Published in final edited form as:

Biometrics. 2018 June ; 74(2): 645–652. doi:10.1111/biom.12782.

Fully-Bayesian spectral methods for imaging data

Brian J Reich¹, Joseph Guinness¹, Simon N Vandekar², Russell T Shinohara², and Ana-Maria Staicu¹

¹North Carolina State University

²University of Pennsylvania

Summary

Medical imaging data with thousands of spatially-correlated data points are common in many fields. Methods that account for spatial correlation often require cumbersome matrix evaluations which are prohibitive for data of this size, and thus current work has either used low-rank approximations or analyzed data in blocks. We propose a method that accounts for nonstationarity, functional connectivity of distant regions of interest, and local signals, and can be applied to large multi-subject datasets using spectral methods combined with Markov Chain Monte Carlo sampling. We illustrate using simulated data that properly accounting for spatial dependence improves precision of estimates and yields valid statistical inference. We apply the new approach to study associations between cortical thickness and Alzheimer's disease, and find several regions of the cortex where patients with Alzheimer's disease are thinner on average than healthy controls.

Keywords

Functional connectivity; Markov chain Monte Carlo; Matérn correlation; Shrinkage priors; Spherical harmonics

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease associated with severe memory impairment and cognitive deficits (McKhann et al., 1984). The prevalence of AD is increasing, with some projections estimating over 100 million cases across the globe by 2050 (Rocca et al., 2011; Brookmeyer et al., 2007; Wimo et al., 2006). To accelerate the pace of preventative and therapeutic strategies, the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Mueller et al., 2005) was established as a large multi-center observational prospective cohort and consortium for sharing clinical, neuroimaging, and other biomarker data for public access. In this paper, we use ADNI data to identify regions of the cortex most affected by AD.

Modeling spatial dependence in neuroimaging studies such as the ADNI study is key to efficient and valid inference (e.g. Spence et al., 1997). A major obstacle in modeling the

9. Supplementary Materials

Web Figures and R code referenced in Sections 2, 5 and 7 are available with this paper at the *Biometrics* website on Wiley Online Library.

spatial dependence is the computational burden associated with handling the spatial covariance. Hyun et al. (2014) and Zhu et al. (2014) modeled spatial dependence with low-rank models and use Wald-based inference. While this is computationally convenient, low-rank approximations can fail to adequately capture local dependence (Stein, 2014). Recently, many approaches have been proposed to model spatiotemporal data. For example, Kang et al. (2012) introduced a spatio-spectral mixed effects model for fMRI time courses that is able to capture the multiscale spatial covariance structure as well as stationary temporal dependence. To handle the temporal correlation the authors work in the Fourier domain (similar to Lange and Zeger, 1997), as the Fourier coefficients are approximately uncorrelated across frequencies. Recently Castruccio et al. (2016) proposed a spatio-temporal model with spatially varying coefficients and multi-resolution error structure. Model estimation and inference is based on a three-step likelihood approximation approach (Castruccio and Guinness, 2017), which is computationally expensive.

A Bayesian approach is appealing because it naturally accounts for uncertainty in all model parameters (Handcock and Stein, 1993). A drawback to the use of Bayesian methods in large imaging problems is their heavy computational burden, and thus previous work in this area focuses largely on computational approximation. Woolrich et al. (2004) proposed a fully-Bayesian model for spatiotemporal fMRI data. To facilitate computation, they assumed spatial independence given the previous timepoint, and thus it is not clear how to apply this method to our case with multiple subjects but only a single image per subject. Bowman (2007) aggregated data to regions of interest and then performed spatiotemporal modeling on these regional summaries. Musgrove et al. (2017) proposed a Bayesian variable selection model that divides the brain into regions and performs spatial variable selection within each region. Zhang et al. (2014) and Zhang et al. (2016) used variational Bayesian techniques to approximate the posterior distribution. While this methodology is computationally feasible and has been shown to provide good point estimates, its downside is that it yields poor estimates of the posterior variances, which may result in misleading inference.

In this paper, we propose a flexible model for cross-sectional multi-subject brain imaging data. The effect of covariates is allowed to vary spatially and is given a spatial shrinkage prior to encourage sparsity. The residual covariance is a combination of stationary covariance for flexible local dependence and a low-rank nonstationary covariance to capture functional connectivities between distant regions of the brain. The full spatial covariance matrix is cumbersome for large imaging datasets, and we therefore project to the spectral domain (Fuentes and Reich, 2010) to decorrelate the responses and facilitate a fully-Bayesian analysis simultaneously for all subjects. We apply this method to simulated multi-subject data and the ADNI data, where we find evidence of nonstationary spatial correlation and that properly accounting for this dependence improves fit.

2. Description of the ADNI data

This work is motivated by data from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. A primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI) can be used to measure

the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD). We consider measures of cortical thickness estimated using FreeSurfer (Fischl and Dale, 2000a) from the MRI acquired in the ADNI (Jack et al., 2008) in 199 subjects with AD, 409 subjects with mild cognitive impairment which is thought to be an early stage of AD in some cases, and 231 healthy controls aged 54-91. The data were collected as part of ADNI1/GO, and all of the subjects were imaged on 1.5T scanners. Cortical thickness is estimated for each subject as the distance between the pial and white matter surfaces at each location on the cortex. Postprocessed maps of cortical thickness were downloaded from the Laboratory of NeuroImaging repository (<http://adni.loni.usc.edu/>). Briefly, the T1-weighted images were inhomogeneity corrected (Sled et al., 1998) and processed using version 4.3 of FreeSurfer (Fischl and Dale, 2000b) to estimate thickness across the cortex. These measurements were then registered across subjects to the fsaverage5 template. The images are smoothed at 10mm on the surface as part of preprocessing. Cortical thickness is known to decrease with normal aging, but more dramatic and spatially dissociable changes have been widely documented in AD (Lerch et al., 2005, 2008; Dickerson et al., 2009). These changes have been assessed primarily using location-specific regression modeling which is known to be suboptimal. We aim to use the proposed spectral method to study the spatially-varying effects of a diagnosis and the ADNI composite memory score (Crane et al., 2012) on cortical thickness.

