8-2015

Thresholding of Statistical Maps in Functional Neuroimaging via Independent Filtering

Jaqueline Kwiasowski
Clemson University, jkwiaso@clemson.edu

Thresholding of Statistical Maps in Functional Neuroimaging via Independent Filtering

---

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Mathematical Sciences

---

by
Jaqueline Kwiasowski
August 2015

Accepted by:
Dr. Andrew Brown, Committee Chair
Dr. Taufiquar Khan
Dr. Patrick Gerard

Abstract

The high dimension of functional magnetic resonance imaging (fMRI) data causes problems in finding effective thresholds for voxelwise test statistics. Due to the enormous number of voxels, adjustment for multiple testing is necessary. But such an adjustment can lead to low statistical power. It has been shown that filtering the test statistics to reduce the number of tests being performed potentially increases the number of discoveries. However, some filter-test combinations can result in loss of control over the false discovery rate. We present an independent filtering approach which avoids this issue. Independent filtering uses filter-test combinations such that the filter is independent from the test statistic, leaving the null distribution of the test statistic unchanged. Applying the procedure to fMRI data, we show that when a voxelwise general linear model is fit, filtering by magnitude of the stimulus coefficient followed by a procedure which controls the FDR even under arbitrary p-value dependence structures, increases the number of discoveries. Thus, we demonstrate that independent filtering has the potential to increase power while controlling the false discovery rate.

Acknowledgments

I would like to express my gratitude to my advisor Dr. Andrew Brown. His guidance, encouragement and expertise have been invaluable throughout all stages of my research and writing. I am extremely thankful to him for helpful suggestions and comments on my thesis drafts which have contributed greatly to the improvement of the work.

I would also like to thank the rest of my committee, Dr. Patrick Gerard and Dr. Taufiquar Khan, for agreeing to serve on my committee, their time and effort.

Being a student in the Mathematical Sciences graduate program for the last year has been a great and valuable experience. I thank the Department of Mathematical Sciences, Dr. Peter Maass and Dr. Ronald Stöver from the University of Bremen for giving me the opportunity to participate in the exchange program. My special thanks to Dr. Taufiquar Khan for providing his help and advice with all difficulties and challenges I had to face as an international student.

Finally, I wish to thank my friends and my sister for moral support, encouragement and cheering me up when needed.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Functional magnetic resonance imaging (fMRI) is a procedure that uses magnetic resonance imaging to indirectly measure neuronal activity by detecting changes in the local oxygenation of blood. The ultimate goal is to map out neuronal connectivity in the brain associated with certain behaviors. When neurons are activated, the amount of blood flow in that area increases. The increase in blood flow causes a relative surplus in local blood oxygenation. FMRI measures signals throughout the brain that depend on the change of the oxygen level. This is called the blood oxygenation level dependent (BOLD) signal. There is much literature providing more detailed information about BOLD signal generation (e.g. Ogawa et al., 1992; Smith et al., 2001; Huettel et al., 2009). However, the exact relation between neuronal activity and the BOLD signal is only partly understood (Poldrack et al., 2011).

FMRI plays an important role in current cognitive, clinical, and social psychology research. It has seen an increasing interest since its development in the 1990s. The main reason is that it has the advantage of being non-invasive and safe in contrast to preceding methods. The first non-invasive neuroimaging method was positron emission tomography (PET). Although its invention was a breakthrough in

neuroscience, PET had the disadvantage that radiation exposure was necessary. Due to safety concerns and the fact that PET systems were barely available, its use was very restricted. FMRI was safe and scanners capable of performing it were common in medical centers. Furthermore, not only was its spatial resolution better, but scans were taken much more quickly (Lindquist et al., 2008; Poldrack et al., 2011).

There are different types of fMRI studies. *Resting-state fMRI* studies the behavior of a brain in a resting state, when no task is performed. These studies can be used to examine functional connectivity such as the default network (Buckner et al., 2008). Other fMRI studies analyze task-related activation. During such an fMRI experiment a subject is exposed to a sequence of stimuli while a series of MRI scans are taken slicewise. To generate a three dimensional brain image, the brain is partitioned into slices, each consisting of a large number of equally-sized volume elements (voxels) arranged in grids. A coordinate system is used to relate each voxel to a spatial location in the brain, where x represents the left-right axis, y represents the anterior-posterior axis, and z represents the inferior-superior axis (Poldrack et al., 2011). The number and size of voxels depend on the desired resolution. An MRI scan measures the BOLD signal at each of often more than 100,000 of voxels repeatedly. The resulting data consist of an enormous number of time series of BOLD signal measurements.

To study these data, it is common to use a voxelwise linear model. In order to detect voxels at which the BOLD signal changes in response to the stimuli, voxelwise hypothesis testing is performed. Results are typically summarized in *statistical parametric maps* where each voxel is colored coded according to the t-statistic value (Friston et al., 1994). Hypothesis tests across all voxels are performed simultaneously.

If many tests are performed simultaneously, the probability among all tests that voxels are identified as active when they are not becomes very high. This phe-

nomenon is known as the *multiple testing problem*. The first attempts to address the problem of multiple comparisons begin with *Fisher's Least Significant Difference Test* (Fisher, 1935). Other methods follow, among these *Tukey's* and *Scheffe's methods* (Tukey, 1949; Scheffe, 1953). One of the currently most well-known methods is the *Bonferroni correction*. The Bonferroni inequality appears in Bonferroni (1935, 1936). However, this is not used as a correction procedure until 1959 (Dunn, 1959, 1961). The method controls the *family wise error rate* (FWER), which is the probability that one or more false positives occur. It adjusts the significance level by dividing it with the total number of tests being performed. The main issue with this method is that it will be too conservative if the number of tests is large, because the adjusted threshold for p-values will be too small. Benjamini and Hochberg (1995) point out that often times it is more important to consider the number of false discoveries than the question of whether any false discovery is made. They present a procedure controlling the *false discovery rate* (FDR), the expected proportion of false positives. Bourgon et al. (2010) argue that any procedure that adjusts for multiple testing can result in low power when applied to large-scale datasets. That is one of the reasons why a more advanced approach is necessary for fMRI.

In this work we study the application of *independent filtering* to fMRI data. Independent filtering is a two-stage approach, which first filters the test statistics and then only tests the remaining set (Bourgon et al., 2010). We analyze filtering fMRI data by two different criteria and then apply an FDR-controlling procedure to the remaining voxels. This approach aims to control the FDR while maintaining power.

