Advances in Modeling and Inference of Neuroimaging Data

by

Hui Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2008

Doctoral Committee:

Research Associate Professor Timothy D. Johnson, Co-Chair
Adjunct Research Associate Professor Thomas E. Nichols, Co-Chair
Professor Jeffrey A. Fessler
Associate Professor Bin Nan

To my family

Acknowledgements

First of all, I would like to express my deepest appreciation to my co-advisors, Drs. Thomas E. Nichols and Timothy D. Johnson for those wonderful research topics, guidance, discussions and support on my Ph.D study. It is really hard to imagine how much I learned from them about neuroimaging analysis, computing skills, writing skills and many other aspects. Without their help, the dissertation would not be successfully finished. I am immensely grateful to Dr. Fessler and Dr. Nan for their insightful comments, ideas to my research.

I am also thankful to all the knowledgeable faculty and smart students in the Department of Biostatistics for making this department an wonderful environment to study. I would also like to thank the staffs, especially Tara, for willing to assist me with many of my needs and for their caring.

At last but not the least, I would like to express my gratitude to my family. Thanks to my parents for all they have done for me, for their strong support and endless love in my life. They always care things in my life and encourage me on my study, even though they are in China. Thanks to my younger sister for helping me take care of my baby during her vacation. Thanks to my husband for tolerating and supporting me over last several years as I have worked towards achieving this goal. Special thanks to my baby daughter that she has to wait for her mother till night. I would also like to thank all my friends for their help.

Table of Contents

List of Figures

List of Tables

List of Appendices

<u>Appendix</u>

Chapter 1

Introduction

Functional Magnetic Resonance Imaging (fMRI) is a relatively newly developed technique used to study neural activity as reflected in changes in blood oxygenation and flow. The results of fMRI data analysis are activation maps which show the specific brain regions related to a mental task. While fMRI has been in use for over 15 years, there remains a need for more sensitive modeling and inference methods. This dissertation consists of three studies to improve sensitivity by developing new theoretical results optimizing existing techniques. In this chapter, we review fMRI techniques and existing fMRI statistical methods.

1.1 Background

fMRI is a technique for studying the relationship between brain activities and human behavior. The blood-oxygen-level-dependent (BOLD) contrast method is the most important method of several methods for obtaining functional information via magnetic resonance imaging (MRI) [Ogawa & Lee, 1992]. It is well known that blood flow and blood oxygen will be changed due to brain activities [Roy & Sherrington, 1890]. When a portion of the human brain has increased neuronal firing activities, the increased demand for oxygen stimulates blood flow. Normally, the blood flow response to the change in blood oxygen level occurs within approximately 1-5 seconds. After that, the

haemodynamic response reaches its peak around 4-5 seconds, followed by a return to its baseline. During increased brain activity, the ratio of oxygenated (HbO_2) to deoxygenated hemoglobin (Hb) is larger in the region activated than at rest [Ogawa *et al.*, 1990; Kwong *et al.*, 1992]. The magnetic susceptibility varies during this process, as deoxygenated Hb is paramagnetic, oxygenated Hb is diamagnetic. In a MRI scanner, this variation alters the homogeneity of the local magnetic field, and results in small changes in image intensity. Figure 1.1 [Springer *et al.*, 1999] simply illustrates this procedure. Part (a) shows the mechanism of the BOLD structure, while Part (b) illustrates the oxygen changes from rest state to active state.

During a fMRI experiment, each subject is presented with two or more conditions, for example a task and a rest condition. The scanner scans the subject's brain in regular intervals during the entire experiment: normally 1-4 seconds. The resulting image is represented in voxels in 3 dimensional space. The size of a voxel is generally 2-by-2mm in plane, an 2-4mm thick. A 3D brain image consists of about 100,000 voxels. Each subject typically has 100 scans or more during an experiment. In addition, the images have to be realigned and motion corrected before analysis. To allow for inter-subject, or second level analysis, the brain images are warped to match a standard brain atlas. This is a crucial step. Next, images are smoothed with a Gaussian kernel in order to increase the signal to noise ratio (SNR) and reduce residual anatomic differences. It is important to note that BOLD effect is not an absolute measure, and is only meaningful as a difference between different conditions.