Cortical thickness is measured at 19,715 locations on a sphere in template space after registration across subjects. Each point on the sphere corresponds to a location on the cerebral cortex as shown in the Supplementary Materials (Figure 1). Our regressions include an intercept and $p = 5$ subject-level covariates: age, gender, ADNI composite memory (ADNI-Mem) score and binary indicators of AD and MCI status. The continuous covariates age and ADNI-Mem are standardized. Age and gender are included as potential confounders, and the main focus is to evaluate associations between the cortical thickness and ADNI-Mem, AD status and MCI status. Figure 1 plots the z-scores for four covariates (excluding the intercept and gender which has small effects) from an ordinary least square analysis conducted separately for each location. There is clearly a spatial pattern in these estimates. For example the largest positive z-scores for the effect of AD in Figure 1 cluster in the left/anterior/inferior region of the cortex.

3. Description of the model in the spatial domain

Let $Y_i(\mathbf{v})$ be the response at spatial location $\mathbf{v} = (v_1, v_2, v_3)^T$ for subject i . Associated with each subject are an intercept $X_{i0} = 1$ and covariates $X_{i1}, \dots, X_{ip}$. We assume that the n locations of interest form a regular grid and we do not specify a model that can extrapolate beyond these n voxels because typically imaging data are defined on an inherently bounded and discrete domain. The model for subject i is

$$Y_i(\mathbf{v}) = \sum_{k=0}^p X_{ik} B_k(\mathbf{v}) + \sum_{j=1}^J Z_j(\mathbf{v}) \gamma_{ij} + E_i(\mathbf{v}) \quad (1)$$

where the B_k are the spatially-varying fixed-effect processes; Z_j are known basis functions that explain the large-scale spatial structure; $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{iJ})^T$ are the corresponding random effects; and $E_i(\mathbf{v})$ are small-scale spatial deviations. The residual spatial process E_i is a mean-zero Gaussian process with isotropic Matérn covariance function (Stein, 1999)

$$\text{Cov}[E_i(\mathbf{v}), E_i(\mathbf{v}')] = \sigma^2 I(\mathbf{v} = \mathbf{v}') + \tau^2 \mathcal{M}_\nu \left(\frac{\|\mathbf{v} - \mathbf{v}'\|}{\phi} \right) \quad (2)$$

where $\mathcal{M}_\nu$ is the 3D-Matérn correlation $\mathcal{M}_\nu(h) = \frac{2^{1-\nu}}{\Gamma(\nu)} (3h\sqrt{\nu})^\nu \mathcal{K}_\nu(3h\sqrt{\nu})$ and $\mathcal{K}$ is the modified Bessel function of the second kind. The Matérn covariance has four parameters, $\boldsymbol{\theta} = (\sigma^2, \tau^2, \phi, \nu)$: σ^2 is the variance of the non-spatial error (nugget); τ^2 is the variance of the spatial process (partial sill); ϕ is the spatial range; and ν dictates the shape of the spatial correlation function. The parameter ν is referred to as the smoothness parameter because for a continuous process defined at an uncountable number of locations (as opposed to our discrete domain), realizations of the process are ν times differentiable.

The large-scale spatial structure is determined by the J random-effect covariates Z_j . Selection of the Z_j depends on prior knowledge of the process and the desired inference. One option is binary indicators for J regions of interest, such as anatomical regions. Another approach is to select data-driven basis functions using principal components. We use outer products of B-spline or Gaussian kernel basis functions, which have the advantage of being defined locally (e.g., each basis function has a unique maximizing voxel) and spanning the entire space of continuous functions as $J \rightarrow \infty$. The random effects are distributed $\boldsymbol{\gamma}_i \sim \text{Normal}(0, \Sigma)$, where $\Sigma = \{\Sigma_{jl}\}$ is a $J \times J$ covariance matrix. The nonstationary component of the covariance is

$$NS(\mathbf{v}, \mathbf{v}') = \sum_{j=1}^J \sum_{l=1}^J Z_j(\mathbf{v}) Z_l(\mathbf{v}') \Sigma_{jl}. \quad (3)$$

The overall covariance combining both large-scale nonstationary covariance and small-scale stationary covariance is

$$\text{Cov}[Y_i(\mathbf{v}), Y_i(\mathbf{v}')] = NS(\mathbf{v}, \mathbf{v}') + \sigma^2 I(\mathbf{v} = \mathbf{v}') + \tau^2 \mathcal{M}_\nu \left(\frac{\|\mathbf{v} - \mathbf{v}'\|}{\phi} \right). \quad (4)$$

The spatially-varying intercept B_0 and slopes B_k for $k = 1, \dots, p$ are modeled as independent (over k) Gaussian processes with mean

$$E[B_k(\mathbf{v})] = \sum_{j=1}^J Z_j(\mathbf{v})\beta_{kj} \quad (5)$$

and Matérn covariance with covariance parameters $\theta_k = (\sigma_k^2, \tau_k^2, \phi_k, \nu_k)$. As with the spatial covariance in (4), the mean functions have both large-scale trends given by the basis functions Z_j and small-scale variation given by a Matérn spatial process. Of course, it is possible to use different basis functions for the mean and covariance, but we select the same basis for notational simplicity.

