False Discovery Rate Control in Magnetic Resonance Imaging Studies via Markov Random Fields

Hien D. Nguyen*, Geoffrey J. McLachlan, Nicolas Cherbuin, and Andrew L. Janke

Abstract—Magnetic resonance imaging (MRI) is widely used to study population effects of factors on brain morphometry. Inference from such studies often require the simultaneous testing of millions of statistical hypotheses. Such scale of inference is known to lead to large numbers of false positive results. Control of the false discovery rate (FDR) is commonly employed to mitigate against such outcomes. However, current methodologies in FDR control only account for the marginal significance of hypotheses, and are not able to explicitly account for spatial relationships, such as those between MRI voxels. In this article, we present novel methods that incorporate spatial dependencies into the process of controlling FDR through the use of Markov random fields. Our method is able to automatically estimate the relationships between spatially dependent hypotheses by means of maximum pseudo-likelihood estimation and the pseudo-likelihood information criterion. We show that our methods have desirable statistical properties with regards to FDR control and are able to outperform noncontextual methods in simulations of dependent hypothesis scenarios. Our method is applied to investigate the effects of aging on brain morphometry using data from the PATH study. Evidence of whole brain and component level effects that correspond to similar findings in the literature is found in our investigation.

Index Terms—Magnetic resonance imaging, Neuroimaging, Mixture model, False discovery rate, Markov random field

I. INTRODUCTION

Magnetic resonance imaging (MRI) is often used by neurological researchers to assess the population effects of factors on brain morphometry. Examples of such studies include investigations regarding the effects of aging (e.g. [1]–[6]), alcoholism (e.g. [5], [6]), education (e.g. [1]), and gender (e.g. [3]) on MRI-measurable morphological features.

In such studies, a test of hypothesis may be carried at each individual voxel of the MRI scans by means of regression or similar statistical procedures. Aggregating across an entire

brain, the number of such simultaneous hypothesis tests can number in the millions.

Such large scales of simultaneous hypothesis testing is known to result in large numbers of false positive results (i.e. incorrectly rejected hypothesis tests). For example, if there are no factor effects, one million hypothesis tests at the usual 5% significance would yield fifty thousand incorrectly rejected hypotheses on average. In the neuroimaging community, this pitfall was famously alluded to in [7], where brain activity was found to be significant in a functional MRI study of a dead salmon.

The problem of false positives in simultaneous hypothesis testing has been well studied in both the statistics literature (e.g. [8]–[10]) and the neuroimaging literature (e.g. [11]–[13]). Among the solutions developed, controlling the false discovery rate (FDR) has emerged to be particularly popular and practical.

The principle of FDR control was first introduced in [14], where it was suggested that researchers should reject only as many hypotheses as to maintain the FDR below a predetermined level, where $\text{FDR} = \mathbb{E}(N_{01}/N_R | N_R > 0) \mathbb{P}(N_R > 0)$ with N_{01} and N_R denoting the number of false positives and the number of rejected hypotheses, respectively. Since the publication of [14], many techniques have been proposed for the control of FDR (e.g. [15]–[21]).

Although these techniques are able to correctly control the FDR when hypotheses have implicit dependencies, they are not able to model the explicit nature of such dependencies. As such, the spatial relationships between hypothesis tests on neighboring voxels cannot be utilized to inform the significance of said voxels.

In the neuroimaging literature, the exploitation of spatial relationships for false positive mitigation is often conducted under the random field theory (RFT) framework (cf. [22]–[25]). Some examples of RFT techniques for FDR control include [26] and [27].

While widely used, RFT techniques are not without their caveats. These include issues such as test dependency (e.g. χ^2 , Gaussian, F , T , or T^2 distributed test statistics require different field models), the assumption of a continuous field over a discrete lattice, and requirements of user-specified smoothing parameters (cf. [13] and [28] for discussions of these issues). As such, RFT techniques can produce incorrect inferences, or be misapplied.

In order to resolve the shortcomings of RFT techniques, we introduce a novel method that uses a test-independent

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The study was supported by the Dementia Collaborative Research Centres and the NHMRC of Australia grant No. 1002160. Nicolas Cherbuin is funded by ARC Fellowship No. 12010227.

*Hien D. Nguyen is with the School of Mathematics and Physics, The University of Queensland, St. Lucia, Brisbane, Australia 4075 (email: h.nguyen7@uq.edu.au).

Geoffrey J. McLachlan is with the School of Mathematics and Physics, The University of Queensland, St. Lucia, Brisbane, Australia 4075.

Nicholas Cherbuin is with the Neuroimaging and Brain Lab, Australian National University, Acton, Canberra, Australia 0200.

Andrew L. Janke is with the Centre for Advanced Imaging, The University of Queensland, St. Lucia, Brisbane, Australia 4075.

discrete-lattice model to spatially inform FDR control. Our method utilizes the mixture model-based procedure of [19] to control FDR at the marginal (spatially uninformed and noncontextual) level. Spatial dependencies between marginally controlled hypothesis tests are then modeled by a Markov random field (MRF); a technique previously used for model-based segmentation of MRIs (e.g. [29]–[32]). As such, we name our collection of techniques: Markov random field FDR control (MRF-FDR).

Usual problems in the application of MRFs are the determination of spatial relationship magnitudes and the distance at which voxels can be considered neighbors. We resolve these issues by using a data-driven process through applications of the pseudo-likelihood (PL) (cf. [33]) and the PL information criterion (PLIC) (cf. [34]), respectively.

The MRF-FDR methods that we present here are a continuation of the work from [35]. In this article, we suggest and theoretically justify a refined version of the technique in [35].

In order to demonstrate and assess the performance of MRF-FDR, we perform a set of simulation studies with the aim of replicating brain MRI-like scenarios, where features of various sizes are statistically significant. We find that our method is able to outperform some commonly available techniques. We then apply our technique to study the effect of aging on the morphometry of the brain in an elderly cohort study using data from the PATH project (cf. [36]). Here, we are able to produce inferences consistent with the neuroimaging literature on aging.

II. NONCONTEXTUAL FDR CONTROL

Suppose that we test a statistical hypothesis at each of the n voxels of an MRI, where the voxels are indexed by $i = 1, \dots, n$. Let H_i be a binary random variable, where $H_i = 0$ if the null hypothesis at voxel i is true, and 1 otherwise, and let P_i be the p -value associated with the hypothesis test at voxel i . If P_i arises from a well-behaved hypothesis test (i.e. $P_i|H_i = 0 \stackrel{i.d.}{\sim} U(0, 1)$), then the probit transformation of P_i yields a z -score $Z_i = \Phi^{-1}(1 - P_i)$, which has the property $Z_i|H_i = 0 \stackrel{i.d.}{\sim} N(0, 1)$, under the well-behaved test assumption. Here, $i.d.$ denotes identically distributed.

In [18], it was suggested that if $0 < \pi_0 < 1$, then Z_i is marginally distributed according to a two-component mixture distribution with density

$$f(z_i) = \pi_0 f_0(z_i) + (1 - \pi_0) f_1(z_i), \quad (1)$$

when H_i is unobserved. Here $\pi_0 = \mathbb{P}(H_i = 0)$, f_0 is the density of $Z_i|H_i = 0$, f_1 is the density of $Z_i|H_i = 1$, and h_i and z_i are realizations of H_i and Z_i , respectively. Since $Z_i|H_i = 0 \stackrel{i.d.}{\sim} N(0, 1)$ under the well-behaved test assumption, we can write $f_0(z_i) = \phi(z_i; 0, 1)$, where $\phi(x; \mu, \sigma^2)$ is a normal density function with mean μ and variance σ^2 . The determination of f_1 is flexible and often depends on situation specifics. However, it was shown in [19] that using $f_1(z_i) = \phi(z_i; \mu_1, \sigma_1^2)$ performs well in general settings, where $\mu_1 \in \mathbb{R}$ and $\sigma_1^2 > 0$. It can be seen that under such conditions, density (1) is a two-component normal mixture model (cf. Chapter 3 of [37] for details).

A. Empirical Null Density Function

The density $f_0(z_i) = \phi(z_i; 0, 1)$ arising from the well-behaved test assumption is referred to as the theoretical null density by [18], and has been broadly tested for veracity. It was found in [18], [38], [39] that in some testing procedures (e.g. permutation tests and correlated test statistics), the assumption that $Z_i|H_i = 0 \stackrel{i.d.}{\sim} N(0, 1)$ is unreasonable.

As an alternative, it was suggested in [18] that the marginal density of $Z_i|H_i = 0$ should be modeled by $f_0(z_i) = \phi(z_i; \mu_0, \sigma_0^2)$ instead, where $\mu_0 \in \mathbb{R}$, $\sigma_0^2 > 0$, and $\mu_0 < \mu_1$. This model for $Z_i|H_i = 0$ is referred to as the empirical null density, and has been shown to improve the fit of the two-component normal mixture model in [19]. As such, we adopt the two-component normal mixture with an empirical null density,

$$f(z_i; \theta) = \pi_0 \phi(z_i; \mu_0, \sigma_0^2) + (1 - \pi_0) \phi(z_i; \mu_1, \sigma_1^2), \quad (2)$$

for noncontextual FDR (NC-FDR) control in the following sections, and shall refer to (2) as the NC-FDR model. Here $\theta = (\pi_0, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2)^T$ is the vector of model parameters, where the T superscript indicates matrix transposition.

B. Rejection Rule

Under the NC-FDR model, Bayes' rule allows us to write the conditional probability of the event $H_i = 0|Z_i = z_i$ as

$$\begin{aligned} \mathbb{P}(H_i = 0|Z_i = z_i) &= \frac{\pi_0 \phi(z_i; \mu_0, \sigma_0^2)}{f(z_i; \theta)} \\ &= \tau(z_i; \theta), \end{aligned} \quad (3)$$

which can be used to derive the rejection rule

$$r_1(z_i; \theta, c_1) = \begin{cases} 1 & \text{if } \tau(z_i; \theta) \leq c_1, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where $0 \leq c_1 \leq 1$. Here, $r_1(z_i; \theta, c_1) = 1$ if the null hypothesis at voxel i is rejected (declared statistically significant), and 0 otherwise.

Expression (3) is referred to as the local FDR by [18], and it can be interpreted as the probability that the null hypothesis at voxel i is true, given knowledge of its z -score. Rejection rule (4) requires that the probability of voxel i being null is sufficiently low (defined through the threshold c_1) before rejecting the hypothesis at this voxel. For brevity, we will suppress the parameter vector θ when convenient, and write $\tau(z_i; \theta)$ and $r_1(z_i; \theta, c_1)$, as $\tau(z_i)$ and $r_1(z_i; c_1)$, respectively.