1.1.1 Experimental Design

A carefully designed experiment is required in order to detect the subtle changes between conditions. During a fMRI experiment, each subject is provided a specific task sequence, such as frequent finger tapping, or random finger tapping. A fMRI experiment
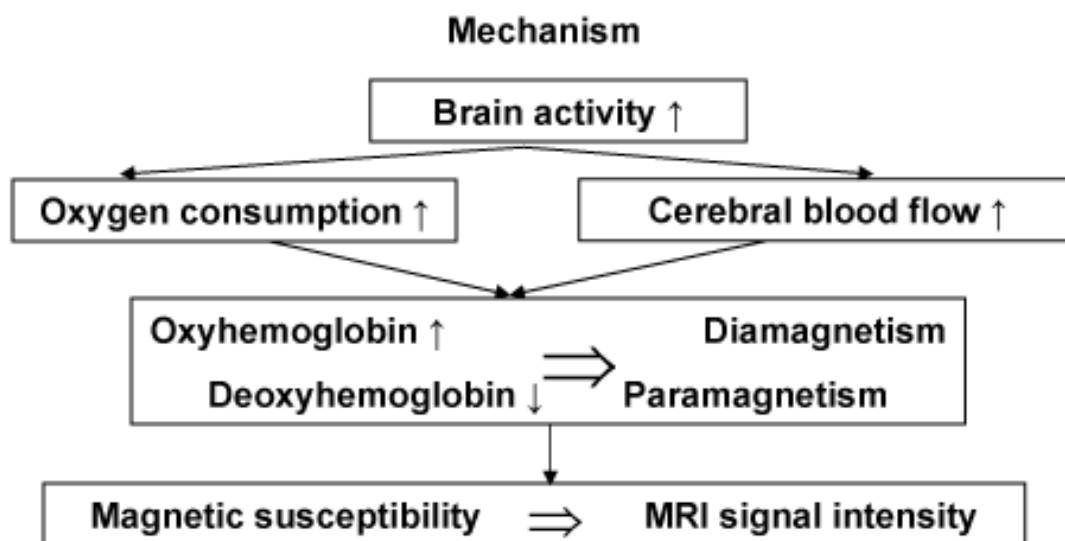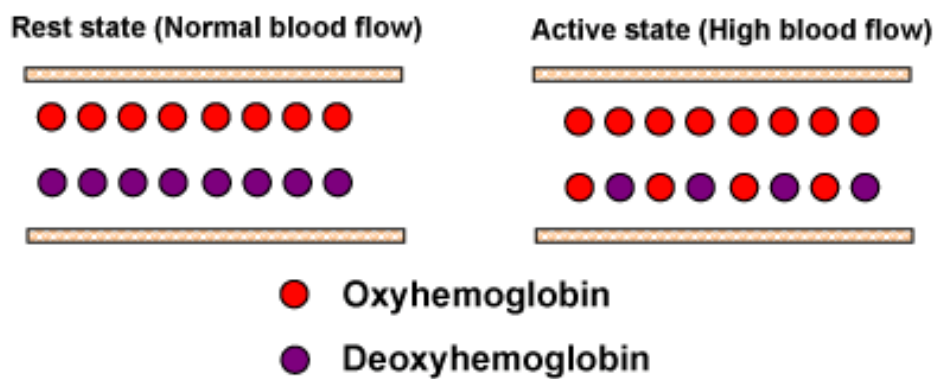
Figure 1.1: Illustration BOLD mechanism structure

is designed according to scientists' research goals. A fMRI experiment normally lasts around one hour. Obviously, the fMRI sequence forms a series of measurements for each voxel over time. The time series measurements relate brain response to the specific task sequence at each voxel.