The thesis is structured as follows. In Chapter 2 we take a closer look at the statistical modeling of fMRI data and some of the challenges that occur with their data analysis. Some methods for dealing with the thresholding problem are introduced. Then we present independent filtering, the two-stage approach mentioned above, in

detail. In Chapter 3, results are presented and discussed when applying our approach to real data. Finally, we summarize our results and conclusions briefly and discuss the need for further research in Chapter 4.

Chapter 2

Methods

In this Chapter we discuss in detail one approach to the statistical analysis of fMRI data. We present a voxelwise general linear model and discuss issues associated with its use. We also introduce our approach for dealing with these issues, which we apply to real fMRI data in Chapter 3.

2.1 The Voxelwise General Linear Model

A common approach to fMRI data analysis is to consider a general linear model for each voxel. The general linear model in its simplest matrix form is

$$\boldsymbol{Y}_{b,v} = \boldsymbol{X}_b\boldsymbol{\beta}_{b,v} + \boldsymbol{\epsilon}_{b,v}, \quad b = 1, \ldots, B; v = 1, \ldots, V_b, \tag{2.1}$$

where B is the number of brains and V_b the number of voxels in brain b. $\boldsymbol{Y}_{b,v}$ is the response, which is the standardized BOLD signal for voxel v in brain b in our case. $\boldsymbol{X}_b$ is the design matrix consisting of predictors for each voxel in brain b, $\boldsymbol{\beta}_{b,v}$ is the unknown coefficients of the predictors, and $\boldsymbol{\epsilon}_{b,v}$ is the error, which we assume to be

normally distributed with mean zero and covariance $\sigma^2_{b,v}\boldsymbol{V}$, where $\boldsymbol{V}$ is a symmetric positive definite matrix (Lazar, 2008). If we assume that the BOLD signal is not temporally correlated within a voxel, $\boldsymbol{V}$ is the identity matrix. For our analysis, we standardize the BOLD signal. In other words, if $Y^*_{b,v,k}$ denotes the raw signal in brain b, voxel v, at time k, where $k = 1, \ldots, T$, then the standardized signal is given by $Y_{b,v,k} = \frac{Y^*_{b,v,k} - \overline{Y^*_{b,v}}}{s^*_{b,v}}$, where $s^*_{b,v}$ is the standard deviation and $\overline{Y^*_{b,v}}$ is the mean of $\boldsymbol{Y}^*_{b,v}$. The design matrix reflects the stimulus that is present / absent at each timepoint and also contains covariates which account for linear trend and motion correction. A detailed description of the design matrix is provided below. We use the voxelwise general linear model to detect those voxels at which the coefficient corresponding to the stimulus is significantly non-zero.

Temporal Correlation

FMRI data are highly spatially and temporally correlated. Temporal correlation occurs within each voxel, whereas spatial correlation exists between voxels. Methods which do not account for either can result in reduced efficiency or misstatement of the false discovery rate (Purdon and Weisskoff, 1998). Models which describe spatial and temporal correlation are more realistic, but the increase in complexity leads to higher computational costs. As fMRI data sets are typically very large, this tradeoff needs to be considered. There are even cases for which the simple general linear model that assumes independence across voxels and timepoints provides useful results (Lazar, 2008).

One way to account for temporal correlation is to extend the general linear model so that the error term becomes an autoregressive process of small order. This means that the the elements of the unknown covariance matrix $\boldsymbol{V}$ depend on an autocovariance function between timepoints (Marchini and Ripley, 2000; Lazar, 2008).

We choose order 1 (AR(1)) for our model, so that the error at time k depends on the error at time $k-1$. Thus, the error term at time k is given by

$$\epsilon_{b,v,k} = \phi\epsilon_{b,v,k-1} + W_{b,v,k},$$

where ϕ is a constant satisfying $|\phi| < 1$ and $W_{b,v,k}$ is white noise (Poldrack et al., 2011; Shumway and Stoffer, 2011). Woolrich et al. (2001) study different methods accounting for temporal correlation and show that differences in efficiency are very small. This leads to the conclusion that as long as we account for temporal correlation, it does not matter which method we choose.

Hemodynamic Response and Convolution

Neuronal activity lasts only milliseconds, whereas the changes in blood flow evolve over several seconds with a delay between the actual activation and the onset of changes in oxygenation. The *hemodynamic response function* (HRF) is a model to describe those changes in the BOLD signal at a single voxel. Within the first two seconds of the onset of neuronal activation, an initial dip in the BOLD signal with relatively small amplitude has been observed in some studies. The blood flow takes about 4-6 seconds to reach its maximum observed amplitude, the peak height. A long undershoot of up to 20 seconds after the onset follows until the signal returns to baseline. The characteristics of the HRF are illustrated in Figure 2.1.

Another important characteristic of the HRF is *linear time invariance*. Studies suggest that the relationship between the BOLD signal and the neuronal response is linear time invariant for the most part (Dale and Buckner, 1997). *Linear* means that if we scale the neuronal response by a constant factor, then the BOLD response will be scaled by the same amount. *Time invariance* is characterized by the BOLD

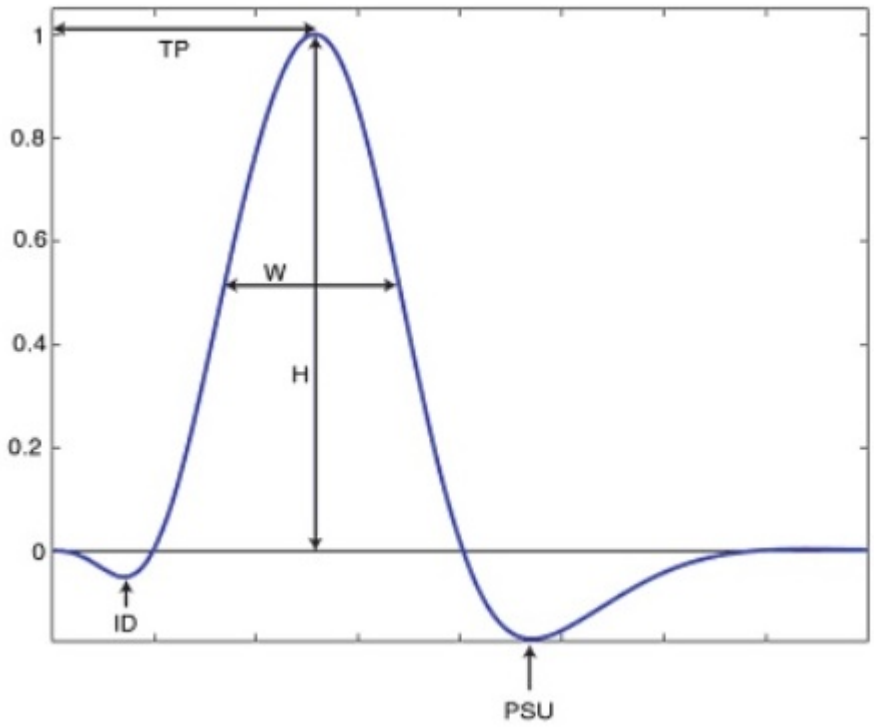

Figure 2.1: Characteristics of the hemodynamic response to describe the HRF are the initial dip (ID), the time from stimulus until peak (TP), the maximum height (H) of the HRF, the width at half the height (W) and the poststimulus undershoot (PSU). Source: Poldrack et al. (2011).

response shifting by the same amount of time as any shift in the time of activation (Poldrack et al., 2011).