Let $I\{A\}$ be the indicator variable, where $I\{A\} = 1$ if proposition A is true, and 0 otherwise. Using the indicator notation, we can express the number of hypotheses rejected by rule (4) as $N_R = \sum_{i=1}^n I\{r_1(Z_i; c_1) = 1\}$, and the number of false positives as $N_{01} = \sum_{i=1}^n I\{r_1(Z_i; c_1) = 1, H_i = 0\}$. Furthermore, we can show that the marginal FDR ($m\text{FDR} = \mathbb{E}N_{01}/\mathbb{E}N_R$) can be consistently estimated (in the probabilistic sense) by

$$\overline{m\text{FDR}} = \frac{\bar{N}_{01}}{\bar{N}_R}, \quad (5)$$

where we define $\bar{N}_R = \sum_{i=1}^n I\{r_1(z_i; c_1) = 1\}$ and $\bar{N}_{01} = \sum_{i=1}^n I\{r_1(z_i; c_1) = 1\} \tau(z_i)$. Here, $\overline{mFDR} = \overline{mFDR}(c_1)$ is a function of c_1 , although we will drop the notation for brevity.

Theorem 1. *For each i , assume that there are at most $\nu < \infty$ indices $i' \neq i$ ($i' = 1, \dots, n$), such that Z_i and H_i are dependent on $Z_{i'}$ and $H_{i'}$. If the probabilities $\mathbb{P}(r_1(Z_i; c_1) = 1, H_i = 0)$ and $\mathbb{P}(r_1(Z_i; c_1) = 1) > 0$ are constant for all i , then $\overline{mFDR} \xrightarrow{P} mFDR$.*

The assumption from Theorem 1 states that although the voxels are not assumed to be independent, we only permit a finite and fixed amount of dependencies. Under the same assumption, we can show that $mFDR$ approaches FDR as n gets large.

Theorem 2. *Under the same assumption as Theorem 1, if the probabilities $\mathbb{P}(r_1(Z_i; c_1) = 1, H_i = 0)$, $\mathbb{P}(r_1(Z_i; c_1) = 1, H_i = 1)$, and $\mathbb{P}(r_1(Z_i; c_1) = 1) > 0$ are constant for all i , then $\lim_{n \rightarrow \infty} |mFDR - FDR| = 0$.*

Proofs of all theoretical results can be found in Appendix A.

C. FDR Control

Using Theorems 1 and 2, we can apply equation (5) to approximate the FDR for any given threshold c_1 . Alternatively, we can also use $\overline{mFDR}$ to approximately control the FDR at a specified level $0 < \alpha \leq 1$ by choosing c_1 using the threshold rule

$$c_1 = \arg \max_{c'_1 \in (0, 1]} \{\overline{mFDR}(c'_1) \leq \alpha\}. \quad (6)$$

For brevity, we will refer to the applications of (4) and (6) for estimation and control of FDR as NC-FDR collectively.

Although rule (6) asymptotically controls FDR at the correct level for hypotheses behaving as conjectured by the NC-FDR model, it cannot exploit the spatial structure of MRI data. However, the NC-FDR model does provide a useful framework for incorporating spatial information, as discussed in the following section.

III. SPATIAL FDR CONTROL

Consider for any voxel i , $\tau(z_i) < 1/2$ indicates that the null hypothesis at the voxel is more likely to be false than not, given z_i is observed. Using this notion we can define the binary random variable $T_i = I\{\tau(Z_i) < 1/2\}$, which equals 0 if the null hypothesis at voxel i is more likely to be true, and 1 otherwise.

Let $\mathbf{s}_i = (s_{1i}, s_{2i}, s_{3i})^T$ be the spatial coordinates and $S_i^d = \{i' \neq i : \delta(\mathbf{s}_i, \mathbf{s}_{i'}) \leq d\}$ be the d -range Moore-type neighborhood of voxel i , where $\delta(\mathbf{s}_i, \mathbf{s}_{i'}) = \max\{|s_{1i} - s_{1i'}|, |s_{2i} - s_{2i'}|, |s_{3i} - s_{3i'}|\}$ is the maximum distance function. Using an MRF argument, a spatial relationship between T_i and its neighbors $T_{(i)} = \{T_{i'} : i' \in S_i\}$ can be induced through the conditional probability statement

$$\begin{aligned} \mathbb{P}(T_i = t_i | T_{(i)} = t_{(i)}) &= \frac{\exp(at_i + bt_i \eta(t_{(i)}))}{1 + \exp(a + b\eta(t_{(i)}))} \\ &= g(t_i, t_{(i)}; \psi), \end{aligned} \quad (7)$$

where $a, b \in \mathbb{R}$ and $\psi = (a, b)^T$ is the parameter vector. Here $\eta(t_{(i)}) = |S_i^d|^{-1} \sum_{i' \in S_i^d} t_{i'}$ is the mean over the neighborhood S_i^d of voxel i , where $|S|$ denotes the cardinality of set S .

Model (7) allows for the probability of T_i to be dependent upon the probability that its neighboring hypotheses are null, and thus, provides a way of incorporating spatial dependencies for FDR control.

A. Rejection Rule

Using (7), we can write the probability of $T_i = 0 | T_{(i)} = t_{(i)}$ as

$$\begin{aligned} \mathbb{P}(T_i = 0 | T_{(i)} = t_{(i)}) &= \frac{1}{1 + \exp(a + b\eta(t_{(i)}))} \\ &= \xi(t_{(i)}; \psi). \end{aligned} \quad (8)$$

Combining expressions (3) and (8), we can specify the spatial informed rejection rule

$$r_2(z_{(i)}; \theta, \psi, c_1, c_2) = \begin{cases} 1 & \text{if } \tau(z_i; \theta) \leq c_1 \\ & \text{and } \xi(t_{(i)}; \psi) \leq c_2, \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

where $0 < c_1, c_2 \leq 1$, and $z_{(i)} = (z_i, t_{(i)})^T$. As with rule (4), $r_2(z_{(i)}; c_1, c_2) = 1$ if the null hypothesis at voxel i is rejected, and 0 otherwise.

Rule (9) extends rule (4) to also take into account the spatial relationship between nearby voxels by rejecting the null hypothesis at voxel i only if the local FDR is below the specified threshold c_1 , and the probability that the null hypothesis at the voxel is false, given the state of its neighbors is also sufficiently low (defined through the threshold c_2). The second condition $\xi(t_{(i)}) \leq c_2$, guarantees that the rejection of the hypothesis at voxel i is spatially coherent with the decisions made at the $(2d+1)^3 - 1$ voxels in its neighborhood S_i^d . Note that if $c_2 = 1$, then rule (9) is exactly rule (4).

B. Theoretical Justification

Under rule (9), we can write the number of rejections and false positives as $N_R = \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1\}$ and $N_{01} = \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1, H_i = 0\}$, respectively. As with equation (5), we can show that

$$\overline{mFDR}' = \frac{\bar{N}_{01}}{\bar{N}_R} \quad (10)$$

can conservatively estimate (i.e. over estimate) the $mFDR$ when we assume that nearby hypotheses are spatially coherent, where $\bar{N}_R = \sum_{i=1}^n I\{r_2(z_{(i)}; c_1, c_2) = 1\}$ and $\bar{N}_{01} = \sum_{i=1}^n I\{r_2(z_{(i)}; c_1, c_2) = 1\} \tau(z_i)$, respectively.

Theorem 3. *For each i , assume that there are at most $\nu < \infty$ indices $i' \neq i$ ($i' = 1, \dots, n$), such that $Z_{(i)}$ and H_i are dependent on $Z_{(i')}$ and $H_{i'}$. Additionally, assume that $\gamma_{1i} \leq \gamma_{2i}$ for all i , where*

$$\gamma_{1i} = \mathbb{E}(I\{H_i = 0\} | r_2(Z_{(i)}; c_1, c_2) = 1),$$

and

$$\begin{aligned}\gamma_{2i} &= \mathbb{E}(I\{H_i = 0\} | r_1(Z; c_1) = 1) \\ &= \tau(Z_i).\end{aligned}$$

If

$$\begin{aligned}\mathbb{P}(r_2(Z_{(i)}; c_1, c_2) = 1) &> 0, \\ \mathbb{P}(r_2(Z_{(i)}; c_1, c_2) = 1, H_i = 0) &, \\ \text{and } \mathbb{E}(I\{r_2(Z_{(i)}; c_1, c_2) = 1\} \tau(Z_i)) &\text{ are constant for all } i, \\ \text{then } \overline{mFDR}' \xrightarrow{P} mFDR' &\geq mFDR.\end{aligned}$$

As with rule (4), we can show that the $mFDR$ for rule (9) also approaches the FDR as n gets large.

Theorem 4. *Under the same assumptions as Theorem 3, if the probabilities*

$$\begin{aligned}\mathbb{P}(r_2(Z_{(i)}; c_1, c_2) = 1, H_i = 0), \\ \mathbb{P}(r_2(Z_{(i)}; c_1, c_2) = 1, H_i = 1),\end{aligned}$$

and $\mathbb{P}(r_2(Z_{(i)}; c_1, c_2) = 1) > 0$ are constant for all i , then $\lim_{n \rightarrow \infty} |mFDR - FDR| = 0$.

Note that the assumption that $\gamma_{1i} \leq \gamma_{2i}$ in Theorem 3 is a formal statement of the conjecture of local coherency among voxels. That is, the assumption states that the null hypothesis at voxel i is less likely to be true if it is known to be in a region of false hypotheses, compared to when this is unknown. Here, we again assume limited dependencies between the voxels; furthermore, notice that ν is only assumed to be finite and may be greater than the size of S_i^d .

C. FDR Control

Combined, Theorems 3 and 4 can be applied to conservatively estimate the FDR by (10), for any given thresholds c_1 and c_2 . Additionally, like (5), equation (10) can be used to conservatively control the FDR by choosing c_1 and c_2 such that

$$(c_1, c_2) = \arg \max_{(c'_1, c'_2) \in (0,1]^2} \{\overline{mFDR}'(c'_1, c'_2) \leq \alpha\}. \quad (11)$$

As with NC-FDR, the usage of (9), (10), and (11) for control or estimation of FDR will henceforth be referred to as MRF-FDR, for brevity. We now proceed to discuss the estimation of the parameter vectors θ and ψ for application in NC-FDR and MRF-FDR.