There are three main types of designs related to fMRI data analysis; block design, event-related design and mixed design. Mixed design which is a combination of block design and event-related design. The most commonly used design is block design. Figure 1.2 (top) shows a simple block design. There are several discrete binary periods in this design. The binary external stimulus is 0 if there is no tasks (rest, or baseline condition) during the experiment or 1 if there is a task. The control task is important since it provides the baseline against which the cognitive tasks are compared. The design consists of the same on-off pattern, although the duration of blocks may vary. A block design normally allows the hemodynamic response function (HRF) reach its maximal value. In addition, the long rest inter-stimulus intervals (ISI) of the block design allows the HRF to return to baseline. The HRF is assumed to have a smooth shape, and a scaled gamma probability distribution function is often used [Lange & Zeger, 1997]. In SPM[1], a mixture of Gamma functions are used in order to mimic the undershoot of HRF [Friston *et al.*, 1998].

The fMRI response is the external stimulus convolved with a HRF (Figure 1.2 (bottom)). Figure 1.2 (middle) illustrates a HRF.

The stimuli sequence in event-related design is arbitrary. Figure 1.3 (top) shows a simple event-related design. Figure 1.3 (bottom) illustrates the response of the corresponding design. This randomization property gives the event-related paradigm greater flexibility than the block design. Moreover, this design is much closer to human being's

---

[1]http://www.fil.ion.ucl.ac.uk/spm

behavior. However, it has less statistical powerful than the block design because of weak contrast-to-noise ratio [Chee *et al.*, 2003; Soltysik & J.S., 2006; Friston *et al.*, 1999].

The mixed design is the third design used in practice. In this design, events are gathered into blocks. This combines the sensitivity of block design with the flexibility of event-related design. The advantage of the mixed design is that it reduces confounding due to the stimulus order. For example, if we have three conditions, the sequence of the stimuli are different but ISI within blocks are the same.
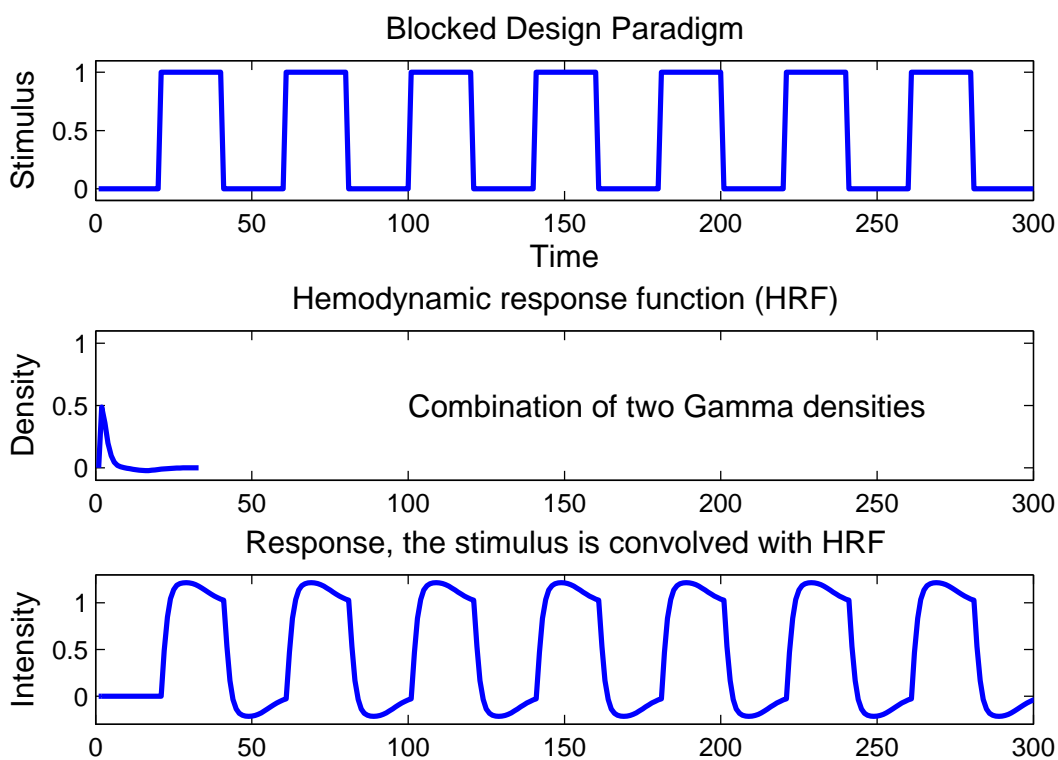
Figure 1.2: Illustration of stimulus in block design