As we expect the intercept to be smooth over space we give normal priors

$\beta_{0j} \stackrel{\text{indep}}{\sim} \text{Normal}(0, \delta_0^2)$. In contrast, for the slope parameters we assume that for most of the spatial domain $B_k(\mathbf{v}) = 0$. To encode this prior, we use a continuous shrinkage prior for the β_{kj} for $k > 0$ that concentrates prior mass near zero to give sparsity, but also has heavy tails to allow for the signal to emerge. The shrinkage prior we select is the horseshoe prior (Carvalho et al., 2010), which is $\beta_{kj} \delta_{kj} \sim \text{Normal}(0, |\delta_{kj}|^2)$ and $\delta_{kj} \sim \text{Cauchy}(0, \delta_k)$.

4. Description of the model in the spectral domain

The covariance in (3) is composed of a low-rank nonstationary component and a full-rank stationary component; the full-rank stationary Matérn process poses the computational challenge for large data sets. To decorrelate the Matérn process and permit an analysis of the entire spatial domain we transform to the spectral domain. Because the data are defined on a sphere, we use the spherical harmonics transformation (SHT) for the B_k and E_i . The $E_i(B_k)$ are expanded similarly) are represented as

$$E_i(\mathbf{v}) = \sum_{\ell=0}^L \sum_{m=-\ell}^{\ell} \frac{1}{\sqrt{2\ell+1}} S_{\ell}^m(s_1, s_2) \mathcal{E}_i(\boldsymbol{\omega}) \quad (6)$$

where $S_{\ell}^m(s_1, s_2)$ are the spherical harmonic functions, $\boldsymbol{\omega} = (\ell m)$, and $\mathcal{E}_i(\boldsymbol{\omega})$ are the unknown coefficients. The spherical harmonic functions are $S_{\ell}^m(s_1, s_2) = P_{\ell}^m[\cos(s_1)] \exp(im s_2)$ where P_{ℓ}^m is the associated Legendre polynomial. The full representation of the process requires $L = \infty$, but we approximate the process with a finite L . Transformations (e.g., from $Y_i(\mathbf{v})$ to $\mathcal{Y}_i(\boldsymbol{\omega})$ or $E_i(s)$ to $\mathcal{E}_i(\boldsymbol{\omega})$ in (6)) are conducted using least squares, and the unique real components $\tilde{Y}_i(\boldsymbol{\omega})$ and $\tilde{E}_i(\boldsymbol{\omega})$ are extracted from the complex $\mathcal{Y}_i(\boldsymbol{\omega})$ and $\mathcal{E}_i(\boldsymbol{\omega})$. That is, the SHT returns complex conjugate pairs of the form $a_i(\boldsymbol{\omega}) + ib_i(\boldsymbol{\omega})$ and $a_i(\boldsymbol{\omega}) - ib_i(\boldsymbol{\omega})$, and we extract the unique real values $a_i(\boldsymbol{\omega})$ and $b_i(\boldsymbol{\omega})$.

Denote $\tilde{Y}_i(\boldsymbol{\omega})$, $\tilde{B}_k(\boldsymbol{\omega})$, $\tilde{Z}_j(\boldsymbol{\omega})$, $\tilde{E}_i(\boldsymbol{\omega})$ as the real-valued (i.e., the unique real values extracted from the complex data in the spectral domain) spectral representations of the processes

$Y_f(\nu)$, $B_k(\nu)$, $Z_f(\nu)$ and $E_f(\nu)$, respectively, for frequency ω . Since the SHT is a linear operator, the spatial model in (1) becomes

$$\tilde{Y}_i(\omega) = \sum_{k=1}^p X_{ik} \tilde{B}_k(\omega) + \sum_{j=1}^J \tilde{Z}_j(\omega) \gamma_{ij} + \tilde{E}_i(\omega). \quad (7)$$

The Gaussian processes E_i and B_k are stationary and defined over a discrete spatial domain and thus their spectral counterparts $\tilde{E}_i$ and $\tilde{B}_k$ can be expressed in terms of independent normal random variables (Yaglom, 2012) (including terms from the same conjugate pair)

$$\begin{aligned} \tilde{E}_i(\omega) &\stackrel{\text{indep}}{\sim} \text{Normal}[0, \tilde{\lambda}(\omega | \theta)] \\ \tilde{B}_k(\omega) &\stackrel{\text{indep}}{\sim} \text{Normal}\left[\sum_{j=1}^J \tilde{Z}_j(\omega) \beta_{kj}, \tilde{\lambda}(\omega | \theta_k)\right] \end{aligned} \quad (8)$$

where the spectral density function $\tilde{\lambda}$ is determined by the Matérn covariance parameters. Following Guinness and Fuentes (2016), we take the variance of $\tilde{E}_i(\omega)$ to be $\lambda_k(\theta) = \sigma^2 + \tau^2 (\phi^{-2} + \ell^2)^{-\nu-1/2}$. This mimics the flexibility of the Matérn covariance function based on Euclidean distance in that ϕ controls the range and ν controls the smoothness of the process.