IV. ESTIMATION OF PARAMETERS AND QUANTITIES

A. Noncontextual Model

Let $z_1, \dots, z_n$ be a set of observed z-scores as defined in Section II, assuming that Z_i is marginally distributed as per (2). The parameter vector (i.e. θ) can be estimated by maximizing the composite marginal likelihood (CML) function

$$CL(\theta; z_1, \dots, z_n) = \prod_{i=1}^n f(z_i; \theta). \quad (12)$$

Because of the mixture form of (2), the parameter vector θ that maximizes (12) cannot be expressed in closed form.

However, it is possible to apply the minorization-maximization (MM) algorithm (cf. [40]) to compute the maximum CML estimate (MCMLE) $\hat{\theta}$ through an iterative scheme. Particulars of the algorithm used are given in Appendix B. Here, the MCMLE is preferred over the maximum likelihood estimate because it can be computed without explicit declaration of the dependencies between the z-scores; see [41] for a treatment on CMLs and [42] for a general overview of composite likelihood methods.

B. Spatial Model

Upon computing the MCMLE, the conditional probabilities $\tau(z_i; \theta)$ can be estimated by $\tau(z_i; \hat{\theta})$, and thus t_i can be approximated by $\hat{t}_i = I\{\tau(z_i; \hat{\theta}) < 1/2\}$. Under the hypothesis that the random variables $\hat{T}_i$ are related to their neighbors $\hat{T}_{(i)}$ through the conditional model (i.e. (7)), we can estimate the parameter vector ψ by maximizing the PL

$$PL(\psi; \hat{t}_1, \dots, \hat{t}_n) = \prod_{i=1}^n g(\hat{t}_i, \hat{t}_{(i)}; \psi). \quad (13)$$

The maximum PL estimation (MPLE) technique was introduced in [33], and the maximum PL estimate (MPLE) can be shown to have desirable statistical properties in a broad range of circumstances (cf. [43] and [44]). As with the estimation of θ , the MPLE $\hat{\psi}$ cannot be estimated in closed form due to the intractability of equation (13). However, we can implement a Newton-type algorithm to iteratively compute $\hat{\psi}$; see Appendix B for details.

C. FDR Estimates

Using the estimates $\hat{\theta}$ and $\hat{\psi}$, we can estimate the marginal FDR expressions (5) and (10) by

$$\widehat{mFDR} = \frac{\sum_{i=1}^n I\{r_1(z_i; \hat{\theta}, c_1) = 1\} \tau(z_i; \hat{\theta})}{\sum_{i=1}^n I\{r_1(z_i; \hat{\theta}, c_1) = 1\}} \quad (14)$$

and

$$\widehat{mFDR}' = \frac{\sum_{i=1}^n I\{r_2(\hat{z}_{(i)}; \hat{\theta}, \hat{\psi}, c_1, c_2) = 1\} \tau(z_i; \hat{\theta})}{\sum_{i=1}^n I\{r_2(\hat{z}_{(i)}; \hat{\theta}, \hat{\psi}, c_1, c_2) = 1\}}, \quad (15)$$

respectively, where $\hat{z}_{(i)} = (z_i, \hat{t}_{(i)})^T$. We can show that (14) and (15) asymptotically approach (5) and (10), respectively, when $\hat{\theta}$ and $\hat{\psi}$ are consistent estimates of the relevant parameters.

Theorem 5. *If $\sum_{i=1}^n I\{r_1(Z_i; \theta, c_1) = 1\} > 0$ and $\hat{\theta} \xrightarrow{P} \theta$, then $\widehat{mFDR} \xrightarrow{D} mFDR$.*

Proof: The proof is similar to that of Theorem 6. ■

Theorem 6. *If $\sum_{i=1}^n I\{r_2(Z_{(i)}; \theta, \psi, c_1, c_2) = 1\} > 0$, $\hat{\theta} \xrightarrow{P} \theta$, and $\hat{\psi} \xrightarrow{P} \psi$, then $\widehat{mFDR}' \xrightarrow{D} mFDR'$.*

The assumptions that $\hat{\theta}$ and $\hat{\psi}$ be consistent are well justified for our models. In the case of the density function

(2), it is well known that if $0 < \pi_0 < 1$ and $\mu_0 < \mu_1$, then the model is identifiable. Under the assumption of Theorem 1, it can be shown that (12) satisfies the regularity conditions for the consistency of an extremum estimator (cf. Theorem 4.1.2 of [45]) with the aid of an appropriate uniform law of large numbers (e.g. Theorem 4 of [46]). Hence, there exist a consistent MCMLE.

To obtain the consistency of $\hat{\psi}$, we need to first establish the identifiability (in the sense of [43]) of the MRF equation (7).

Lemma 7. *If $\psi^* \neq \psi$, then there exists t_i and $t_{(i)}$ such that $g(t_i, t_{(i)}; \psi^*) \neq g(t_i, t_{(i)}; \psi)$, where $\psi^* = (a^*, b^*)^T$.*

The consistency of the MPLE $\hat{\psi}$ can be obtained as follows.

Proposition 8. *If $T_1, \dots, T_n$ are random observations from the MRF (7), and if (7) is identifiable, then $\hat{\psi} \xrightarrow{P} \psi$.*

Proof: See the consistency of pseudo-likelihood theorem of [43]. ■

The results above allow for the use of equations (14) and (15) in the place of (5) and (10) for FDR estimation and control, respectively.

D. Range Estimation

Thus far, the range (i.e. d) of the neighborhoods S_i^d has been assumed constant. Due to the intractability of its inclusion in the MPLE process, it is necessary to determine d by means of an external method. We can use the PLIC for such a purpose.

The PLIC at any d is defined as

$$\text{PLIC}(d) = \log PL(\hat{\psi}_d; \hat{t}_1, \dots, \hat{t}_n) - 2^{-1} k_d \log n,$$

where $\hat{\psi}_d$ is the MPLE, and k_d is the number of parameters for a d -ranged MRF (here, $k_d = 2$ for all d). According to [34], the optimal range can be estimated consistently via the rule

$$d = \arg \max_{d'=1,2,\dots} \text{PLIC}(d'). \quad (16)$$

Rule (16) has been applied successfully in imaging applications (e.g. [47]), and can be viewed as the PL equivalent of the Bayesian Information Criterion of [48]. Because an exhaustive search over all $d = 1, 2, \dots$ is not feasible, we can instead apply a greedy version of (16) in practice. We do this by initializing at $d' = 1$ and iteratively increasing d' until $\text{PLIC}(d' + 1) - \text{PLIC}(d') \leq 0$, whereupon we set $d = d'$. The use of this method is implicit in the ensuing discussions.

V. SIMULATION STUDY

In order to discuss the application and assess the performance of MRF-FDR, we explore a number of simulation scenarios. These scenarios are coded by the size of their spatial features (S1-S3, where S1 has the smallest spatial features and S3 has the largest), and the size of their statistical effect (E1-E3, where E1 has the smallest statistical effect, and E3 has the largest). For example, scenario S1E1 has small spatial features and statistical effects.

Figure 1. Reference schematics for simulation study.

A. Setup

In each of the 9 scenarios, a cube of $n = 120^3$ voxels was simulated with intensities generated independently from the $N(0, 1)$ distribution. The total size of the simulation was chosen to replicate the number of voxels in a typical MRI study. For S3, the cube is partitioned into 8 sub-cubes (as per schema B of Fig. 1), each of size 60^3 . The dark sub-cubes are then replaced with intensities generated independently from $N(\Delta, 1)$. In scenario S2, the cube is partitioned into 27 sub-cubes (as per schema A), each of size 40^3 , whereby the dark sub-cubes are replaced with independent intensities from $N(\Delta, 1)$.

Scenario S1 is a modification of S3. In S1, a type B schema is used, whereby each of the black sub-cubes are replaced with a type A sub-cube with sub-sub-cubes of size 20^3 . Here, the white sub-sub-cubes are unchanged, and the black sub-sub-cubes are replaced by independent intensities from $N(\Delta, 1)$. If we call the voxels with intensities from $N(0, 1)$ null, and those with intensities from $N(\Delta, 1)$ alternative, then there are 56×20^3 , 14×40^3 , and 4×60^3 alternative voxels in S1-S3, respectively.

The statistical effect can be thought of as the difference between a null and an alternative voxel. Here, the statistical effect is determined by the Δ in the alternative intensity generation. We use $\Delta = 1, 2, 3$, for scenarios E1-E3, respectively.

Let $\zeta_1, \dots, \zeta_n$ be the intensities of the n voxels, as per the simulation described. Using ζ_i , we can write the p -value for the test of the null hypothesis $\Delta = 0$, against the alternative hypothesis $\Delta > 0$, as $p_i = 1 - \Phi(\zeta_i)$, for each voxel i . This is simply a one-side z-test for a normal mean.

B. Threshold Selection

Integral to the application of MRF-FDR is the selection of the marginal and spatial thresholds, c_1 and c_2 , respectively. In practice, the choice of combinations of the two thresholds

can be difficult due to the number of options. However, a decision can be made based on two primary statistics. Firstly, (15) can be used to conservatively estimate the FDR for the combination of thresholds, and secondly,

$$\hat{N}_R = \sum_{i=1}^n I \left\{ r_2 \left(\hat{z}_{(i)}; \hat{\theta}, \hat{\psi}, c_1, c_2 \right) = 1 \right\} \quad (17)$$

can be used to evaluate the number of voxels rejected under the choice of threshold. The tradeoff in these two quantities, with respect to the thresholds can be assessed by graphical means.

To demonstrate the process of threshold selection, we simulated a single instance of S1E2, S2E2, and S3E2. The MRF-FDR method was then applied in each case using threshold levels, $c_1, c_2 = 0.01, 0.02, \dots, 1$. Level plots of $\widehat{mFDR}'$ and $\hat{N}_R/n \times 100\%$ (the percentage of hypotheses rejected) were then produced, and are given in Fig. 2. We can make some general observations from the figure. Firstly, for a fixed value of c_2 , higher values c_1 correspond to increases in both $\widehat{mFDR}'$ and $\hat{N}_R$. Secondly, the effect is the same for higher levels of c_2 when c_1 is fixed. We also observe that in all scenarios, $\widehat{mFDR}'$ and $\hat{N}_R$ appear to behave strangely when $c_2 < 0.1$. This is potentially due to a discordance between γ_{1i} and γ_{2i} (from Theorem 3) at very small threshold levels. Lastly, we note that the joint effects of c_1 and c_2 appear to be linear in the sense that there are no paradoxical decreases in FDR or rejections when both thresholds are increased in tandem.