1.1.2 Temporal Autocorrelation

A subject's brain is imaged repeatedly at very short intervals, for approximately 2 seconds. At each voxel, the data thus comprise a time series (see Figure 1.4 for an example).
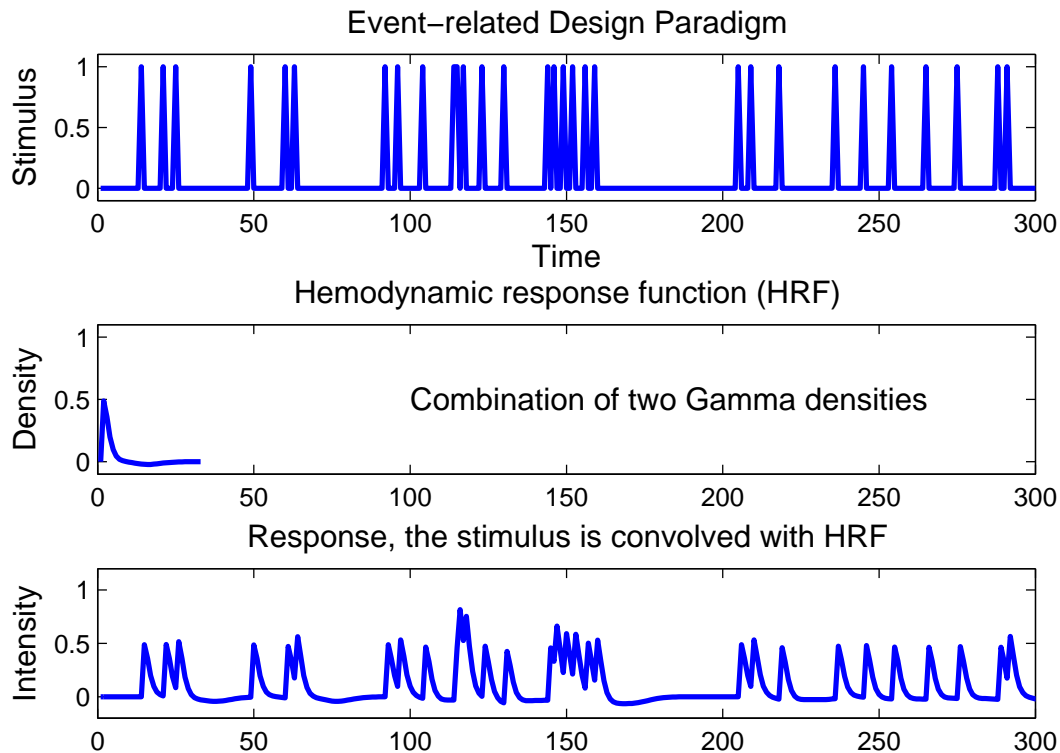
Figure 1.3: Illustration of stimulus in event-related design

Many authors have investigated the temporal autocorrelation in fMRI time series [Bullmore *et al.*, 1996; Marchini & Ripley, 2000; Woolrich *et al.*, 2001]. The general method is to use autocorrelation with one parameter (AR(1)) model to describe the underlying correlation structure [Worsley *et al.*, 2002]. Although this model can be extended to an AR(p) model, the complex autocorrelation structure increases computational requirements. The other issue related to temporal autocorrelation is whether to assume that the temporal correlation for all voxels is same. Since there are more than 100,000 voxels in an image and we have to analyze all voxels simultaneously, the computational complexity is huge. Thus some software packages only estimate a pooled AR(1) model (e.g. SPM), while FSL[2] estimates a local autocorrelation function (ACF). In reality, we use one model for all voxels.