The basis functions $Z_f(\nu)$ are also projected into the spectral domain. The spectral processes $\tilde{Z}_j(\omega)$ will be nonzero for all j and ω . However, since the B-spline basis functions are smooth functions, they can be approximated accurately with a small number of low-frequency terms with $\omega \in \mathcal{L}$. Therefore, to improve computational efficiency, we set $\tilde{Z}_j(\omega) = 0$ for terms with $\omega \notin \mathcal{L}$, and select $\mathcal{L}$ to include roughly J terms.

5. Priors and computing details

The model laid out in Sections 3 and 4 can be written succinctly as follows (with all terms being defined conditionally on lower levels of the hierarchy). For terms with $\omega \in \mathcal{L}$,

$$\begin{aligned}
\tilde{Y}_i(\omega) &\stackrel{\text{indep}}{\sim} \text{Normal} \left[\sum_{k=0}^p X_{ik} \tilde{B}_k(\omega) + \sum_{j=1}^J \tilde{Z}_j(\omega) \gamma_{ij}, \tilde{\lambda}(\omega | \theta) \right] \\
\gamma_i &\stackrel{\text{indep}}{\sim} \text{Normal}(0, \Sigma) \\
\tilde{B}_l(\omega) &\stackrel{\text{indep}}{\sim} \text{Normal} \left[\sum_{j=1}^J \tilde{Z}_j(\omega) \beta_{lj}, \tilde{\lambda}(\omega | \theta_l) \right] \quad \text{for } l = 0, 1, \dots, p \\
\beta_{0j} &\stackrel{\text{indep}}{\sim} \text{Normal}(0, \delta_0^2) \\
\beta_{kj} &\stackrel{\text{indep}}{\sim} \text{Normal}(0, \delta_{kj}^2) \quad \text{and} \quad \delta_{kj} \sim \text{HalfCauchy}(\delta_k) \quad \text{for } k > 0.
\end{aligned} \tag{9}$$

For terms with $\omega \in \mathcal{L}$,

$$\begin{aligned}
\tilde{Y}_i(\omega) &\stackrel{\text{indep}}{\sim} \text{Normal} \left[\sum_{k=1}^p X_{ik} \tilde{B}_k(\omega), \tilde{\lambda}(\omega | \theta) \right] \\
\tilde{B}_l(\omega) &\stackrel{\text{indep}}{\sim} \text{Normal} [0, \tilde{\lambda}(\omega | \theta_l)] \quad \text{for } l = 1, 2.
\end{aligned} \tag{10}$$

After the spectral transformation, $\tilde{Y}_i(\omega)$ and $\tilde{B}_l(\omega)$ are independent across ω . The Matérn parameters are reparameterized from $\theta = (\sigma^2, \tau^2, \phi, \nu)$ (and similarly for the θ_k) to overall variance $\nu = \sigma^2 + \tau^2$, the logit of the spatial variance proportion $r = \text{logit}[\tau^2/(\sigma^2 + \tau^2)]$, and the log range and smoothness, $\phi' = \log(\phi)$ and $\nu' = \log(\nu)$. To complete the Bayesian model we specify the following uninformative priors: $\Sigma \sim \text{InvWishart}(J + \varepsilon, \frac{\nu_3}{J + \varepsilon} I_J)$, $\nu, \nu_0, \nu_1, \delta_0, \delta_1 \sim \text{InvGamma}(\varepsilon, \varepsilon)$, and $\phi', \phi'_0, \phi'_1 \sim \text{Normal}(0, 1/\varepsilon^2)$ with $\varepsilon = 0.1$; $r, r_0, r_1 \sim \text{Normal}(0, 1)$ so the signal-to-noise ratio has prior mass spread over $[0, 1]$; and $\nu', \nu'_0, \nu'_1 \sim \text{Normal}(-2, 1)$ to avoid too much mass on large smoothness parameters.

We explore the posterior with a standard combination of Gibbs and Metropolis sampling. $\tilde{B}_k(\omega)$, γ_i , Σ , and ν_k are updated from their conjugate full conditional distributions. The remaining Matérn parameters are updated using random-walk Metropolis sampling with Gaussian candidate distributions tuned to give acceptance probability around 0.4. R code is available in the online supplementary materials.

6. Simulation study

In this section we conduct a simulation study to explore the effects of model misspecification on Bayesian inference. Each simulated dataset includes m subjects and $p = 1$ binary covariate. The first $m/2$ subjects have $X_i = 0$ and the remaining have $X_i = 1$. Data are generated on the unit sphere (i.e., $\nu_1^2 + \nu_2^2 + \nu_3^2 = 1$) at a regular 30×30 grid of $n = 900$

points defined by longitude ($s_1 \in [0, 2\pi]$) and latitude ($s_2 \in [0, \pi]$). The intercept $B_0(\mathbf{v})$ and the residuals E_j are generated from a Gaussian process with mean zero and Matérn covariance defined by Euclidean distance in $\mathcal{R}^3$ and $\boldsymbol{\theta}_0 = (0.01, 0.99, 0.1, 1)$ and $\boldsymbol{\theta} = (0.1, 0.9, 0.1, 1)$, respectively. Data are generated with $J = 2$ basis functions: $Z_1(\mathbf{v}) = \mathbf{I}(s_1 < \pi)$ and $Z_2(\mathbf{v}) = 1 - Z_1(\mathbf{v})$. Data generation scenarios vary by the number of subjects $m = \{50, 100\}$; the magnitude of nonstationarity with $\Sigma = 0$ (“RE = N”) or $\Sigma = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ (“RE = Y”); and the form of the signal with either $B_1(\mathbf{v}) = \cos(2s_1) \sin(5s_2)$ (“HotSpot = N”) so the signal is non-sparse or $B_1(\mathbf{v}) = 2\exp\{-3[(s_1 - \pi/2)^2 + (s_2 - \pi/2)^2]\}$ (“HotSpot = Y”) so that the signal large only near $(s_1, s_2) = (\pi/2, \pi/2)$.