These results appear to confirm the common sense notion that rejecting hypotheses under more stringent conditions would decrease the number of false positives, but also diminish the ability to reject potentially significant hypotheses. If the desired level of FDR and number of rejections are known, then the level plots from Fig. 2 can be used to decide upon a combination of thresholds for the study. For example, if one chooses to control FDR at the 0.1 level, and make $\hat{N}_R/n \times 100\% = 40\%$ in scenario S2E2, then a possible choice of thresholds is $c_1 = 0.3$ and $c_2 = 0.5$. However, there are many other choices that would yield approximately the same combination of outcomes.

C. False Discovery Rate Control

In practice, one does not determine the number of rejections to make. As such, the only decision to make is to control the FDR at an appropriate level α . Here, as in the previous section, the choice is not obvious due to the multitude of combinations of thresholds which lead to similar FDR levels. To alleviate the issue, we propose the use of three thresholding rules: (6), (11), and

$$c_1 = \arg \max_{c'_1 \in (0,1]} \left\{ \widehat{mFDR}'(c'_1, 0.5) \leq \alpha \right\}. \quad (18)$$

The first rule is the NC-FDR marginal rule which ignores the spatial information available for FDR control. The second is the MRF-FDR method discussed in Section III, which allows for dependencies in the rejections of neighboring voxels, and thirdly, rule (18) is a stringent criterion, whereby it is insisted that voxels are only rejected if their neighborhood indicates a lower chance of the null being true than false.

D. Assessing Performance

In order to assess the performance of suggested rules, we simulated 10 repetitions each of scenarios S1E1 to S3E3 controlling α at the scientifically-typical levels 0.005, 0.01, 0.05 and 0.1. For each repetition, we compute the observed FDR ($oFDR = oN_{01}/oN_R$), and observed true positive rate ($oTPR = oN_{11}/oN_1$), where oN_{01} , oN_{11} , oN_R and oN_1 are the numbers of observed false positives, true positives (correctly rejected hypotheses), rejections and number of false null hypotheses, respectively. We then average over the ten repetitions of each scenario and α level.

The $oFDR$ and $oTPR$ measurements were selected for their interpretability. Intuitively, $oFDR$ measures the number of false positives per rejected hypothesis and should be near α if the technique assessed is correctly controlling the FDR level. The $oTPR$ instead measures the power of the procedure to reject hypotheses where the null is false and should be as large as possible. The most desirable property for a control method is to achieve $oFDR$ close to α , and a high $oTPR$ simultaneously, over the scenarios.

For additional comparison, we also assess the performances of the [14] and [15] methods which we shall refer to as BH and BY, respectively.

E. Results

The results of the simulation described above are reported in Table I. From the results we can make a number of observations. We can firstly rank the rules based on the number of scenarios where they achieved the best results. In terms of $oFDR$, the order from best to worst (counting all ties, and not counting E1 scenarios, where $\alpha = 0.005$ or 0.01) is (18), BY, (11), and BH tied with (6). For $oTPR$, the order from best to worst is (11), (6), (18), BH, and BY.

We firstly note that in E1 scenarios, where $\alpha = 0.005$ or 0.01, all methods were either incapable of rejecting any hypotheses, or were only able to reject negligible numbers of alternative voxels. This is due to the conservative nature of all the tested methods, when α is very low. When ignoring these conditions, we find that the mixture model family of rules (i.e. (6), (11), and (18)) tended to outperform BH and BY in terms of $oTPR$, and that the two spatially informed rules (i.e. (11) and (18)) tended to outperform or equal the marginal rules in many scenarios, while also achieving lower $oFDR$ values.

Additionally, we note some specific observations. It can be seen that only BH and BY were meaningful in S1E2 for $\alpha = 0.005$, indicating that these methods may be more appropriate than the mixture model family when controlling for very low levels of FDR when there are smaller spatial features. This observation is supported by observing the performance of BH in terms of $oTPR$ in S1E3. However, we also observe that BH and BY tend to conservatively control FDR over all other scenarios, and thus suffer in terms of $oTPR$ due to the lack of rejections made.

Next, we observe that rules (6) and (11) tend to behave similarly in terms of both $oFDR$ and $oTPR$ for low α values and low levels of effects, although increasing either of these two quantities results in (11) outperforming (6) in both criteria.

Figure 2. Sub-figures A1-A3 and B1-B3 display the relationships between c_1 and c_2 , versus $\widehat{mFDR}'$ and $\hat{N}_R/n \times 100\%$ (percentage of hypotheses rejected), respectively, for simulations of scenarios S1E2-S3E2. The white space indicates values of which $\widehat{mFDR}'$ could not be evaluated (i.e. when there are no rejections).

Table I

AVERAGE α FDR AND α TPR VALUES (OVER 10 REPETITIONS) FOR THE TESTED RULES UNDER VARIOUS SIMULATION SCENARIOS. BOLD VALUES INDICATE THE BEST PERFORMANCE WITHIN EACH SCENARIO (LOWEST FDR AND HIGHEST TPR). DASHED LINES INDICATE NO REJECTIONS MADE BY THE METHOD IN THE SPECIFIED SCENARIO.

Scenario		S1E1		S1E2		S1E3		S2E1		S2E2		S2E3		S3E1		S3E2		S3E3	
α	Rule	<i>o</i> FDR	<i>o</i> TPR																
0.005	(6)	-	-	-	-	.003	.352	-	-	.005	.104	.004	.590	-	-	.005	.097	.004	.580
	(11)	-	-	-	-	.003	.352	-	-	.005	.104	.004	.590	-	-	.004	.097	.004	.580
	(18)	-	-	-	-	.000	.337	-	-	.000	.104	.000	.590	-	-	.000	.097	.000	.580
	BH	-	-	.004	.021	.004	.386	-	-	.002	.049	.002	.494	-	-	.003	.046	.002	.488
	BY	-	-	.002	.000	.000	.096	-	-	.000	.001	.000	.147	-	-	.000	.001	.000	.144
0.01	(6)	-	-	.006	.034	.010	.536	.000	.000	.009	.182	.009	.707	.010	.000	.009	.172	.009	.698
	(11)	-	-	.002	.022	.010	.536	.000	.000	.009	.182	.009	.707	.000	.000	.009	.172	.009	.698
	(18)	-	-	.000	.027	.000	.513	.000	.000	.000	.181	.000	.706	.000	.000	.000	.172	.000	.698
	BH	-	-	.008	.048	.007	.495	.000	.000	.005	.101	.005	.606	.000	.000	.005	.097	.005	.600
	BY	-	-	.001	.001	.001	.147	-	-	.000	.003	.000	.214	-	-	.000	.003	.000	.210
0.05	(6)	.049	.001	.049	.284	.050	.793	.044	.013	.049	.551	.049	.913	.042	.011	.049	.531	.049	.907
	(11)	.013	.001	.049	.284	.020	.803	.031	.011	.049	.551	.003	.932	.017	.006	.049	.537	.002	.926
	(18)	-	-	.000	.227	.000	.772	.001	.013	.001	.549	.000	.932	.000	.011	.000	.535	.000	.926
	BH	.039	.000	.037	.230	.037	.751	.024	.002	.024	.370	.024	.839	.024	.002	.025	.361	.025	.835
	BY	-	-	.003	.012	.002	.329	-	-	.001	.030	.002	.431	-	-	.002	.029	.002	.425
0.1	(6)	.094	.008	.100	.460	.100	.880	.095	.101	.098	.738	.099	.963	.096	.090	.098	.718	.100	.960
	(11)	.016	.004	.087	.462	.020	.908	.095	.101	.020	.764	.021	.993	.096	.090	.051	.738	.024	.989
	(18)	-	-	.000	.376	.000	.878	.002	.099	.002	.763	.003	.999	.001	.088	.000	.746	.001	.999
	BH	.072	.002	.074	.382	.074	.845	.049	.018	.048	.546	.048	.912	.048	.017	.050	.537	.050	.909
	BY	-	-	.005	.030	.005	.431	-	-	.003	.067	.003	.541	-	-	.003	.064	.003	.535

Unlike rule (6) which controls FDR at the desired α level, rules (11) and (18) tend to be conservative like BH and BY, supporting the assumption that $\gamma_{1i} \leq \gamma_{2i}$, in Theorem 3. This does not appear to impact upon the ϕ TPR results, as exemplified by the numerous scenarios where (18) achieved the best results in both criteria. Upon inspection, we observe that the aforementioned scenarios where (18) performs well are cases where the effect sizes or spatial features are large (E3 or S3); both of which are rare in practice. Furthermore, the ϕ FDR is always much smaller than the allowed level α , indicating that rule (18) may be overly conservative.

Finally, we note that in the scenarios where (18) achieves the best ϕ TPR, (11) tends to produce a comparable result. The closeness in ϕ TPR in combination with the ability of rule (11) to better use the FDR allowance indicates that (11) may be a better scenario across all scenarios and α levels.

Thus, we conclude that the mixture family of rules tends to be better than BH and BY, and should be preferred to these methods in almost all scenarios. In cases where there are small effects or when controlling at moderately low α levels, the spatial relationships can be ignored since (6) tends to perform well. However, when there are very large effect sizes or spatial features, rule (18) produces both low FDR levels and high power. In all other circumstances, and in general, rule (11) is able to control FDR at below the desired level, and provides highly comparable power. Thus, we recommend rule (11) for general use, and apply it to a real data study in the following section.

VI. PATH STUDY

PATH is a large longitudinal study of aging aimed at investigating the course of mood disorders, cognition, health, and other individual characteristics across the lifespan (cf. [36]). It surveys 7485 individuals in three age groups of 20 to 24, 40 to 44, and 60 to 64 years at baseline. Follow-up is every four years over a period of 20 years. The data from the PATH project were obtained from the normative older age group (2551 at baseline). PATH surveys residents of the city of Canberra and the adjacent town of Queanbeyan, Australia, who were randomly recruited through the electoral roll. Enrollment to vote is compulsory for Australian citizens; consequently this cohort is representative of the population. The study was approved by the Australian National University Ethics Committee.

A. Data Description

Of the 2551 randomly selected older PATH participants included in the study in wave 1, 2076 consented to be contacted regarding an MRI scan. From these, a randomly selected sub-sample of 622 participants were offered an MRI scan and 478 participants eventually completed MRI scanning; 360 participants were rescanned at wave 3.