There are different approaches to estimate the hemodynamic response function (HRF). The most common ones are parametric approaches, where some family of statistical distributions model the BOLD response curve. The first family suggested was the Poisson family (Friston et al., 1994). But the gamma family, proposed by Lange and Zeger (1997), is used more often as it is continuous and more flexible than the Poisson family, which only has one parameter to describe the HRF. Since the BOLD response is continuous over time, a continuous family seems to be a better choice than a discrete one like Poisson (Lazar, 2008). As it is difficult to model all

the fine details of the BOLD response with a single gamma family, Friston et al. (1998) suggest taking the difference of two gamma functions and using the parameter estimates given by Glover (1999). We follow this suggestion and thus model the HRF at time t by

$$h(t) = \left(\frac{t}{\tau_1}\right)^{\delta_1} \exp\left[-\frac{\delta_1}{\tau_1}(t-\tau_1)\right] - c\left(\frac{t}{\tau_2}\right)^{\delta_2} \exp\left[-\frac{\delta_2}{\tau_2}(t-\tau_2)\right],$$

where $\tau_1 = 0.9\delta_1, \tau_2 = 0.9\delta_2, \delta_1 = 6, \delta_2 = 12$ and $c = 0.35$. Those values correspond to assumptions regarding the peak time and undershoot (Glover, 1999).

As explained above, the HRF is assumed to be time invariant. This assumption allows us to use a convolution operation to create the convolved stimulus trail that is used as a predictor for our general linear model (Poldrack et al., 2011). The predictor is obtained by a convolution of the stimulus trail $s(t)$, a binary time series corresponding to the blocks of fixation and stimulus, and a model for the hemodynamic response function, $h(t)$, so that

$$x(t) = \int_0^\infty h(u)\, s(t-u)\, du.$$

Thus the predictor which is included in the design matrix of the general linear model will be the convolution at each time t. Say scans are taken T times, the corresponding column in our design matrix is then $(x(1), x(2), \ldots, x(T))'$ (Lazar, 2008). The convolved stimulus is illustrated in Figure 2.2.

The Design Matrix

The design matrix incorporates the convolved stimulus, a correction for linear trend and six estimated correction parameters for head motion. The correction parameters

are specified for each brain individually, since each person moves their head differently throughout the experiment. They account for translational and rotational changes in head position. In more detail, the design matrix $\boldsymbol{X}_b$ consists of the following columns:

- $\boldsymbol{t}$: the vector of T timepoints $\boldsymbol{t} = (1, 2, \ldots, T)'$ is used to correct for a linear trend in the BOLD signal
- the convolved stimulus trail $(x(1), x(2), \ldots, x(T))'$
- $\boldsymbol{z}_{b,\text{trans},x}, \boldsymbol{z}_{b,\text{trans},y}, \boldsymbol{z}_{b,\text{trans},z}$: motion correction parameters for translation in the x-,y- and z-axis
- $\boldsymbol{z}_{b,\text{rot},x}, \boldsymbol{z}_{b,\text{rot},y}, \boldsymbol{z}_{b,\text{rot},z}$: motion correction parameters for rotation around the x-,y- and z-axis

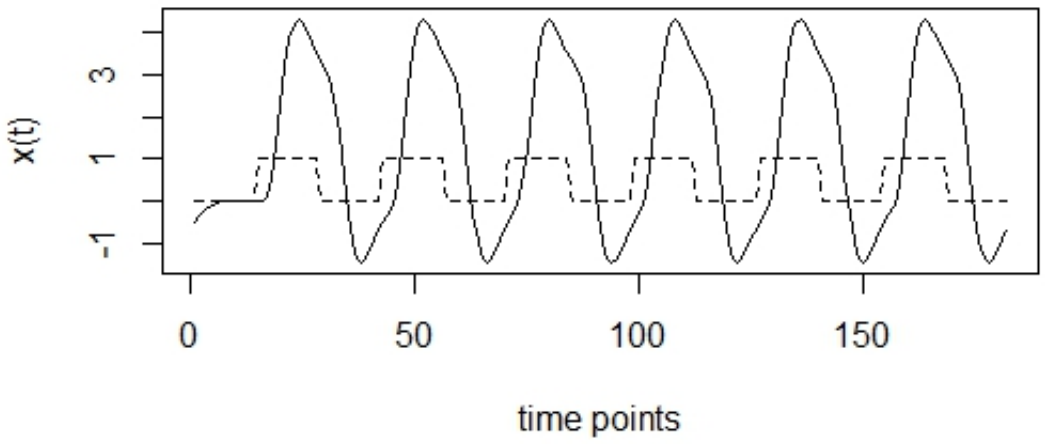

Figure 2.2: An example of a convolved stimulus trail and the original stimulus trail (dashed line). The heights of the original stimulus trail are dummy indicator variables indicating when the stimulus is present and when it is absent, thus they do not have a meaning.

The model we consider can be written as

$$\begin{aligned}
\boldsymbol{Y}_{b,v} &= \boldsymbol{X}_b\boldsymbol{\beta}_{b,v} + \boldsymbol{\epsilon}_{b,v} \\
&= \beta_{0,b,v}\mathbf{1} + \beta_{1,b,v}\boldsymbol{t} + \beta_{2,b,v}\boldsymbol{x} + \beta_{3,b,v}\boldsymbol{z}_{b,\text{trans},x} + \beta_{4,b,v}\boldsymbol{z}_{b,\text{trans},y} \\
&\quad + \beta_{5,b,v}\boldsymbol{z}_{b,\text{trans},z} + \beta_{6,b,v}\boldsymbol{z}_{b,\text{rot},x} + \beta_{7,b,v}\boldsymbol{z}_{b,\text{rot},y} + \beta_{8,b,v}\boldsymbol{z}_{b,\text{rot},z} + \boldsymbol{\epsilon}_{b,v},
\end{aligned} \tag{2.2}$$

where $\boldsymbol{x} = (x(1), x(2), \ldots, x(T))'$ and the error structure is AR(1) as described before.