We fit several versions of the model given in Sections 3 and 4. For all models we approximate the process with $L = 29$ levels giving $n = 900$ terms. The $J = 25$ basis functions $Z_j(s)$ are taken to be Gaussian kernel basis functions with J knots forming a 5×5 rectangular grid over the observed latitudes and longitudes and $\mathcal{L} = \{\omega | l < 5\}$. The kernel bandwidth is set to the minimum distance between knots, and distances between knots and data locations are computed using spherical distance. For each dataset we fit the model with ($\tau > 0$, “Matérn”) and without ($\tau = 0$, “Independent”) spatially-dependent residuals; with ($\Sigma \neq 0$, “Nonstationary”) and without ($\Sigma = 0$, “Stationary”) nonstationarity; and with ($\delta_1 > 0$, “SP”) and without ($\delta_1 = 0$, “NSP”) sparsity in the prior for B_1 . For each method and dataset we generate 10,000 MCMC iterations and discard the first 2,000 as burn-in. The full model takes around 8 seconds per 1,000 MCMC iterations for $m = 50$ subjects and 11 seconds per 1,000 MCMC iterations for $m = 100$ subjects using R on an ordinary PC.

For each dataset we compute the posterior mean $\bar{B}_1(\mathbf{v})$ and 95% equal-tailed credible set $[L_1(\mathbf{v}), U_1(\mathbf{v})]$ for $B_1(\mathbf{v})$ for all voxels and then compute the spatial average mean squared error $\Sigma_v [B_1(\mathbf{v}) - \bar{B}_1(\mathbf{v})]^2/n$ and coverage $\Sigma_v \mathbf{I}([L_1(\mathbf{v}), U_1(\mathbf{v})] \ni B_1(\mathbf{v})) / n$. Table 1 reports the average of the mean squared error and coverage across the 100 simulated datasets for each scenario. The stationary model with independent residuals often has high MSE and low coverage. For the models with independent residuals, adding the low-rank non-stationary covariance improves performance, but coverage remains low in some cases. Coverage is at or near the nominal 0.95 level for the full model for all scenarios. The inclusion of non-stationarity and horseshoe priors improve performance in the anticipated settings.

7. Analysis of the ADNI data

7.1 Model comparisons

In this section we apply the model proposed in Section 3 to the ADNI data. Unlike Section 3, the covariate effects do not have mean zero because for this application we do not expect all fixed effects to be centered on zero (e.g., the confounder age should have an increasing effect throughout the brain). Therefore we modify (5) to be $E[B_k(\mathbf{v})] = \bar{B}_k + \sum_{j=1}^J Z_j(\mathbf{v})\beta_{kj}$, where the $\bar{B}_k$ is the overall mean and the β_{kj} are have the same prior in as Section 3.

For our analysis we use $L = 30$ giving a reduction from 18,715 observations for each subject in the spatial domain to 931 observations in the spectral domain. Exploratory analysis in the Supplementary Materials (Figure 1) shows that terms beyond $L = 30$ do not appear to

contain additional spatial signal. As in Section 3, we allow for a different set of spatial covariance parameters for the residual process E_i and the fixed effects B_k . Preliminary analysis shows that the nugget variance is estimated to be near zero and so it is excluded. Also, while we allow the six fixed effects terms to have a different variance σ_k^2 , we assume they have the same spatial range (ϕ) and smoothness parameter (ν) to improve MCMC convergence. The overall effect $\tilde{B}_k$ have normal priors with mean zero and variance 1,000, and for the remaining hyperparameters we use the same uninformative priors as in Section 6. The MCMC algorithm is run for 100,000 iterations, the first 10,000 are discarded as burn-in, and the remaining samples are thinned by 10 to remove autocorrelation. The full model with Matérn correlation, $J = 225$ basis functions, Horseshoe priors and nonstationary covariance fit to these data takes around 11 minutes per 1,000 iterations.

We compare several methods using test set prediction. We randomly (across subject and resolution) remove 10% of the $\tilde{Y}_i$, fit the model to the remaining 90%, and evaluate the agreement of the posterior predictive distributions and the test set observations. The models fit vary by the specification of the residual correlation, fixed and random effects as in the simulation study, and also the number of basis function J . Methods are compared in Table 2 using mean squared error of the posterior predictive means, the average (across all observations) of the posterior predictive variances, the empirical coverage of the 90% credible sets, and the posterior predictive mean of the log-likelihood of the test set observations. Comparisons are made by Legendre resolution in Supplementary Materials (Figure 2). Inclusion of the Matérn covariance is the most important factor in determining the predictive log-likelihood, and including the nonstationary covariance with $J = 225$ also improves fit. For these data, the Gaussian prior for fixed effects is sufficient, and we proceed with Gaussian priors for the fixed effects and covariance that includes both the stationary Matérn component and non-stationary component with $J = 225$ basis functions.

7.2 Results of the final model

Figure 2 summarizes the nonstationary component $NS(\nu, \nu')$ of the covariance in (4). The standard deviation of the random effects ranges from 0.1–0.4; the standard deviation of the response (across subject and location) is 0.53, so the nonstationary random effects account for a substantial proportion of the variation. Figure 2 also maps the correlation in the nonstationary component between three locations (indicated with black dots) and all other locations, and show some interesting patterns. In Figure 2's second row, all sites in the superior/posterior region are strongly correlated. The bottom two rows reveal symmetries between the left and right sides of the cortex in the posterior and anterior regions, respectively, that clearly could not be captured by the stationary Matérn correlation function.