From the 360 rescanned individuals, we excluded 18 participants from the current study due to gross brain abnormalities evident in their MRI scans; another 30 participants were excluded due to a history of epilepsy, Parkinson's disease or stroke. Furthermore, 46 individuals were diagnosed with mild

cognitive impairment or dementia (based on full neurological assessment meeting clinical criteria), and were subsequently excluded. Our data consists of the wave 1 scans of the remaining 266 participants.

The 266 participants were scanned using a 1.5 Tesla Philips Gyroscan ACS-NT scanner (Philips Medical Systems, Best, the Netherlands) for T1-weighted three-dimensional structural MRI, and all participant identifiers were stripped from the resulting scans. This 266 subjects sub-sample of scans constitutes the data for our investigation. The sub-sample has mean age 63.16 (standard deviation 1.47) and a comparable gender ratio to that of the greater PATH cohort.

All data were first converted to the MINC format and then intensity non-uniformity corrected using the N3 algorithm (cf. [49]). Data were then roughly normalized by clamping the image between histogram percent critical thresholds of 0.5 and 99.5% and rescaling the resulting data between 0 and 100. The histogram clamping step is to remove small peripheral image artifacts. Data were then linearly registered to an internal average. The resulting MRIs have dimensions $264 \times 264 \times 199$ with a total brain volume of 1387661 voxels, as determined by atlas based registration to a common template (cf. [50] and [51]).

B. Hypothesis Testing

Let x_j be the age of individual $j = 1, \dots, m$, and y_{ij} be the intensity of voxel i from the MRI of individual j . Here, $m = 266$ and $n = 1387661$.

At each voxel, we model the relationship between age and intensity by the linear regression model $Y_{ij} = \beta_{i0} + \beta_{i1}x_j + E_{ij}$, where β_{i0} and β_{i1} are voxel specific parameters, and $E_{ij} \sim N(0, \sigma_i^2)$. This model allows us to perform n separate regression computations.

Since we only model the intensities by age, there is a potential for inferential bias due to the absence of other relevant factors such as education, gender, and neighboring voxel intensities in each regression. Thus, instead of the usual least-squares based inference, we opted to use the misspecification-adjusted standard errors of [52] to correct for these effects (see Appendix B of [35] for computational details).

The misspecification-adjusted standard errors allow us to perform the n voxel-specific regressions separately even though they may be related, and in absence of additional factors; this is necessary since modeling the full dependency structure among the n regression models would not be computationally feasible. The major drawbacks of the approach are as follows: the true effect-size of the modeled factor upon the voxel intensity will be biased, and the standard errors of the regression parameters are conservative (i.e. inefficient) in comparison to the correctly-specified model. The first issue is not a concern considering we are only interested in the significance (i.e. p -value) of the effect of aging upon each voxel; the p -value is still correct even if the sizes of the effect is biased. The second issue is not problematic considering it is unlikely that the correctly-specified model is determinable (with regards to the inclusion of all relevant factors, considering the complexity of biological processes); see [52] and

[53] for further details regarding the analysis of misspecified models.

Based on the misspecification-adjusted standard errors, we can compute p -values (i.e. p_i) for each voxel to test the null hypothesis that $\beta_{i1} = 0$, using a Wald-type test. That is, we test whether or not age has an effect on the intensity at each voxel, correcting for the potential effects of excluding other intensity-affecting factors and between-voxel dependencies.

C. Technical Results

After performing the hypothesis tests described above, we then assessed the p -values for indications of voxels, for which the null hypothesis was false. In Fig. 3 A, we note that the p -values do not appear uniformly distributed, thus implying that not all null hypotheses are true, under the well-behaved assumption. This implies that there may be an effect of age on intensity across some of the analyzed voxels.

The z -scores were then computed and the NC-FDR model with empirical null density was fitted to the data (see Fig. 3 B). The MCMLE of the parameter vector for density (2) was found to be

$$\begin{aligned}\hat{\theta} &= (\hat{\pi}_0, \hat{\mu}_0, \hat{\sigma}_0^2, \hat{\mu}_1, \hat{\sigma}_1^2)^T \\ &= (0.306, -0.492, 0.860, 0.915, 0.752)^T,\end{aligned}$$

which indicates that there is a large proportion of age affect voxels, and that the choice to use an empirical null was appropriate in this situation.

The MRF of Section III was then estimated with the PLIC selected neighborhood range of $d = 1$, indicating that the relationship between age-effected voxels appears to be very local in nature. The potential FDR and number of rejections for various threshold levels were then assessed via level plots (Fig. 3 C and D). The graphs suggest that controlling the FDR at $\alpha = 0.1$ would yield a rejection proportion of approximately 60% of the tested voxels.

We choose to control FDR at the $\alpha = 0.1$ level due to the potentially high power from the large proportion of rejections, and the conservative nature of rule (11). This resulted in the choice of thresholds $c_1 = 0.38$ and $c_2 = 0.21$, which indicates a strong preference towards spatial coherency of the rejected hypotheses.

Applying rule (11) with the chosen thresholds yields 826687 (59.6%) rejections, which can be compared to NC-FDR which yields 883117 (63.6%) at the same α level. Both of these results could be anticipated from Fig. 3 D. Images of rejected voxels under MRF-FDR, and NC-FDR can be found in Fig. 4.

On inspection of Fig. 4, we note that the NC-FDR images show that there are a large number of isolated rejected regions when compared to the MRF-FDR images which appear to be more spatially coherent. This is exemplified in the comparison of Sub-figure A1 to B1. In order to numerically assess the spatial coherency of the two sets of rejections, we can compute the number of isolated voxels (voxels with no same-class neighbors in its 1-range Moore-type neighborhood). Using the aforementioned metric, we find that MRF-FDR produces 177

(0.013%) isolated voxels, compared to 336 (0.024%) by NC-FDR. Thus, MRF-FDR has half as many isolated voxels as NC-FDR, which implies lower FDR levels under the spatial coherency assumption of Theorem 3.

D. Clinical Results

Using the ANIMAL+INSECT technique for structural segmentation (cf. [54]), we were able to compute the proportion of rejected voxels in various major anatomical components (see Table II).

Our finding of whole-brain effects due to age in the previous section is clinically consistent with the brain aging literature (i.e. the clinical inferences drawn from our results are compatible with those drawn from the quantitative results of articles in the literature; e.g. [2], [3], [55]–[59]), as are the age effects on the caudate nucleus (e.g. [55], [59], [60]), cerebellum (e.g. [3], [55], [61]), frontal lobe (e.g. [2]–[4]), globus pallidus (e.g. [59]), occipital lobe (e.g. [58]), parietal lobe (e.g. [2], [3], [57]), putamen (e.g. [59]), temporal lobe (e.g. [2], [3]), and thalamus (e.g. [59], [62], [63]).

VII. CONCLUSION

We have developed a spatially informed technique that is able to conservatively control the FDR by complementing the technique of [19] with an MRF structure. This was achieved through the use of MPLE and PLIC for estimation and model selection, respectively. The resulting method was shown to outperform traditional marginal techniques in terms of both false positive mitigation and power over various simulations of spatially dependent hypotheses.

Applications of MRF-FDR to the PATH study data showed that it was able to produce more spatially coherent rejections than those rejections made by noncontextual methods. This implies that there is less potential for false discovery when adopting an assumption that voxel effects are spatially dependent.

Using an anatomic segmentation, we also investigated clinical findings of our study. Our results corresponded well with the literature in terms of both whole brain aging and aging in various anatomic components.

This article has improved upon the work of [35] by suggesting a set of more interpretable rules and proving a number of technical theorems regarding the workings of the methodology; however, there are still some limitations, which need addressing in future work. For instance, we have assumed uniformity with respect to the spatial dependencies between voxels which may be untrue for the effects of some factors. To address such issues, we may allow for a separate set of MRF parameters (i.e. ψ) for different regions of interest (e.g. the different anatomical components from Section VI-D). Another solution may be to allow for each voxel to have a different set of MRF parameters (i.e. ψ_i) which can be treated as a random effect, or a random variable with a prior distribution (in the Bayesian sense); this would allow for each voxel to be differently related to its neighbors.

Furthermore, our current method has been developed under an assumption of information austerity (i.e. we only assume

Figure 3. Sub-figures A and B displays the p -values and z -scores histograms for the PATH data hypothesis tests, where the blue, red and black lines are graphs of $\pi_0 f_0(z_i)$, $(1 - \pi_0) f_1(z_i)$, and $f(z_i)$, respectively (as per Section II). Sub-figures C and D show $\widehat{mFDR}'$ and $\hat{N}_R/n \times 100\%$ for various threshold combinations, respectively.

Figure 4. Sub-figures A1, A2 and A3 show FDR control by MRF-FDR using rule (11) at the $\alpha = 0.1$ level, and B1 B2 and B3 show control by NC-FDR using rule (6) at the sample level. Sub-figures A1 and B1 are midsagittal slices, A2 and B2 are midcoronal slices and A3 and B3 are midhorizontal slices. Red indicates rejected voxels, and vice versa for black.

Table II
ANATOMICAL DECOMPOSITION WITH PROPORTIONS OF REJECTED VOXELS (FDR LEVEL $\alpha = 0.1$) AND VOLUMES (IN VOXELS).

Component	Left		Right		Total	
	Proportion	Volume	Proportion	Volume	Proportion	Volume
Brainstem	-	-	-	-	0.603	43109
Caudate Nucleus	0.733	4242	0.108	4362	0.416	8604
Cerebellum	0.617	79762	0.617	80091	0.617	159853
Fornix	0.160	873	0.105	962	0.131	1835
Frontal Lobe	0.499	259165	0.495	256852	0.497	516017
Globus Pallidus	0.998	936	0.991	932	0.995	1868
Occipital Lobe	0.555	77162	0.588	75281	0.571	152443
Parietal Lobe	0.426	100907	0.519	128996	0.478	229903
Putamen	0.989	4049	0.968	4076	0.979	8125
Temporal Lobe	0.585	153812	0.589	128952	0.587	282764
Thalamus	0.674	5636	0.435	7589	0.537	13225

knowledge of the p -values for the effect of the factor of interest at each voxel and the spatial location of said voxels) and therefore we have not discussed any issues regarding the use of complementary data. However, given complementary data is available, there are many ways in which our methodology can be expanded upon. For example, given contiguous data (i.e. information regarding the voxels which is independent of the factor of interest), we could perform a clustering of the voxels using the method of [64] before computing p -values, and conducting FDR control using MRF-FDR. Additionally, factors that are not directly of interest can be incorporated into the prior probability for each z-score via a mixture-of-experts construction within the noncontextual-phase of our method (cf. [65] and Section 5.12 of [37] regarding mixtures-of-experts); this would allow for individual-specific FDR control. We believe that these avenues represent interesting future directions.