2.2 Thresholding Statistical Parametric Maps

It is common to detect active voxels in neuroimaging data by performing voxelwise hypothesis tests and thresholding the resulting test statistics. After fitting the general linear model to the data, we can perform hypothesis tests. Testing voxel by voxel, the null hypothesis is that the voxel is not activated in response to the task whereas rejecting it implies that the voxel shows task-related activation. A test statistic is computed for each voxel and those voxels corresponding to the test statistics exceeding the threshold are declared active.

The results for each slice are often displayed in statistical parametric maps (Friston et al., 1994). In a *t-map*, each voxel is colored based on the t-statistic value corresponding to the estimated coefficient of the convolved stimulus. To generate the t-maps in our study, we first fit model (2.2) to our data and use the test statistic $T_{\widehat{\beta_{2,b,v}}} = \frac{\widehat{\beta}_{2,b,v}}{\text{se}(\widehat{\beta}_{2,b,v})}$, where $\text{se}(\widehat{\beta}_{2,b,v})$ is the standard error of the estimated coefficient $\widehat{\beta}_{2,b,v}$ corresponding to the convolved stimulus timecourse $x(t)$. An example of a t-map is shown in Figure 2.3. Such a t-map can give some information about the activity structure, but should be used carefully (Lindquist et al., 2008). We need to consider that we have a large number of voxels. Our data, for example, contain 1920 voxels

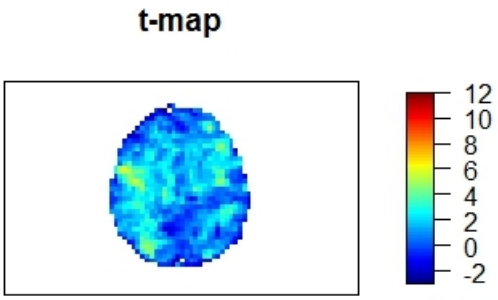

Figure 2.3: Example of a t-map of the test statistics generated from fitting voxelwise general linear models to observed BOLD signals. Each voxel is colored based on its t-value.

per slice, before masking out voxels that are outside of the brain.

Thresholding and False Discovery Rates

In general, we are mainly interested in controlling the proportion of hypothesis rejections that are false positives. The four possible statistical decisions are summarized in Table 2.1 (Genovese et al., 2002).

One of the main problems in the statistical analysis of fMRI data is to choose a threshold for which the voxels corresponding to the exceeding test statistics are declared active. There are different methods to deal with this problem and to correct for multiple comparisons. The common goal is to limit the number of false discoveries, V_{10}. One traditional approach is the Bonferroni correction for the FWER which is described in Chapter 1. This is a very conservative method and, in practice, too

	Declared active	Declared inactive	Total
Truly active	V_{00}	V_{01}	$V_{0.}$
Truly inactive	V_{10}	V_{11}	$V_{1.}$
Total	$V_{.0}$	$V_{.1}$	V

Table 2.1: Tabulation of outcomes when performing multiple hypothesis tests

conservative when applied to fMRI data. Due to the large number of tests required, the Bonferroni adjustment is not only preventing false discoveries but also true ones, resulting in low statistical power. Over the last decade it has become popular in fMRI research to use methods which control the expected proportion of false discoveries, the *false discovery rate* (FDR), rather than just the probability that even one false discovery is made. Using the notation from Table 2.1, the FDR is given by

$$\text{FDR} = \frac{V_{10}}{V_{00} + V_{10}} = \frac{V_{10}}{V_{\cdot 0}}.$$

If no hypothesis is rejected, the FDR is set to zero, by definition (Benjamini and Hochberg, 1995; Lazar, 2008). The procedure for controlling the FDR introduced by Benjamini and Hochberg (1995) is as follows:
To test n null hypotheses $H_0^1, H_0^2, \ldots, H_0^n$ with p-values $p_1, p_2, \ldots, p_n$,

1. Order p-values so that $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(n)}$ and let $H_{(j)}$ be the null hypothesis corresponding to $p_{(j)}$.

2. Let l be the largest j for which $p_{(j)} \leq \frac{j}{n}q$, where q is a constant.

3. Reject $H_{(1)}, H_{(2)}, \ldots, H_{(l)}$.

It can be shown that if the test statistics are independent, this procedure controls the FDR at q, which means that $\mathbb{E}\,(\text{FDR}) \leq q$, where $\mathbb{E}\,(\text{FDR})$ is the expected value of the FDR (Benjamini and Hochberg, 1995). It is important to point out that, according to Benjamini and Hochberg (1995), there is a gain in power compared to controlling the FWER. However, if the null hypothesis is true everywhere, the FDR is identical to the FWER. Many more approaches toward controlling the FDR exist. Some of them are described by Lazar (2008).

Independent Filtering

Corrections for multiple testing procedures can lead to relatively low statistical power. The reduction in power gets worse with the number of performed hypotheses tests. Bourgon et al. (2010) suggest a two-stage procedure that tests only those variables which pass a certain filter. Filtering can potentially increase the number of rejections by reducing the number of tests performed. However, some filter-test pairs might result in a misstatement of the FDR and overly optimistic p-values. An increase in discoveries is only helpful if the FDR is still under control, so that the rejections tend to be associated with true non-null cases. Bourgon et al. (2010) argue that *independent filtering* prevents this issue and can indeed increase power. Independent filtering means that filter-test pairs "are independent under null hypothesis but correlated under the alternative"(Bourgon et al., 2010). The goal of this procedure is to increase the number of discoveries while maintaining control over the FDR.

To apply the procedure, an appropriate choice of filter and test statistic needs to be found. In the first stage, variables are filtered by the chosen filtering criterion. This is done for a range of cutoffs, so that different sizes of fractions are filtered out. In the second stage, the chosen test is applied to the remaining sets. Generally, an efficient choice for a filter is one which reduces the number of hypothesis tests to be performed in stage two, while only filtering out variables which are of no interest. However, to guarantee control over the FDR, further conditions need to hold. Bourgon et al. (2010) examine those conditions and found that a filter-test combination should satisfy the *marginal independence criterion*. The marginal independence criterion is that for true null hypotheses, "the conditional marginal distributions of the test statistics after filtering are the same as the unconditional distributions before filtering" (Bourgon et al., 2010). Thus the independent filtering method uses filters that are

independent of the test statistic, so that after filtering the null distribution of the test statistic remains unchanged.