Posterior z-scores plotted in Figure 3. For the z-score maps, the threshold produced using the method of Sun et al. (2015) to control Bayesian false discovery rate at 0.01 (separate for all covariates) are given in the caption and statistically significant regions base on this threshold are displayed in the Supplementary Materials (Figure 5). In this multiple testing problem, the one-sided null and alternative hypotheses are $H_0 : \beta_k(s) = 0$ and $H_1 : \beta_k(s) < 0$, respectively, for each covariate except ADNI-Mem which uses $H_0 : \beta_k(s) = 0$ and $H_1 : \beta_k(s) > 0$. These plots broadly resemble the least squares results in Figure 1, but are smoother and

in some cases reveal stronger results. Age and ADNI-Mem are the strongest predictors. Age is negatively correlated with the response throughout the cortex and has the strongest estimates in the anterior, specifically in the bilateral parahippocampal gyrus, temporal pole, dorsolateral prefrontal cortex, and superior frontal cortex. ADNI-Mem is positively correlated throughout the brain with the largest estimates in the anterior/inferior region, including bilaterally in precuneus, middle and inferior temporal lobes, and medial orbitofrontal gyrus.

Much of the spatial variation in the voxel-at-a-time MCI regression coefficients (Figure 1, bottom row) is smoothed out by the spatial model, and as a result none of the posterior z-scores are significant in the spatial model. Also, the largest least squares effects for AD are smoothed towards zero, suggesting these positive effects AD may be spurious associations. However, AD is associated with significantly thinner cortex bilaterally in the inferior temporal lobes (including the parahippocampal gyrus), the occipitoparietal junction and dorsal prefrontal cortex. These regions overlap with well-established findings of parahippocampal atrophy in the ADNI by Greene et al. (2010), and are understood to be involved in the earlier stages of pathology (Braak and Braak, 1991). Atrophy in occipitoparietal junction is another well-documented change, and pronounced pathology in this area has been associated with a posterior cortical variant of AD (e.g., Crutch et al., 2012) which manifests in more visuospatial and visuo-perceptive deficiencies. Predominant atrophy of the frontal cortex accompanies the frontal variant of AD which manifests in behavioral changes (Dubois et al., 2014). Many of these regions overlap with those found to be associated with ADNI-Mem scores (Nho et al., 2012).

8. Discussion

In this paper, we proposed a spectral spatial model for large imaging data sets. The model accommodates features often seen in brain imaging data including nonstationary spatial covariance and local covariate effects. The spherical harmonics transformation permit a fully-Bayesian application to large datasets while respecting the spherical nature of the data. Using simulated data we show that properly accounting for residual spatial correlation is necessary for efficient estimation and valid inference. When applied to the ADNI data we find that both stationary Matérn and low-rank nonstationary covariance terms improve fit, and that spatial methods identify regions of the cortex that differ by Alzheimer's disease status.

We have focused on multi-subject studies, however our methods could be applied for a single-subject study where data is collected over time. For example, after a discrete Fourier transform over time at each voxel (Kang et al., 2012) the images for different frequencies can be considered as independent (analogous to independence across subjects) and the proposed method can be applied. Another area of future work is to improve computation by exploiting parallel computing. The likelihood factors across subjects and frequencies, and therefore likelihood evaluations are embarrassingly parallelizable and many parameters could be updated in parallel which might improve computation time by an order of magnitude.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was partially supported NIH awards RO1NS085211 (AS and RTS), R21NS093349 (RTS), and T32MH065218 (SNV) and NSF award MS-1613219 (JG). Data collection and sharing was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (DOD W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California (USC). ADNI data are disseminated by the Laboratory for Neuro Imaging at USC. The authors thank http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/%ADNI_Acknowledgement_List.pdf

References

- Bowman FD. Spatiotemporal models for region of interest analyses of functional neuroimaging data. *Journal of the American Statistical Association*. 2007; 102:442–453.
- Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*. 1991; 82:239–259. [PubMed: 1759558]
- Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. *Alzheimer's & Dementia*. 2007; 3:186–191.
- Carvalho CM, Polson NG, Scott JG. The horseshoe estimator for sparse signals. *Biometrika*. 2010; 97:465–480.
- Castruccio S, Guinness J. An evolutionary spectrum approach to incorporate large-scale geographical descriptors on global processes. *Journal of the Royal Statistical Society: Series C*. 2017 in press.
- Castruccio, S., Ombao, H., Genton, M. A multi-resolution spatio-temporal model for brain activation and connectivity in fMRI data. 2016. <https://arxiv.org/abs/1602.02435>
- Crane PK, Carle A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging and Behavior*. 2012; 6:502–516. [PubMed: 22782295]
- Crutch SJ, Lehmann M, Schott JM, Rabinovici GD, Rossor MN, Fox NC. Posterior cortical atrophy. *The Lancet Neurology*. 2012; 11:170–178. [PubMed: 22265212]
- Dickerson BC, Feczko E, Augustinack JC, Pacheco J, Morris JC, Fischl B, Buckner RL. Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiology of Aging*. 2009; 30:432–440. [PubMed: 17869384]
- Dubois B, Feldman HH, Jacova C, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *The Lancet Neurology*. 2014; 13:614–629. [PubMed: 24849862]
- Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*. 2000a; 97:11050–11055.
- Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*. 2000b; 26:11050–11055.
- Fuentes, M., Reich, BJ. Spectral domain. In: Gelfand, AE, Diggle, PJ, Fuentes, M., Guttorp, P., editors. *Handbook of Spatial Statistics*. Vol. chapter 5. Chapman & Hall/CRC Press; 2010.
- Greene SJ, Killiany RJ. Alzheimer's Disease Neuroimaging Initiative (ADNI). Subregions of the inferior parietal lobule are affected in the progression to Alzheimer's disease. *Neurobiology of Aging*. 2010; 31:1304–1311. [PubMed: 20570398]
- Guinness J, Fuentes M. Isotropic covariance functions on spheres: Some properties and modeling considerations. *Journal of Multivariate Analysis*. 2016; 143:143–152.
- Handcock MS, Stein ML. A Bayesian analysis of Kriging. *Technometrics*. 1993; 35:403–410.

