2011. *Statistics for Spatio-temporal Data*. Wiley.
- Derado, G., Bowman, F.D., Kilts, C.D., 2010. Modeling the spatial and temporal dependence in fMRI data. *Biometrics* 66 (3), 949–957.
- Derado, G., Bowman, F.D., Zhang, L., et al., 2013. Predicting brain activity using a Bayesian spatial model. *Stat. Methods Med. Res.* 22 (4), 382–397.
- Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., et al., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8 (4), 272–284.
- Friston, K., Holmes, A., Worsley, K., Poline, J., Frith, C., Frackowiak, R., 1995. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210.
- Friston, K.J., Stephan, K.E., Lund, T.E., Morcom, A., Kiebel, S., 2005. Mixed-effects and fMRI studies. *NeuroImage* 24 (1), 244–252.
- Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al., 2013. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* 80, 105–124.
- Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., Petersen, S.E., 2016. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* 26 (1), 288–303.
- Harrison, L.M., Green, G.G., 2010. A Bayesian spatiotemporal model for very large data sets. *NeuroImage* 50 (3), 1126–1141.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning*. Springer.
- Holodny, A.I., Shevzov-Zebrun, N., Brennan, N., Peck, K.K., 2011. Motor and sensory mapping. *Neurosurg. Clin. N. Am.* 22 (2), 207–218.
- Hyun, J.W., Li, Y., Gilmore, J.H., Lu, Z., Styner, M., Zhu, H., 2014. SGPP: spatial Gaussian predictive process models for neuroimaging data. *NeuroImage* 89, 70–80.
- Kang, H., Ombao, H., Linkletter, C., Long, N., Badre, D., 2012. Spatio-spectral mixed-effects model for functional magnetic resonance imaging data. *J. Am. Stat. Assoc.* 107 (498), 568–577.
- Li, Y., Zhu, H., Shen, D., Lin, W., Gilmore, J.H., Ibrahim, J.G., 2011. Multiscale adaptive regression models for neuroimaging data. *J. R. Stat. Soc. Ser. B (Stat Methodol.)* 73 (4), 559–578.
- Lindquist, M.A., 2008. The statistical analysis of fMRI data. *Stat. Sci.* 23 (4), 439–464.
- Lindquist, M.A., Krishnan, A., Lopez-Sola, M., Jepma, M., Woo, C.-W., Koban, L., Roy, M., Atlas, L.Y., Chang, L.J., Losin, L., et al., 2015. Group-regularized individual prediction: theory and application to pain. *NeuroImage* <http://dx.doi.org/10.1016/j.neuroimage.2015.10.074>.
- McCulloch, C.E., Searle, S.R., Neuhaus, J.M., 2008. *Generalized Linear Mixed Models*. second ed. Wiley.
- Meier, J.D., Afalo, T.N., Kastner, S., Graziano, M.S., 2008. Complex organization of human primary motor cortex: a high-resolution fMRI study. *J. Neurophysiol.* 100 (4), 1800–1812.
- Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.-M., Malave, V.L., Mason, R.A., Just, M.A., 2008. Predicting human brain activity associated with the meanings of nouns. *Science* 320 (5880), 1191–1195.
- Mumford, J.A., Nichols, T., 2009. Simple group fMRI modeling and inference. *NeuroImage* 47 (4), 1469–1475.
- Neggler, S., Langerak, T., Schutter, D., Mandl, R., Ramsey, N., Lemmens, P., Postma, A., 2004. A stereotactic method for image-guided transcranial magnetic stimulation validated with fMRI and motor-evoked potentials. *NeuroImage* 21 (4), 1805–1817.
- Penny, W.D., Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic press.
- Petersson, K.M., Nichols, T.E., Poline, J.-B., Holmes, A.P., 1999. Statistical limitations in functional neuroimaging II. Signal detection and statistical inference. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 354 (1387), 1261–1281.
- Pillai, J., 2010. The evolution of clinical functional imaging during the past 2 decades and its current impact on neurosurgical planning. *Am. J. Neuroradiol.* 31 (2), 219–225.
- Ruffini, G., Fox, M.D., Ripolles, O., Miranda, P.C., Pascual-Leone, A., 2014. Optimization of multifocal transcranial current stimulation for weighted cortical pattern targeting from realistic modeling of electric fields. *NeuroImage* 89, 216–225.
- Searle, S.R., Casella, G., McCulloch, C.E., 1992. *Variance Components*. vol. 391. John Wiley & Sons.
- Shou, H., Eloyan, A., Nebel, M.B., Mejia, A., Pekar, J.J., Mostofsky, S., Caffo, B., Lindquist, M.A., Crainiceanu, C.M., 2014. Shrinkage prediction of seed-voxel brain connectivity using resting state fMRI. *NeuroImage* 102, 938–944.
- Stevens, W.D., Buckner, R.L., Schacter, D.L., 2010. Correlated low-frequency BOLD fluctuations in the resting human brain are modulated by recent experience in category-preferential visual regions. *Cereb. Cortex* 20 (8), 1997–2006.
- Van Essen, D.C., Uebulir, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., et al., 2012. The Human Connectome Project: a data acquisition perspective. *NeuroImage* 62 (4), 2222–2231.
- Wig, G.S., Grafton, S.T., Demos, K.E., Kelley, W.M., 2005. Reductions in neural activity underlie behavioral components of repetition priming. *Nat. Neurosci.* 8 (9), 1228–1233.



- Woolrich, M.W., Behrens, T.E., Beckmann, C.F., Jenkinson, M., Smith, S.M., 2004. Multilevel linear modelling for fMRI group analysis using Bayesian inference. *NeuroImage* 21 (4), 1732–1747.
- Worsley, K.J., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., Evans, A., 2002. A general statistical analysis for fMRI data. *NeuroImage* 15 (1), 1–15.
- Xu, L., Johnson, T.D., Nichols, T.E., Nee, D.E., 2009. Modeling inter-subject variability in fMRI activation location: a Bayesian hierarchical spatial model. *Biometrics* 65 (4), 1041–1051.
- Zhang, L., Guindani, M., Versace, F., Vannucci, M., 2014. A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. *NeuroImage* 95, 162–175.
- Zhang, L., Guindani, M., Vannucci, M., 2015. Bayesian models for functional magnetic resonance imaging data analysis. *Wiley Interdiscip. Rev. Comput. Stat.* 7 (1), 21–41.
- Zhu, H., Fan, J., Kong, L., 2014. Spatially varying coefficient model for neuroimaging data with jump discontinuities. *J. Am. Stat. Assoc.* 109 (507), 1084–1098.