Further from the aforementioned future directions, we hope to apply this work to other brain study scenarios to assess its robustness under different experimental conditions, and implement the method in major open-source software projects to encourage its use.

APPENDIX A PROOF OF THEOREMS

Proof of Theorem 1

The equation

$$\begin{aligned} & I \{r_1(Z_i; c_1) = 1\} \tau(Z_i) \\ &= I \{r_1(Z_i; c_1) = 1\} \mathbb{E}(H_i = 0 | I \{r_1(Z_i; c_1) = 1\}), \end{aligned}$$

implies

$$\begin{aligned} & \mathbb{E}(I \{r_1(Z_i; c_1) = 1\} \tau(Z_i)) \\ &= \mathbb{E}I \{r_1(Z_i; c_1) = 1, H_i = 0\}, \end{aligned}$$

by the law of iterative expectations.

Under the theorem's assumption we have $\text{var}(n^{-1}\bar{N}_{01}) \rightarrow 0$ and $\text{var}(n^{-1}\bar{N}_R) \rightarrow 0$, as $n \rightarrow \infty$. Thus, by application of a dependent-variable law of large numbers (e.g. Theorem 5.3 of [66]), we have $n^{-1}\bar{N}_{01} \xrightarrow{P} \mathbb{E}I \{r_1(Z_i; c_1) = 1, H_i = 0\}$ and

$n^{-1}\bar{N}_R \xrightarrow{P} \mathbb{E}I \{r_1(Z_i; c_1) = 1\}$. It follows that

$$\begin{aligned} \overline{m\text{FDR}} &= \frac{\bar{N}_{01}}{\bar{N}_R} \\ &= \frac{n^{-1}\bar{N}_{01}}{n^{-1}\bar{N}_R} \\ &\xrightarrow{P} \frac{\mathbb{E}I \{r_1(Z_i; c_1) = 1, H_i = 0\}}{\mathbb{E}I \{r_1(Z_i; c_1) = 1\}} \\ &= \frac{n\mathbb{E}I \{r_1(Z_i; c_1) = 1, H_i = 0\}}{n\mathbb{E}I \{r_1(Z_i; c_1) = 1\}} \\ &= \frac{\mathbb{E} \sum_{i=1}^n I \{r_1(Z_i; c_1) = 1, H_i = 0\}}{\mathbb{E} \sum_{i=1}^n I \{r_1(Z_i; c_1) = 1\}} \\ &= \frac{\mathbb{E}N_{01}}{\mathbb{E}N_R} \\ &= m\text{FDR}, \end{aligned}$$

by continuous mapping.

Proof of Theorem 2

Let $U_0 = \sum_{i=1}^n I \{r_1(Z_i; c_1) = 1, H_i = 0\}$ and $U_1 = \sum_{i=1}^n I \{r_1(Z_i; c_1) = 1, H_i = 1\}$, and define $V_0 = n^{-1}\mathbb{E}U_0$ and $V_1 = n^{-1}\mathbb{E}U_1$. Using this notation, we can write

$$\begin{aligned} m\text{FDR} &= \frac{\mathbb{E}U_0}{\mathbb{E}U_0 + \mathbb{E}U_1} \\ &= \frac{V_0}{V_0 + V_1} \end{aligned}$$

and

$$\text{FDR} = \mathbb{E} \left(\frac{U_0}{U_0 + U_1} \right).$$

Let $\kappa(u_0, u_1) = u_0 / (u_0 + u_1)$. By Taylor's theorem, we can write

$$\begin{aligned} \text{FDR} &= \mathbb{E}\kappa(U_0, U_1) \\ &= \frac{V_0}{V_0 + V_1} + R_1 + R_2 \\ &= m\text{FDR} + R_1 + R_2, \end{aligned}$$

where

$$\begin{aligned} R_1 &= \frac{nV_1}{(nV_0 + nV_1)^2} \mathbb{E}(U_0 - nV_0) \\ &\quad - \frac{nV_0}{(nV_0 + nV_1)^2} \mathbb{E}(U_1 - nV_1), \end{aligned}$$

$$R_2 = \sum_{k_0+k_1=2} \frac{2}{k_0!k_1!} \mathbb{E} \left((U_0 - nV_0)^{k_0} (U_1 - nV_1)^{k_1} \rho_{k_0k_1} \right),$$

and

$$\begin{aligned} & \rho_{k_0k_1}(U_0, U_1, V_0, V_1) \\ &= \int_0^1 (1-t)^2 \frac{\partial^2 \kappa}{\partial u_0^{k_0} \partial u_1^{k_1}} (tU_0 + (t-1)nV_0, tU_1 + (t-1)nV_1) dt. \end{aligned}$$

Firstly, note that $R_1 = 0$, because $\mathbb{E}(U_0 - nV_0) = \mathbb{E}U_0 - nV_0 = 0$, and similarly $\mathbb{E}(U_1 - nV_1) = 0$ by definition of V_0 and V_1 .

Now, we need to bound $|R_2|$ by a function which shrinks to zero as $n \rightarrow \infty$. Consider that for each k_0 and k_1 such that $k_0 + k_1 = 2$, we have

$$\begin{aligned} & \left| (1-t)^2 \frac{\partial^2 \kappa}{\partial u_0^{k_0} \partial u_1^{k_1}} (tU_0 + (t-1)nV_0, tU_1 + (t-1)nV_1) \right| \\ & \leq \frac{2C(1-t)^2 \max\{tU_0 + (t-1)nV_0, tU_1 + (t-1)nV_1\}}{(t(U_0 + U_1) + (1-t)(V_0 + V_1))^3} \\ & \leq \frac{2C}{(t(t-1)^{-1}(U_0 + U_1) + n(V_0 + V_1))^2} \\ & \leq \frac{2C}{(V_0 + V_1)^2 n^2}, \end{aligned}$$

and hence $|\rho_{k_0k_1}(U_0, U_1, V_0, V_1)| \leq n^{-2}C'(V_0, V_1)$, where C' is a constant, and C' is a function of V_0 and V_1 which does not depend on n . Thus, by the Cauchy-Schwarz inequality, for each k_0 and k_1 , $\left| \mathbb{E} \left((U_0 - nV_0)^{k_0} (U_1 - nV_1)^{k_1} \rho_{k_0k_1} \right) \right| \leq n^{-2}C'(V_0, V_1) \left| \mathbb{E} \left((U_0 - nV_0)^{k_0} (U_1 - nV_1)^{k_1} \right) \right|$.

Observe that for $k_0 = k_1 = 1$, we have

$$\begin{aligned} & \left| \mathbb{E}((U_0 - nV_0)(U_1 - nV_1)) \right| \\ &= \left| \mathbb{E} \sum_{i=1}^n \sum_{i'=1}^n W_{i0} W_{i1} \right| \\ &= \left| \sum_{i=1}^n \sum_{i'=1}^n \mathbb{E}(W_{i0} W_{i1}) \right| \\ &\leq n(\nu + 1), \end{aligned}$$

where $W_{ij} = I\{r_1(Z_i; c_1) = 1, H_j = 0\} - V_j$ for $j = 1, 2$, since each voxel i is at most dependent on the ν other voxels, $|\mathbb{E}(W_{i0} W_{i'1})| \leq 1$, and $\mathbb{E}W_{i0} \mathbb{E}W_{i'1} = 0$. By similar reasoning, we also have $\mathbb{E}((U_0 - nV_0)^2) \leq n(\nu + 1)$ and $\mathbb{E}((U_1 - nV_1)^2) \leq n(\nu + 1)$.

Finally, by the triangle inequality,

$$\begin{aligned} |R_2| &\leq \sum_{k_0+k_1=2} \frac{2}{k_0!k_1!} \frac{C'(V_0, V_1)}{n^2} n(\nu + 1) \\ &= \frac{C'(V_0, V_1)(\nu + 1)}{n} \\ &\rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$, as required.

Proof of Theorem 3

Like Theorem 1, the limited dependency assumption implies that $\text{var}(n^{-1}\bar{N}_{01}) \rightarrow 0$ and $\text{var}(n^{-1}\bar{N}_R) \rightarrow 0$, as $n \rightarrow \infty$. Theorem 5.3 of [66] is then applied to

yield $n^{-1}\bar{N}_{01} \xrightarrow{P} \mathbb{E}(I\{r_2(Z_{(i)}; c_1, c_2) = 1\} \tau(Z_i))$ and $n^{-1}\bar{N}_R \xrightarrow{P} \mathbb{E}I\{r_2(Z; c_1, c_2) = 1\}$. Similarly to Theorem 1, we have

$$\begin{aligned} \overline{m\text{FDR}}' &\xrightarrow{P} \frac{\mathbb{E} \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1\} \tau(Z_i)}{\mathbb{E} \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1\}} \\ &= m\text{FDR}', \end{aligned}$$

by continuous mapping.

Now, $\gamma_{1i} \leq \gamma_{2i}$ implies

$$\begin{aligned} & \mathbb{E} \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1\} \tau(Z_i) \\ &= \mathbb{E} \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1\} \gamma_{2i} \\ &\geq \mathbb{E} \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1\} \gamma_{1i} \\ &= \mathbb{E} \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1, H_i = 0\}, \end{aligned}$$

by the law of iterative expectations. Finally, we have

$$\begin{aligned} m\text{FDR}' &\geq \frac{\mathbb{E} \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1, H_i = 0\}}{\mathbb{E} \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1\}} \\ &= m\text{FDR}, \end{aligned}$$

as required. Here, $m\text{FDR}$ must exist by the theorem's assumption.

Proof of Theorem 4

In the proof of Theorem 2, set

$$\begin{aligned} U_0 &= \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1, H_i = 0\}, \\ U_1 &= \sum_{i=1}^n I\{r_2(Z_{(i)}; c_1, c_2) = 1, H_i = 1\}, \end{aligned}$$

and $W_{ij} = I\{r_2(Z_{(i)}; c_1, c_2) = 1, H_j = 0\} - V_j$, for each i and $j = 1, 2$.