- Hyun JW, Li YM, Gilmore J, et al. SGPP: Spatial Gaussian predictive process models for neuroimaging data. *NeuroImage*. 2014; 89:70–80. [PubMed: 24269800]
- Jack CR, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*. 2008; 27:685–691. [PubMed: 18302232]
- Kang H, Ombao H, Linkletter C, Long N, Badre D. Spatio-spectral mixed-effects model for functional magnetic resonance imaging data. *Journal of the American Statistical Association*. 2012; 107:568–577. [PubMed: 25400305]
- Lange N, Zeger SL. Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 1997; 46:1–29.
- Lerch JP, Pruessner J, Zijdenbos AP, Collins DL, Teipel SJ, Hampel H, Evans AC. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiology of Aging*. 2008; 29:23–30. [PubMed: 17097767]
- Lerch JP, Pruessner JC, Zijdenbos A, Hampel H, Teipel SJ, Evans AC. Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. *Cerebral Cortex*. 2005; 15:995–1001. [PubMed: 15537673]
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984; 34:939–939. [PubMed: 6610841]
- Mueller SG, Weiner MW, Thal LJ, et al. Ways toward an early diagnosis in Alzheimers disease: the Alzheimers Disease Neuroimaging Initiative (ADNI). *Alzheimer's & Dementia*. 2005; 1:55–66.
- Musgrove DR, Hughes H, Eberly LE. Fast, fully Bayesian spatiotemporal inference for fMRI data. *Biostatistics*. 2017 In press.
- Nho K, Risacher SL, Crane PK, et al. Voxel and surface-based topography of memory and executive deficits in mild cognitive impairment and Alzheimer's disease. *Brain Imaging and Behavior*. 2012; 6:551–567. [PubMed: 23070747]
- Rocca WA, Petersen RC, Knopman DS, Hebert LE, et al. Trends in the incidence and prevalence of Alzheimers disease, dementia, and cognitive impairment in the United States. *Alzheimer's & Dementia*. 2011; 7:80–93.
- Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*. 1998; 17:87–97. [PubMed: 9617910]
- Spence J, Carmack P, Gunst R, Schucany W, Woodward W, Haley R. Accounting for spatial dependence in the analysis of SPECT brain imaging data. *Journal of the American Statistical Association*. 1997; 102:464–473.
- Stein, M. Statistical interpolation of spatial data: Some theory for Kriging. Springer; New York: 1999.
- Stein ML. Limitations on low rank approximations for covariance matrices of spatial data. *Spatial Statistics*. 2014; 8:1–19.
- Sun W, Reich BJ, Cai T, Guindani M, Schwartzman A. False discovery control in large-scale spatial multiple testing. *Journal of the Royal Statistical Society: Series B*. 2015; 77:59–77.
- Wimo A, Jonsson L, Winblad B. An estimate of the worldwide prevalence and direct costs of dementia in 2003. *Dementia and Geriatric Cognitive Disorders*. 2006; 21:175–181. [PubMed: 16401889]
- Woolrich M, Jenkinson M, Brady J, Smith S. Fully Bayesian spatio-temporal modeling of fMRI data. *IEEE Transactions on Medical Imaging*. 2004; 23:213–231. [PubMed: 14964566]
- Yaglom, AM. Correlation theory of stationary and related random functions. Springer Science & Business Media; 2012.
- Zhang L, Guindani M, Versace F, et al. A spatiotemporal non-parametric Bayesian model of multi-subject fMRI data. *The Annals of Applied Statistics*. 2016; 10:638–666.
- Zhang L, Guindani M, Versace F, Vannucci M. A spatio-temporal non-parametric Bayesian variable selection model of fMRI data for clustering correlated time courses. *NeuroImage*. 2014; 95:162–175. [PubMed: 24650600]

Zhu H, Fan J, Kong L. Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association*. 2014; 109:1084–1098. [PubMed: 25435598]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Z-scores for the covariate effects from ordinary least squares conducted separately for each location (the intercept and gender effect are omitted). The outlined regions in the center of the images in the middle two columns is the medial wall where the two hemispheres meet and does not include measures of cortical thickness. This figure appears in color in the electronic version of this article.

Figure 2.

Summaries of the posterior mean of the nonstationary covariance $NS(\mathbf{v}; \mathbf{v}')$ in (3). The top row gives the standard deviation $\sqrt{NS(\mathbf{v}, \mathbf{v})}$ and the remaining rows plot the correlation $NS(\mathbf{v}, \mathbf{v}') / \sqrt{NS(\mathbf{v}, \mathbf{v})NS(\mathbf{v}', \mathbf{v}')}$ across $\mathbf{v}$ for three points $\mathbf{v}'$ denoted with black dots. The outlined regions in the center of the images in the middle two columns is the medial wall where the two hemispheres meet and does not include measures of cortical thickness. This figure appears in color in the electronic version of this article.