The marginal independence criterion ensures that even after filtering we can use the unconditional null distribution to find p-values for an FDR-controlling procedure, because the conditional probabilities to compute the p-values after filtering are equivalent to the unconditional probabilties before filtering. FDR-controlling procedures which make no requirements for the joint distribution of the p-values can therefore be applied directly to the p-values after filtering. Bourgon et al. (2010) point out that the method introduced by Benjamini and Yekutieli (2001) is one such procedure.

2.3 Two-stage Approach: Filtering and FDR-control

We first filter the data and then apply an FDR-controlling method to the remaining set. Our goal is to maximize the number of discoveries while still controlling the FDR. We consider two different filters. For the first one, we fit a general linear model to the data without the convolved stimulus, so that only linear trend and motion correction are applied. Then we filter by residual variance setting θ-quantiles as cutoffs, where θ is in $[0, 1]$. Those voxels corresponding to the lowest residual variances, that means the lowest $\theta(100)\%$ residual variances, are removed from the set. Then the full general linear model (2.2) is used to fit the remaining voxels. For the other filter, we use model (2.2) and then filter by magnitude of $\widehat{\beta}_{2,b,v}$, the estimated regression coefficients corresponding to the convolved stimulus. Again, we filter out those voxels corresponding to the lowest $\theta(100)\%$ coefficient magnitudes. In the next step we apply an FDR-controlling procedure to the test statistics corresponding to the remaining voxels. The procedure that we are using is originally introduced by

Benjamini and Yekutieli (2001) and then suggested for fMRI data by Genovese et al. (2002). The method is guaranteed to control the FDR so that

$$\mathbb{E}\left(\text{FDR}\right) \leq \frac{V_1}{V} q \leq q,$$

where $\mathbb{E}$ denotes expected value. Moreover, it controls the FDR even under arbitrary dependencies of the p-values. As filtering drastically alters the correlation structure among the test statistics, this is an important feature necessary for our approach. The procedure can be described as follows:

1. Pick an FDR bound q, where $0 \leq q \leq 1$. q is the maximum false discovery rate that the researcher is willing to tolerate.

2. Sort p-values in ascending order: $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(V)}$.

3. Let k be the largest i, so that the following is satisfied:

$$p_{(i)} \leq \frac{i}{V} \frac{q}{c_V},$$

   where c_V is a constant determined based on the assumptions about the distribution of p-values across voxels.

4. The voxels $v_{(1)} \ldots v_{(k)}$ corresponding to the p-values $p_{(1)} \ldots p_{(k)}$ are declared active.

When assuming that all p-values across voxels are independent the constant should be $c_V = 1$. Note that if we we make this choice, the procedure is the same as the original procedure propsed by Benjamini and Hochberg (1995). For any other joint distribution of p-values the constant should be set as $c_V = \sum_{i=1}^{V} \frac{1}{i}$. This constant is the modification made by Benjamini and Yekutieli (2001) leading to an algorithm

which is guaranteed to control the FDR under arbitrary dependence structure of the p-values. Our general linear model accounts for temporal correlation only. However, literature suggests that the BOLD signal is also spatially correlated across voxels (e.g., Purdon et al., 2001; Lazar, 2008; Poldrack et al., 2011). Thus, $c_V = \sum_{i=1}^{V} \frac{1}{i}$ is a better choice. By choosing this constant, our approach accounts for both temporal and spatial correlation.

Chapter 3

Results

We consider fMRI data obtained from a study with 64 subjects. We have two samples, the control sample consisting of 32 healthy subjects and the case sample consisting of 32 schizophrenia patients. Each subject was given a sequence of visual stimuli during an MRI. The sequence of visual stimuli alternated between seven blocks at 28 seconds fixation and six blocks at 28 seconds prosaccade trials. A *prosaccade* is an eye movement task. The subject is asked to keep their head still and look in the direction of the target, which is moved from directly in front of the subject to the left or right. (Experiments involving prosaccade are studied in Dyckman et al., 2007; Camchong et al., 2008, and elsewhere.) Fixation required the patient to look at a fixed target in front of them. The stimulus pattern is displayed in Figure 3.1.

For each brain, data were generated for 38 slices. Each slice consists of 1920 voxels with size 3.4x3.4x4mm, arranged in a 40x48 grid. Every 2 seconds the BOLD signal was measured at each voxel, resulting in a time series consisting of 182 measurements. Preprocessing steps were applied to the resulting data. Preprocessing included despiking, slice timing correction, alignment of functional images to anatomical images, scaling of voxel time series to the mean of 100, and warping the functional

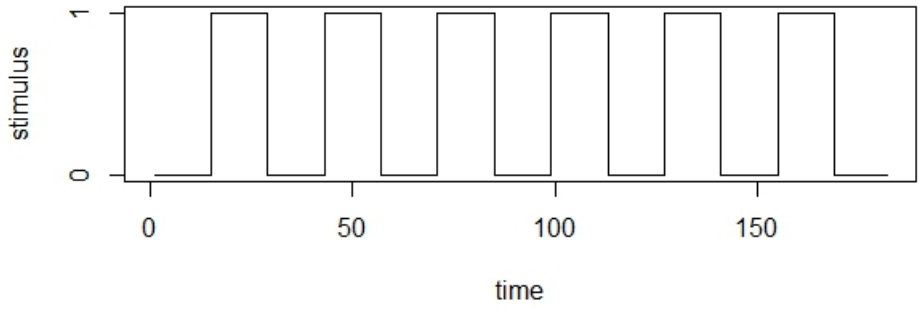

Figure 3.1: Stimulus over time for the fMRI data analyzed here, where fixation is indicated by 0 and prosaccade by 1.

images into Talaraich space, a three-dimensional coordinate space defined by a set of anatomical landmarks (Poldrack et al., 2011), to make them comparable across brains. In our study, we mask out the voxels lying outside the brain and thus only consider the time series at each voxel corresponding to the brain. Throughout this study we only consider a single slice of one subject's brain from the control group, consisting of 897 voxels inside the brain region. All results presented in this Chapter are based on the observed BOLD signals of this particular brainslice.

Figure 3.2 shows t-maps of our considered brainslice when the model (2.2) with an AR(1) error structure is fit to the data. On the right, the model (2.2) is fit but this time we let the errors be independent and identically normally distributed (i.i.d.) with mean zero and variance σ_v^2.(Note that we omit the index for brain since we are only considering one.) A general linear model with i.i.d. error structure is sometimes used in practice. Since the Figure shows higher values for the t-statistics, we can conclude that fitting the data assuming independence results in more voxels being declared active. But as discussed above, ignoring the temporal correlation leads to

a rather unrealistic model and we can conclude that the results are overly optimistic and thus not useful.