Proof of Theorem 6

Observe that $\hat{\theta} \xrightarrow{P} \theta$ implies $\tau(Z_i; \hat{\theta}) \xrightarrow{D} \tau(Z_i; \theta)$ by Slutsky's theorem, and thus $\hat{T}_i \xrightarrow{D} T_i$ by the portmanteau lemma. Now consider that $\hat{T}_i \xrightarrow{D} T_i$ and $\hat{\psi} \xrightarrow{P} \psi$ imply that $\xi(\hat{T}_i, \hat{T}_{(i)}; \hat{\psi}) \xrightarrow{D} \xi(T_i, T_{(i)}; \psi)$ by Slutsky's theorem, and hence, by the portmanteau lemma,

$$\begin{aligned} & I\{r_2(\hat{Z}_{(i)}; \hat{\theta}, \hat{\psi}, c_1, c_2) = 1\} \\ &\xrightarrow{D} I\{r_2(Z_{(i)}; \theta, \psi, c_1, c_2) = 1\}. \end{aligned}$$

Lastly, as required, $\widehat{m\text{FDR}}' \xrightarrow{D} \overline{m\text{FDR}}'$ by the continuous mapping theorem.

Proof of Lemma 7

If we suppose that $g(t_i, t_{(i)}; \psi^*) = g(t_i, t_{(i)}; \psi)$ for all t_i and $t_{(i)}$, and $\psi^* \neq \psi$, then we have

$$\frac{g(1, t_{(i)}; \psi^*)}{g(0, t_{(i)}; \psi^*)} = \frac{g(1, t_{(i)}; \psi)}{g(0, t_{(i)}; \psi)},$$

which implies

$$\exp(a^* + b^* \eta(t_{(i)})) = \exp(a + b \eta(t_{(i)})).$$

Upon simplification of the expression above, we have

$$(a^* - a) + \eta(t_{(i)}) (b^* - b) = 0,$$

which only holds if $\psi^* = \psi$, or $\eta(t_{(i)}) = 0$. Thus, the result follows by contradiction.

APPENDIX B ALGORITHM DETAILS

MM Algorithm for Mixture Models

The MM algorithm for estimating θ proceeds by initializing the parameter vector by $\theta^{(0)}$. Upon letting $\theta^{(k)} = (\pi_0^{(k)}, \mu_0^{(k)}, \sigma_0^{(k)2}, \mu_1^{(k)}, \sigma_1^{(k)2})^T$ be the estimates on the k th iteration, the $(k+1)$ th iteration of the algorithm is conducted in two steps: the minorization (min) and maximization (max)-step. In the min-step, $\tau_{i0}^{(k+1)} = \tau(z_i; \theta^{(k)}) = 1 - \tau_{i1}^{(k+1)}$ is computed for each i . The max-step then requires the evaluation of $\pi_0^{(k+1)} = n^{-1} \sum_{i=1}^n \tau_{i0}^{(k+1)}$, and

$$\mu_j^{(k+1)} = \frac{\sum_{i=1}^n \tau_{ij}^{(k+1)} z_i}{\sum_{i=1}^n \tau_{ij}^{(k+1)}},$$

and

$$\sigma_j^{(k+1)2} = \frac{\sum_{i=1}^n \tau_{ij}^{(k+1)} (z_i - \mu_j^{(k+1)})^2}{\sum_{i=1}^n \tau_{ij}^{(k+1)}},$$

for each $j = 0, 1$.

The steps are iterated until convergence is achieved, whereupon the final vector of estimates is declared the MCMLE $\hat{\theta}$. Upon letting $\pi_1 = 1 - \pi_0$, the algorithm arises through applications of the minorizer:

$$\log \left(\sum_{j=0}^1 A_{ij} \right) \geq \sum_{j=0}^1 \left(\frac{B_{ij}}{\sum_{j'=0}^1 B_{ij'}} \right) \log \left(\frac{\sum_{j'=0}^1 B_{ij'} A_{ij'}}{B_{ij}} \right),$$

where $A_{ij} = \pi_j \phi(z_i; \mu_j, \sigma_j^2)$ and $\tau_{ij}^{(k)} = B_{ij} / \sum_{j'=0}^1 B_{ij'}$, for each i and j (cf. [67]). Because the algorithm is constructed via minorization, it must monotonically increase the CML at each iteration.

Newton Algorithm for Spatial Model

The PL of equation (13) can be maximized indirectly by considering the logarithm

$$\begin{aligned} \log PL(\psi; \hat{t}_1, \dots, \hat{t}_n) &= \sum_{i=1}^n \log g(\hat{t}_i, \hat{t}_{(i)}; \psi) \\ &= a \sum_{i=1}^n \hat{t}_i + b \sum_{i=1}^n \hat{t}_i \eta(\hat{t}_{(i)}) \\ &\quad - \sum_{i=1}^n \log(1 + \exp(a + b \eta(\hat{t}_{(i)}))). \end{aligned} \quad (19)$$

Let $\nabla \log PL$ and $\nabla^2 \log PL$ be the gradient and Hessian of (19), respectively, with components

$$\nabla_a \log PL = \sum_{i=1}^n \hat{t}_i - \sum_{i=1}^n \bar{\xi}(\hat{t}_i, \hat{t}_{(i)}; \psi),$$

$$\nabla_b \log PL = \sum_{i=1}^n \hat{t}_i \eta(\hat{t}_{(i)}) - \sum_{i=1}^n \eta(\hat{t}_{(i)}) \bar{\xi}(\hat{t}_i, \hat{t}_{(i)}; \psi),$$

$$\nabla_a^2 \log PL = - \sum_{i=1}^n \bar{\xi}(\hat{t}_i, \hat{t}_{(i)}; \psi) + \sum_{i=1}^n \bar{\xi}(\hat{t}_i, \hat{t}_{(i)}; \psi)^2,$$

$$\begin{aligned} \nabla_{ab}^2 \log PL &= - \sum_{i=1}^n \eta(\hat{t}_{(i)}) \bar{\xi}(\hat{t}_i, \hat{t}_{(i)}; \psi) \\ &\quad + \sum_{i=1}^n \eta(\hat{t}_{(i)}) \bar{\xi}(\hat{t}_i, \hat{t}_{(i)}; \psi)^2, \end{aligned}$$

and

$$\begin{aligned} \nabla_{b^2}^2 \log PL &= - \sum_{i=1}^n \eta(\hat{t}_{(i)})^2 \bar{\xi}(\hat{t}_i, \hat{t}_{(i)}; \psi) \\ &\quad + \sum_{i=1}^n (\eta(\hat{t}_{(i)}) \bar{\xi}(\hat{t}_i, \hat{t}_{(i)}; \psi))^2, \end{aligned}$$

where $\bar{\xi}(t_i, t_{(i)}; \psi) = 1 - \xi(t_i, t_{(i)}; \psi)$.

Using an initial estimate $\psi^{(0)}$, equation (19) can be maximized through iterative computations of the Newton step

$$\psi^{(k+1)} = \psi^{(k)} - \left(\nabla^2 \log PL(\psi^{(k)}) \right)^{-1} \nabla \log PL(\psi^{(k)}),$$

where $\psi^{(k)} = (a^{(k)}, b^{(k)})^T$ is the parameter estimate at the k th iteration. Upon convergence of the sequence of estimates, the final iterate is declared to be the MPLE $\hat{\psi}$.

ACKNOWLEDGMENT

The authors are grateful to Kaarin Anstey, Peter Butterworth, Helen Christensen, Simon Eastale, Patricia Jacomb, Luke Lloyd-Jones, Karen Maxwell, Perminder Sachdev and Ian Wood for helpful discussions on various aspects of the manuscript, and to the PATH interviewers. Furthermore, the authors are indebted to the reviewers for providing many useful comments which have lead to numerous improvements in the manuscript.