---

[2]http://www.fmrib.ox.ac.uk/fsl
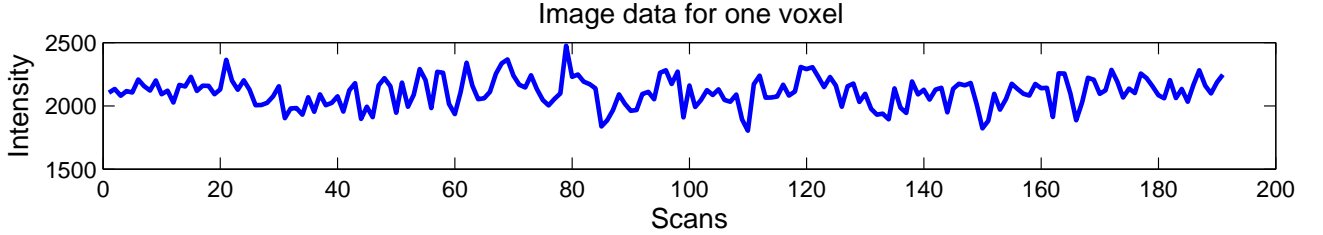
Figure 1.4: Image data for one voxel

1.2 Statistical Analysis

1.2.1 Methods

The main goal for fMRI data analysis is to investigate the tasks related activated brain regions. A general linear model (GLM) is applied to each voxel for fMRI data analysis. In a mass univariate data analysis,

$$Y_i = X\beta_i + \varepsilon_i$$

is fitted for each voxel $i = 1, \ldots, I$, where Y_i is a $N \times 1$ vector of responses, X is a common $N \times q$ design matrix of predictors, β_i is a $q \times 1$ vector of unknown parameters and ε_i is a $N \times 1$ vector of random errors. Typically, at each voxel, errors are assumed to be independent and identically distributed $N(0, \sigma_i^2)$ random variates, though dependent errors can be accommodated [Luo & Nichols, 2003]. The ordinary least squares estimator of β_i is $\widehat{\beta}_i = (X^\mathrm{T}X)^{-1}X^\mathrm{T}Y_i$, and of σ_i^2 is $\widehat{\sigma}_i^2 = e_i^\mathrm{T}e_i/\eta$, where $e_i = Y_i - X\widehat{\beta}_i$ and where η is the error degrees of freedom. Then, the Student's t statistic at voxel i is

$$T_i = c\widehat{\beta}_i \left(c(X^\mathrm{T}X)^{-1}c^\mathrm{T}\widehat{\sigma}_i^2 \right)^{-1/2}$$

where c is a contrast of interest (row vector). We write the statistic image as $\mathbf{T} = \{T_i\}_{i=1}^{I}$.

Performing inference for each voxel independently is known as voxel-wise inference. If the t statistic of one voxel is greater than a cutoff point, that voxel is considered statistically significant. An alternative approach is to apply an arbitrary threshold and

form clusters, and then assess significance based on the spatial extent of the cluster. Such cluster-wise inference is generally more sensitive than voxel-wise inference, especially for spatially extended signals.

Whether voxel-wise or cluster-wise inference is used, there is a serious multiple testing issue when searching the brain for activation. If we use Bonferroni's method, there is little power, as Bonferroni is conservative under dependence. Random Field Theory (RFT) and permutation methods provide inferences that adapt to the smoothness of the data.

1.2.2 Nonparametric Permutation Method

The classical statistical inference methods are normally based on some assumptions on the distribution of the underlying population. Once the distribution of a statistic is derived, the corresponding parameters, such as the mean and variance, will be fully stated parametrically. All conclusions based on these inference methods on these statistics are valid as long as these assumptions are substantiated. However, in reality, the distribution of the underlying population is unknown and not easily derived. In addition, the distributional assumptions are sometimes not reasonable. Nonparametric methods were developed to address these shortcomings of classical parametric methods.