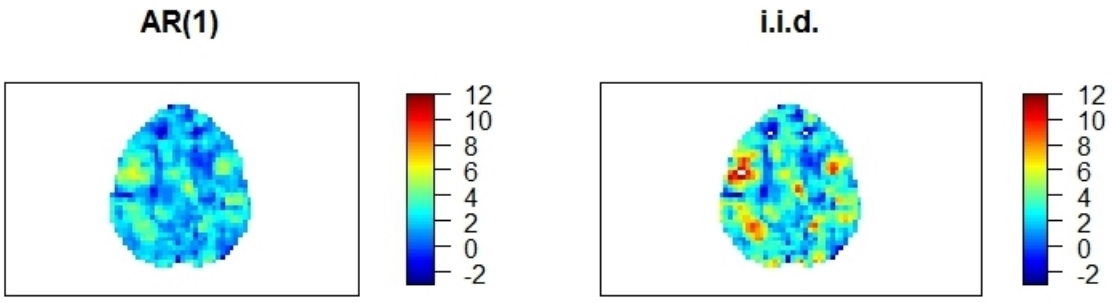

Figure 3.2: T-maps displaying values of test statistics corresponding to the change in BOLD signal due to the stimulus at each voxel

We apply our two-stage approach described in Chapter 2 to the data. To account for an arbitrary p-value distribution, we choose the constant $c_V = \sum_{i=1}^{V} \frac{1}{i}$ for the threshold. Furthermore, we choose the bound $q = 0.05$, which is the maximum FDR that we are willing to tolerate, on average.

We consider two different filters. One is filtering by residual variance when using the general linear model without the predictor, the convolved stimulus trail. We set quantiles ranging from 0-quantile, which leaves the set unfiltered, to 0.95-quantile as cutoffs. For each θ-quantile, where $\theta \in [0, 0.95]$, voxels associated with the residual variances that are smaller than the specified quantile are removed from the set. Thus, for each cutoff we are left with a $\theta(100)\%$ fraction from the original set. The voxels in the remaining sets are then used to fit the full general linear model including the predictor corresponding to the stimulus. In the second stage we apply the FDR-controlling procedure, as described in Chapter 2. The other filter criterion we consider is the magnitude of $\widehat{\beta}_{2,v}$, the estimated regression coefficients

corresponding to the convolved stimulus, when the general linear model (2.2) is fit. Again, we set quantiles ranging from 0-quantile to 0.95-quantile as cutoffs and for each cutoff, we remove those voxels corresponding to the $\theta(100)\%$ lowest magnitudes of $\widehat{\beta}_{2,v}$. Then the FDR-controlling procedure is applied to the remaining sets of voxels.

The resulting binary activation maps are displayed in Figures 3.3 and 3.4. Comparing those Figures, we can see that both filters result in similar estimated activation patterns. However, the plots in Figure 3.4 show more red areas than those in Figure 3.3. We can conclude that when the filter criterion of residual variance is applied, less voxels are declared active than when the approach is used with filtering by the magnitude of $\widehat{\beta}_{2,v}$. We can further compare the two filters by looking at Figures 3.5 and 3.6. They show the number of rejections corresponding to the different quantile cutoffs. After a modest increase in the beginning, the curve in Figure 3.5 is decreasing. Filtering by residual variance the number of null hypotheses rejected increases by a small number if 1% of the original set is filtered out, but filtering out larger fractions results in less discoveries. The curve in Figure 3.6, however, has a different shape. It increases first and reaches its peak around the 0.7-quantile cutoff and then starts to decrease. Removing voxels corresponding to the lowest 69% of the estimated stimulus coefficients, the number of discoveries increases by roughly 44%. We can conclude that the filter by residual variance is not an effective choice with our approach. The increase in discoveries is minor, especially compared to the increase depicted in Figure 3.6. Using our two-stage approach with filtering on magnitude of $\widehat{\beta}_{2,v}$ results in more discoveries and we can conclude a gain in power. Additionally, FDR-control is guaranteed by using the FDR-controlling procedure under arbitrary dependence in the second stage.