REFERENCES

- [1] C. E. Coffey, J. A. Saxton, G. Ratcliff, R. N. Bryan, and J. F. Lucke, "Relation of education to brain size in normal aging," *Neurology*, vol. 53, pp. 189–196, 1999.
- [2] E. R. Sowell, B. S. Peterson, P. M. Thompson, S. E. Welcome, A. L. Henkenius, and A. W. Toga, "Mapping cortical change across the human life span," *Nature Neuroscience*, vol. 6, pp. 309–315, 2003.
- [3] C. D. Smith, H. Chebrolu, D. R. Wekstein, F. A. Schmitt, and W. R. Markesbery, "Age and gender effects on human brain anatomy: A voxel-based morphometric study in healthy elderly," *Neurology of Aging*, vol. 28, pp. 1075–1087, 2007.
- [4] D. J. Tisserand, J. C. Pruessner, E. J. Sanz Arigita, M. P. J. van Bostel, A. C. Evans, J. Jolles, and H. B. M. Uylings, "Regional Frontal Cortical Volumes Decrease Differentially in Aging: An MRI Study to Compare Volumetric Approaches and Voxel-Based Morphometry," *NeuroImage*, vol. 17, pp. 657–689, 2002.
- [5] A. Pfefferbaum, K. O. Lim, R. B. Zipursky, D. H. Mathalon, M. J. Rosenbloom, B. Lane, C. N. Ha, and E. V. Sullivan, "Brain Gray and White Matter Volume Loss Accelerates with Aging in Chronic Alcoholics: A Quantitative MRI Study," *Alcoholism: Clinical and Experimental Research*, vol. 16, pp. 1078–1089, 1992.
- [6] A. Pfefferbaum, E. V. Sullivan, D. H. Mathalon, and K. O. Lim, "Frontal Lobe Volume Loss Observed with Magnetic Resonance Imaging in Older Chronic Alcoholics," *Alcoholism: Clinical and Experimental Research*, vol. 21, pp. 521–529, 1997.
- [7] C. M. Bennett, A. A. Baird, M. B. Miller, and G. L. Wolford, "Neuro correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction," in *15th Annual meeting of the Organization for Human Brain Mapping*, 2009.
- [8] R. G. Miller, *Simultaneous Statistical Inference*. Springer, 1981.
- [9] S. Dudoit and M. J. van der Laan, *Multiple Testing Procedures with Applications to Genomics*. Springer, 2008.
- [10] B. Efron, *Large-Scale Inference*. Cambridge University Press, 2010.
- [11] F. E. Turkheimer, C. B. Smith, and K. Schmidt, "Estimation of the Number of "True" Null Hypotheses in Multivariate Analysis of Neuroimaging Data," *Neuroimage*, vol. 13, pp. 920–930, 2001.
- [12] C. R. Genovese, N. A. Lazar, and T. Nichols, "Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate," *NeuroImage*, vol. 15, pp. 870–878, 2002.
- [13] T. Nichols and S. Hayasaka, "Controlling the familywise error rate in functional neuroimaging: a comparative review," *Statistical Methods in Medical Research*, vol. 12, p. 419–446, 2003.
- [14] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society Series B*, vol. 57, pp. 289–300, 1995.
- [15] Y. Benjamini and D. Yekutieli, "The Control of the False Discovery Rate in Multiple Testing under Dependency," *Annals of Statistics*, vol. 29, pp. 1165–1188, 2001.
- [16] C. Genovese and L. Wasserman, "Operating characteristics and extensions of the false discovery rate procedure," *Journal of the Royal Statistical Society Series B*, vol. 64, pp. 499–517, 2002.
- [17] J. D. Storey, "A direct approach to false discovery rates," *Journal of the Royal Statistical Society Series B*, vol. 64, pp. 479–498, 2002.
- [18] B. Efron, "Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis," *Journal of the American Statistical Society*, vol. 99, pp. 96–104, 2004.
- [19] G. J. McLachlan, R. W. Bean, and L. Ben-Tovim Jones, "A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays," *Bioinformatics*, vol. 22, pp. 1608–1615, 2006.
- [20] W. Sun and T. T. Cai, "Oracle and Adaptive Compound Decision Rules for False Discovery Rate Control," *Journal of the American Statistical Society*, vol. 102, pp. 901–912, 2007.
- [21] J. Xie, T. T. Cai, J. Maris, and H. Li, "Optimal false discovery rate control for dependent data," *Statistics and Its Interface*, vol. 4, pp. 417–430, 2011.
- [22] K. J. Friston, P. Jezzard, and R. Turner, "Analysis of Functional MRI Time-Series," *Human Brain Mapping*, vol. 1, pp. 153–171, 1994.
- [23] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical Parametric Maps in Functional Imaging: A General Linear Approach," *Human Brain Mapping*, vol. 2, pp. 189–210, 1995.
- [24] K. J. Worsley and K. J. Friston, "Analysis of fMRI Time-Series Revisited-Again," *NeuroImage*, vol. 2, pp. 173–181, 1995.
- [25] K. J. Worsley, J. E. Taylor, F. Tomaiuolo, and J. Lerch, "Unified univariate and multivariate random field theory," *NeuroImage*, vol. 23, pp. S189–S195, 2004.
- [26] M. P. Pacifico, C. Genovese, I. Verdinielli, and L. Wasserman, "False Discovery Control for Random Fields," *Journal of the American Statistical Society*, vol. 99, pp. 1002–1014, 2004.
- [27] J. R. Chumbley and K. J. Friston, "False discovery rate revisited: FDR and topological inference using Gaussian random fields," *NeuroImage*, vol. 44, pp. 62–70, 2009.
- [28] C. M. Bennett, G. L. Wolford, and M. B. Miller, "The principled control of false positives in neuroimaging," *Social Cognitive and Affective Neuroscience*, vol. 4, pp. 417–422, 2009.
- [29] Y. Zhang, M. Brady, and S. Smith, "Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, pp. 45–57, 2001.
- [30] S.-K. Ng and G. J. McLachlan, "Speeding up the EM algorithm for mixture model-based segmentation of magnetic resonance images," *Pattern Recognition*, vol. 37, pp. 1573–1589, 2004.
- [31] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, pp. 839–851, 2005.
- [32] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated Model-Based Tissue Classification of MR Images of the Brain," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 897–908, 1999.
- [33] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society Series B*, vol. 36, pp. 192–236, 1974.
- [34] C. Ji and L. Seymour, "A Consistent Model Selection Procedure for Markov Random Fields Based on Penalized Pseudolikelihood," *Annals of Applied Probability*, vol. 6, pp. 423–443, 1996.
- [35] H. D. Nguyen, G. J. McLachlan, A. L. Janke, N. Cherbuin, P. Sachdev, and K. J. Anstey, "Spatial False Discovery Rate Control for Magnetic Resonance Imaging Studies," in *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications*, 2013.
- [36] K. J. Anstey, H. Christensen, P. Butterworth, S. Easteal, A. Mackinnon, T. Jacomb, K. Maxwell, B. Rodgers, T. Windsor, N. Cherbuin, and A. F. Jorm, "Cohort Profile: The PATH through life project," *International Journal of Epidemiology*, vol. 41, pp. 951–960, 2012.
- [37] G. J. McLachlan and D. Peel, *Finite Mixture Models*. New York: Wiley, 2000.
- [38] B. Efron, "Size, Power and False Discovery Rates," *Annals of Statistics*, vol. 35, pp. 1351–1377, 2007.
- [39] —, "Correlation and Large-Scale Simultaneous Significance Testing," *Journal of the American Statistical Society*, vol. 102, pp. 93–103, 2007.
- [40] D. R. Hunter and K. Lange, "A Tutorial on MM Algorithms," *The American Statistician*, vol. 58, pp. 30–37, 2004.
- [41] C. Varin, "On composite marginal likelihoods," *Advances in Statistical Analysis*, vol. 92, pp. 1–28, 2008.
- [42] C. Varin, N. Reid, and D. Firth, "An Overview of Composite Likelihood Methods," *Statistica Sinica*, vol. 21, pp. 5–42, 2011.
- [43] S. Geman and C. Graffigne, "Markov Random Field Image Models and Their Applications to Computer Vision," in *Proceedings of the International Congress of Mathematicians*, 1986, pp. 1496–1517.
- [44] S. Mase, "Consistency of The Maximum Pseudo-Likelihood Estimator of Continuous State Space Gibbs Processes," *Annals of Applied Probability*, vol. 5, pp. 603–612, 1995.
- [45] T. Amemiya, *Advanced Econometrics*. Cambridge: Harvard University Press, 1985.
- [46] D. W. K. Andrews, "Generic Uniform Convergence," *Econometric Theory*, vol. 8, pp. 241–257, 1992.
- [47] C. D. Stanford and A. E. Raftery, "Approximate Bayes Factor for Image Segmentation: The Pseudolikelihood Information Criterion (PLIC)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1517–1520, 2002.
- [48] G. Schwarz, "Estimating the Dimensions of a Model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.
- [49] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 87–97, 1998.
- [50] V. S. Fonov, A. C. Evans, R. C. McKinstry, C. R. Alml, and D. L. Collins, "Unbiased nonlinear average age-appropriate brain templates from birth to adulthood," *NeuroImage*, vol. 47, p. S102, 2009.
- [51] V. Fonov, A. C. Evans, K. Botteron, C. R. Alml, R. C. McKinstry, D. L. Collins, and the Brain Development Cooperative Group, "Unbiased average age-appropriate atlases for pediatric studies," *NeuroImage*, vol. 54, pp. 313–327, 2011.

- [52] H. White, "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, vol. 50, pp. 1–25, 1982.
- [53] —, *Estimation, Inference and Specification Analysis*. Cambridge University Press, 1994.
- [54] D. L. Collins, A. P. Zijdenbos, W. F. C. Baare, and A. C. Evans, "ANIMAL+INSECT: Improved Cortical Structure Segmentation," in *IPMI Lecture Notes in Computer Science*. Springer, 1999, vol. 1613, pp. 210–223.
- [55] N. Raz, U. Lindenberger, K. M. Rodrigue, K. M. Kennedy, D. Head, A. Williamson, C. Dahle, and D. G. J. D. Acker, "Regional Brain Changes in Aging Healthy Adults: General Trends, Individual Differences and Modifiers," *Cerebral Cortex*, vol. 15, pp. 1676–1689, 2005.
- [56] N. Raz and K. M. Rodrigue, "Differential aging of the brain: Patterns, cognitive correlates and modifiers," *Neuroscience and Biobehavioral Reviews*, vol. 30, pp. 730–748, 2006.
- [57] E. Courchesne, H. J. Chisum, J. Townsend, A. Cowles, J. Covington, B. Egaas, S. Hinds, and G. A. Press, "Normal Brain Development and Aging: Quantitative Analysis at in Vivo MR Imaging in Healthy Volunteers," *Neuroradiology*, vol. 216, pp. 672–682, 2000.
- [58] D. H. Salat, R. L. Buckner, A. Z. Snyder, D. N. Greve, R. S. R. Desikan, E. Busa, J. C. Morris, A. M. Dale, and B. Fischl, "Thinning of the Cerebral Cortex in Aging," *Cerebral Cortex*, vol. 14, pp. 721–730, 2004.
- [59] K. B. Walhovd, A. M. Fjell, I. Reinvang, A. Lundervold, A. M. Dale, D. E. Eilertsen, B. T. Quinn, D. Salat, N. Makris, and B. Fischl, "Effects of age on volumes of cortex, white matter and subcortical structures," *Neurobiology of Aging*, vol. 26, pp. 1261–1270, 2005.
- [60] K. Yamashita, T. Yoshiura, A. Hiwatash, T. Noguchi, O. Togao, Y. Takayama, E. Nagao, H. Kamano, M. Hatakenaka, and H. Honda, "Volumetric asymmetry and differential aging effect of the human caudate nucleus in normal individuals: a prospective MR imaging study," *Journal of Neuroimaging*, vol. 21, pp. 34–37, 2011.
- [61] A. R. Luft, M. Skalej, J. B. Schulz, D. Welte, R. Kolb, K. Burk, T. Klockgether, and K. Voigt, "Patterns of Age-related Shrinkage in Cerebellum and Brainstem Observed In Vivo Using Three-dimensional MRI Volumetry," *Cerebral Cortex*, vol. 9, pp. 712–721, 1999.
- [62] E. J. Hughes, J. Bond, P. Svrckova, A. Makropoulos, G. Ball, D. J. Sharp, A. D. Edwards, J. V. Hajnal, and S. J. Counsell, "Regional changes in thalamic shape and volume with increasing age," *NeuroImage*, vol. 63, pp. 1134–1142, 2012.
- [63] E. V. Sullivan, M. Rosenbloom, K. L. Serventi, and A. Pfefferbaum, "Effects of age and sex on volumes of the thalamus, pons, and cortex," *Neurobiology of Aging*, vol. 25, pp. 185–192, 2004.
- [64] R. Heller, D. Stanley, D. Yekutieli, N. Rubin, and Y. Benjamini, "Cluster-based analysis of FMRI data," *NeuroImage*, vol. 33, pp. 599–608, 2006.
- [65] M. I. Jordan and R. A. Jacobs, "Hierarchical Mixtures of Experts and the EM Algorithm," *Neural Computation*, vol. 6, pp. 181–214, 1994.
- [66] D. D. Boos and L. A. Stefanski, *Essential Statistical Inference: Theory and Methods*. Springer, 2013.
- [67] H. Zhou and K. Lange, "MM Algorithms for Some Discrete Multivariate Distributions," *Journal of Computational and Graphical Statistics*, vol. 19, pp. 645–665, 2010.