Figure 3.

Posterior z-scores (i.e., the ratio of the posterior mean and standard deviation) for the fixed effects $\bar{B}_k + B_k(s)$ for the ADNI data (the intercept and gender effect are omitted). We declare results statistically significant while controlling Bayesian false discovery rate at 0.01 if $z < -1.16$ for age, $z > 0.88$ for ADNI-Mem, $z < -1.70$ for AD; there are no significant results for MCI using this criteria. The outlined regions in the center of the images in the middle two columns is the medial wall where the two hemispheres meet and does not include measures of cortical thickness. This figure appears in color in the electronic version of this article.

Table 1

Simulation study mean squared error and pointwise coverage probability for $B_1(\mathbf{v})$ for **data on a sphere**. Data sets are generated with m subjects, with or without random effects (“RE”) to induce nonstationarity, and with the true signal following the fixed hot-spot pattern (“Hot spot= Y ”) or from a Gaussian process (“Hot spot= N ”). The model is fit with independent or Matérn small-scale residuals; with or without random effects to allow for nonstationarity, and with (“SP”) or without (“NSP”) Bayesian shrinkage priors. All values are multiplied by 100 and the maximum of each row’s standard error is given in the final column.

(a) Mean squared error												
Data generation				Fitted model								
m	RE	Hot spot	Independent			Stationary			Matérn			Max SE
			NSP	SP	NSP	NSP	SP	NSP	NSP	SP	NSP	
50	N	N	4.9	4.9	4.7	4.7	4.7	5.5	5.6	5.2	5.4	0.1
		Y	3.3	3.1	3.8	3.2	2.9	2.6	3.1	2.6	2.6	0.1
	Y	N	12.8	13.1	5.2	6.3	5.2	6.0	5.9	6.9	6.9	0.8
		Y	9.4	9.1	6.0	5.2	7.1	5.1	5.4	4.1	4.1	0.8
100	N	N	2.9	3.0	2.8	2.8	3.1	3.2	3.0	3.1	<0.1	
		Y	2.2	2.2	2.3	2.1	2.0	1.8	2.0	1.9	<0.1	
	Y	N	7.7	7.8	3.3	3.6	3.3	3.6	3.5	3.8	0.4	
		Y	5.9	5.7	4.4	3.8	5.3	3.6	4.0	3.2	0.4	
(b) Coverage of 95% intervals												
Data generation				Fitted model								
m	RE	Hot spot	Independent			Stationary			Matérn			Max SE
			NSP	SP	NSP	NSP	SP	NSP	NSP	SP	NSP	
50	N	N	84	84	90	90	91	90	91	91	91	<1
		Y	86	85	92	93	96	96	96	96	96	<1
	Y	N	71	70	90	90	95	95	91	92	92	2
		Y	59	47	88	89	90	95	92	95	95	3
100	N	N	87	87	91	91	91	91	91	92	92	<1
		Y	89	88	93	93	95	95	95	95	96	<1
	Y	N	73	73	91	92	95	96	92	93	93	1

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

(b) Coverage of 95% intervals											
Data generation						Fitted model					
			Independent			Matérn					
			Stationary	Nonstationary	Stationary	Stationary	Nonstationary	Stationary	Nonstationary	Stationary	Max
			NSP	SP	NSP	SP	NSP	SP	NSP	SP	SE
<i>m</i>	RE	Hot spot	59	49	86	87	89	94	89	93	3
		Y									

Table 2

Cross validation results for the ADNI data. The methods fit to the data vary by the residual correlation of $\tilde{E}$; the prior for the fixed effects (“FE”) $\tilde{B}$, the inclusion of the nonstationary subject random effects (“nonstationary”) $\tilde{Z}$, and the number of basis functions J . Methods are compared by the average (over all test set observations) of log prediction likelihood (“Log Like”), mean squared prediction error (“MSE”), average prediction variance (“Var”), and coverage of 90% prediction intervals (“Cov”).

Residuals	FE	Nonstationary	J	Log Like	MSE	Var	Cov90
Independent	Gaussian	No	25	-0.945	0.372	0.591	0.959
Independent	Gaussian	No	225	-0.945	0.372	0.591	0.959
Independent	Gaussian	Yes	25	-0.907	0.435	0.334	0.919
Independent	Gaussian	Yes	225	-1.087	0.327	0.295	0.895
Independent	Horseshoe	No	25	-0.945	0.372	0.591	0.959
Independent	Horseshoe	No	225	-0.945	0.372	0.590	0.960
Independent	Horseshoe	Yes	25	-0.907	0.435	0.334	0.919
Independent	Horseshoe	Yes	225	-1.088	0.329	0.295	0.895
Matern	Gaussian	No	25	-0.520	0.371	0.530	0.901
Matern	Gaussian	No	225	-0.520	0.371	0.530	0.901
Matern	Gaussian	Yes	25	-0.513	0.385	0.342	0.901
Matern	Gaussian	Yes	225	-0.529	0.313	0.302	0.900
Matern	Horseshoe	No	25	-0.520	0.372	0.530	0.901
Matern	Horseshoe	No	225	-0.520	0.372	0.530	0.902
Matern	Horseshoe	Yes	25	-0.512	0.385	0.342	0.901
Matern	Horseshoe	Yes	225	-0.529	0.313	0.301	0.900