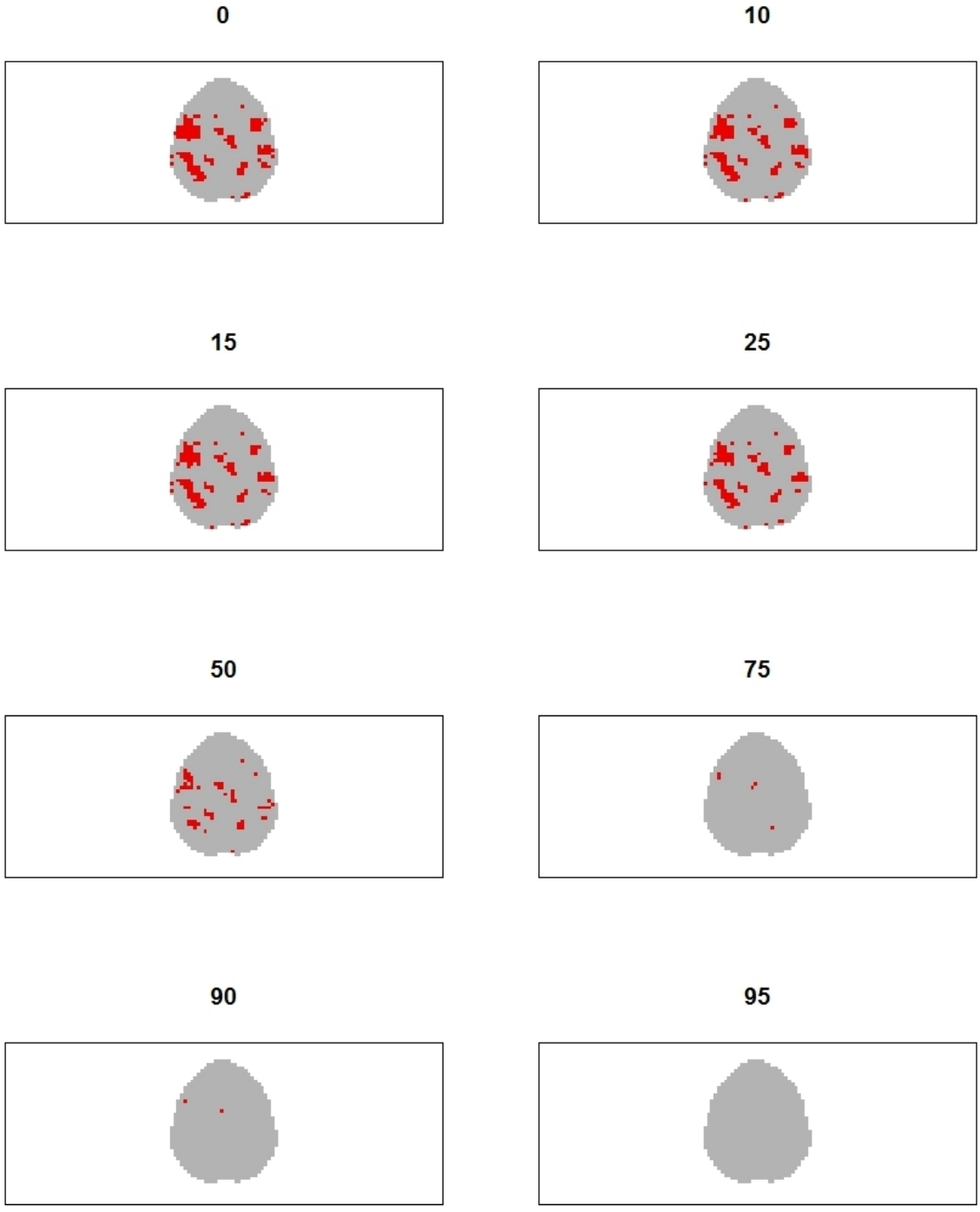

Figure 3.3: Binary activation maps display voxels that are declared active (red) when the two-stage approach with filtering by residual variance is applied. Voxels corresponding to the lowest $\theta(100)\%$ residual variances are removed from the set. The number above each plot indicates this percentage of voxels that are filtered out.

Figure 3.4: Binary activation maps display voxels that are declared active (red) when the suggested two-stage approach with filtering by magnitude of $\widehat{\beta}_{2,v}$, the estimated regression coefficients corresponding to the convolved stimulus, is applied. Voxels corresponding to the lowest $\theta(100)\%$ coefficient magnitudes are removed from the set. The number above each plot indicates this percentage of voxels that are filtered out.

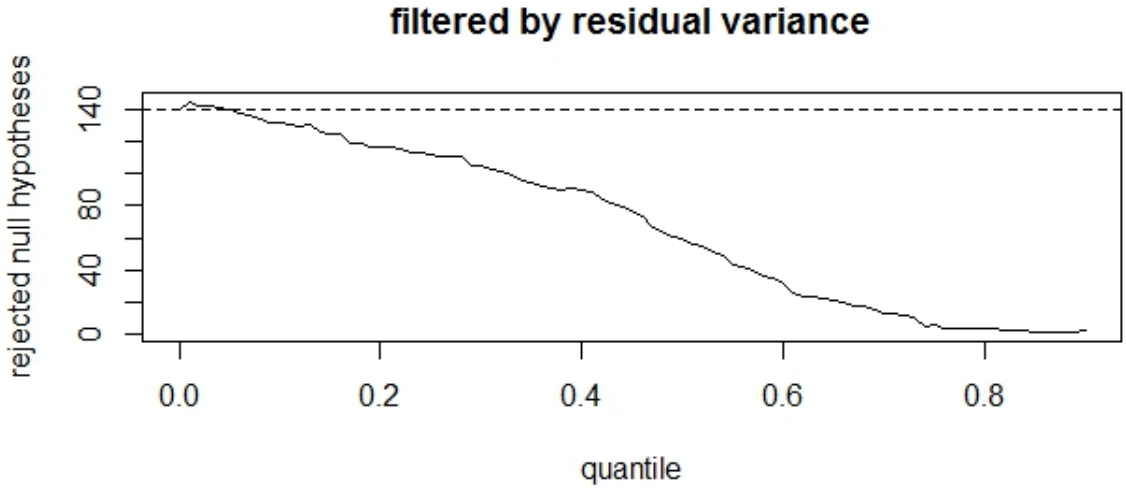

Figure 3.5: Number of rejections (voxels declared active) when the independent filtering method is applied with filtering by residual variance in the first step and an FDR-controlling procedure in the second step

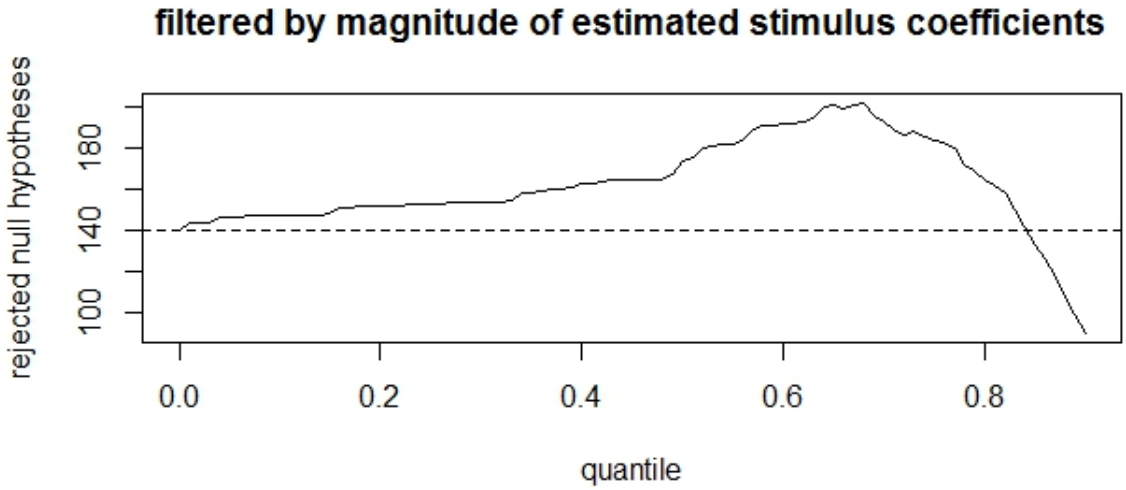

Figure 3.6: Number of rejections (voxels declared active) when the independent filtering method is applied with filtering by magnitude of estimated stimulus coefficients in the first step and an FDR-controlling procedure in the second step

Chapter 4

Discussion

We found that our suggested two-stage approach with filtering by magnitude of the estimated stimulus coefficients leads to more discoveries. By applying this approach, we were able to increase power while controlling the false discovery rate. However, filtering by residual variance resulted in less discoveries for the most part. Therefore the filter by coefficient magnitude should be used with our method. Although the approach leads to the conclusion that we can attain a gain in power, a simulation study should be conducted to investigate the method's impact on power in general.

We made assumptions when using the general linear model which are common but nevertheless unrealistic. One of them is that we assumed the same model for every voxel. Considering the physiological complexity of the brain, this assumption is doubtful. Furthermore it is well known that fMRI data are not only temporally but also spatially correlated. The voxelwise general linear model we used accounts only for temporal correlation. Instead of describing spatial correlation with the model, we dealt with it by choosing an appropriate threshold within the FDR-controlling procedure that we apply. In this way our approach accounts for both temporal and

spatial correlation.

In our study we filtered by two different filter criteria. While filtering by magnitude of the estimated stimulus coefficients led to satisfying results, filtering by residual variance did not. The only multiple testing adjustment we applied in stage two is the FDR-controlling procedure that is suggested by (Genovese et al., 2002). The results summarized in Chapter 3 were obtained with the FDR bound $q = 0.05$, which can be interpreted as the maximum proportion of false discoveries we are willing to tolerate. Other choices of q could be considered as well. Although it is common to select this bound similar to typical significance levels, there is no good reason to do so. There are situations in which more conservative or higher choices are appropriate, so it might be worth considering different FDR bounds (Lazar, 2008). The application of independent filtering with other filter and multiple testing approaches should be analyzed. Moreover it should be considered that filtering drastically alters spatial correlation structure among data. So far the standard way to handle this is to use a multiple testing procedure that makes no assumption about the dependence structure. Interesting future work would be to find filtering methods that preserve local dependences among the statistics, or a satisfactory way of modeling the new correlation structure obtained after filtering.

We presented the two-stage approach in the context of fMRI data. In general, the procedure could be applied to any high-dimensional data, where the multiple testing problem often arises. Knowledge about the data's nature and scientific background can help to find an effective filter. Studies analyzing independent filtering in neuroimaging have been rare so far. However, due to its potential to gain power while controlling the false discovery rate, its simple implementation and the chance to incorporate scientific background information, further analysis of independent filtering methods could lead to a valuable contribution to future fMRI research.

Bibliography

Benjamini, Y. and Hochberg, Y. (1995), "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society, Series B (Methodological)*, 289–300.

Benjamini, Y. and Yekutieli, D. (2001), "The control of the false discovery rate in multiple testing under dependency," *Annals of Statistics*, 1165–1188.

Bonferroni, C. E. (1935), *Il calcolo delle assicurazioni su gruppi di teste*, Rome.

— (1936), "Teoria statistica delle classi e calcolo delle probabilità," *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8, 3–62.

Bourgon, R., Gentleman, R., and Huber, W. (2010), "Independent filtering increases detection power for high-throughput experiments," *Proceedings of the National Academy of Sciences*, 107, 9546–9551.

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008), "The brain's default network," *Annals of the New York Academy of Sciences*, 1124, 1–38.

Camchong, J., Dyckman, K. A., Austin, B. P., Clementz, B. A., and McDowell, J. E. (2008), "Common neural circuitry supporting volitional saccades and its disruption in schizophrenia patients and relatives," *Biological Psychiatry*, 64, 1042–1050.

Dale, A. M. and Buckner, R. L. (1997), "Selective averaging of rapidly presented individual trials using fMRI," *Human Brain Mapping*, 5, 329–340.

Dunn, O. J. (1959), "Estimation of the medians for dependent variables," *The Annals of Mathematical Statistics*, 192–197.

— (1961), "Multiple comparisons among means," *Journal of the American Statistical Association*, 56, 52–64.

Dyckman, K. A., Camchong, J., Clementz, B. A., and McDowell, J. E. (2007), "An effect of context on saccade-related behavior and brain activity," *NeuroImage*, 36, 774–784.

Fisher, R. (1935), *The Design of Experiments*, Edinburgh: Oliver and Boyd.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J., Frith, C. D., and Frackowiak, R. S. (1994), "Statistical parametric maps in functional imaging: a general linear approach," *Human Brain Mapping*, 2, 189–210.

Friston, K. J., Jezzard, P., and Turner, R. (1994), "Analysis of functional MRI time-series," *Human Brain Mapping*, 1, 153–171.

Friston, K. J., Josephs, O., Rees, G., and Turner, R. (1998), "Nonlinear event-related responses in fMRI," *Magnetic Resonance in Medicine*, 39, 41–52.

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002), "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," *NeuroImage*, 15, 870–878.

Glover, G. H. (1999), "Deconvolution of Impulse Response in Event-Related BOLD fMRI," *NeuroImage*, 9, 416–429.

Huettel, S. A., Song, A. W., and McCarthy, G. (2009), *Functional Magnetic Resonance Imaging*, Sunderland: Sinauer Associates, 2nd ed.

Lange, N. and Zeger, S. L. (1997), "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)*, 46, 1–29.

Lazar, N. A. (2008), *The Statistical Analysis of Functional MRI Data*, New York: Springer.

Lindquist, M. A. et al. (2008), "The statistical analysis of fMRI data," *Statistical Science*, 23, 439–464.

Marchini, J. L. and Ripley, B. D. (2000), "A new statistical approach to detecting significant activation in functional MRI," *NeuroImage*, 12, 366–380.

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., and Ugurbil, K. (1992), "Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging," *Proceedings of the National Academy of Sciences*, 89, 5951–5955.

Poldrack, R. A., Mumford, J., and Nichols, T. (2011), *Handbook of Functional MRI Data Analysis*, Cambridge: Cambridge University Press.

Purdon, P. L., Solo, V., Weisskoff, R. M., and Brown, E. N. (2001), "Locally regularized spatiotemporal modeling and model comparison for functional MRI," *NeuroImage*, 14, 912–923.

Purdon, P. L. and Weisskoff, R. M. (1998), "Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI," *Human Brain Mapping*, 6, 239–249.

Scheffe, H. (1953), "A method for judging all contrasts in the analysis of variance," *Biometrika*, 40, 87–110.

Shumway, R. H. and Stoffer, D. S. (2011), *Time Series Analysis and its Applications: With R Examples*, Springer Texts in Statistics, New York: Springer, 3rd ed.

Smith, S. M., Matthews, P. M., and Jezzard, P. (2001), *Functional MRI: an Introduction to Methods*, Oxford: Oxford University Press.

Tukey, J. W. (1949), "Comparing individual means in the analysis of variance," *Biometrics*, 99–114.

Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M. (2001), "Temporal autocorrelation in univariate linear modeling of FMRI data," *NeuroImage*, 14, 1370–1386.