A permutation test is one type of nonparametric inference. It is based on a resampling of the data without replacement, in contrast to bootstrap methods which sample data with replacement. All possible values of rearrangements of the data, which are equivalent under the null hypothesis, are calculated in order to find the distribution of the corresponding test statistic. The basic requirement for this method is that the data are exchangeable under the null hypothesis. The advantage of this method is that it does not assume anything about the distribution of the test statistic. It is widely used for newly defined or complicated statistics. The limitation for this method is that

data are sometimes not exchangeable, as in intra-subject fMRI data, where temporal autocorrelation invalidates exchangeability. In this dissertation, we use a permutation method only for group-level fMRI data and not for voxels, since it will change the spatial structure if we permut voxels.

1.3 Existing inference methods and problems

Poline *et al.* [Poline *et al.*, 1997] developed the parametric joint distribution of cluster extent and suprathreshold peak height intensity. Their minimum P value approach is sensitive to clusters with large cluster extent or high suprathreshold peak height intensity. In addition, Bullmore *et al* [Bullmore *et al.*, 1999] proposed a nonparametric cluster mass inference method. Cluster mass is defined as the integral of suprathreshold intensities within a cluster. They used the permutation framework to obtain cluster mass P values for all defined clusters. This method is generally considered as a more sensitive method than cluster extent or voxel intensity inference method since the cluster mass statistic naturally combines the information from cluster extent and voxel intensity. Other combining functions were also developed to draw benefits from the two statistics [Hayasaka & Nichols, 2004]. To date, there is no study that combines methods with these two statistics.

Although the nonparametric cluster mass inference method is considered a more sensitive method than the other combining methods, it can not be used for a single subject fMRI data analysis. The data for a single subject are not exchangeable because a parametric autocorrelation model is needed. In addition, nuisance covariates are not easy to accommodate since the permutation scheme depends on experimental design and not on the design matrix. Thus, a parametric cluster mass inference method is needed. Moreover, random field theory provides a platform for the development of the

parametric cluster mass inference method.

For a small group with less than 20 subjects, the critical value is higher than that for a large group because of the small degrees of freedom. This fact will lead to fewer significant region. Previous studies found that the smoothed variance t-test has a larger degrees of freedom and better detection sensitivity than the standard t-test [Nichols & Holmes, 2002; Hayasaka & Nichols, 2003]. However, the magnitude of the smoothness parameter applied to variance images in order to achieve this is unknown. In addition, the true variance image was considered a constant [Worsley *et al.*, 2002], which is a questionable assumption.

In this dissertation, we study the problems mentioned above. In Chapter 2, the performance of the combining method with suprathreshold average intensity and cluster extent is evaluated by simulation and real data analysis. In Chapter 3, the parametric distribution of the cluster mass statistic is fully developed via random field theory. In Chapter 4, the relationship between three smoothness parameters, smoothing kernel used for data, smoothing kernel from a true variance image and smoothing kernel size applied to the variance image, and the effective degrees of freedom of the smoothed variance t-test are studied.

Chapter 2

Combining Suprathreshold Average Voxel Intensity and Cluster Extent with Permutation Test Framework

Neuroimaging inferences are generally based on two statistics, cluster extent, the number of voxels within a cluster, and voxel intensity, the maximum voxel intensity in a cluster. Voxel wise inference method focuses on signal intensity, while cluster wise inference method focuses on signal spatial extent. To leverage the strength from both statistics, some combining functions were developed. Poline *et al.* [Poline *et al.*, 1997] developed parametric joint distribution of cluster extent and suprathreshold peak height intensity. Their minimum P value approach is sensitive to clusters with large cluster extent or high suprathreshold peak height intensity. In addition, Bullmore *et al* [Bullmore *et al.*, 1999] proposed nonparametric cluster mass inference method. Cluster mass is defined as the integral of suprathreshold intensities within a cluster. They used permutation framework to obtain cluster mass P values for all defined clusters. This method is generally considered as a more sensitive method than cluster extent or voxel intensity inference method since the cluster mass statistic naturally combines the information from cluster extent and voxel intensity. Other combining functions were also developed to benefit from the two statistics [Hayasaka & Nichols, 2004]. Since cluster mass is the product of cluster extent and suprathreshold average intensity within a cluster, we

study some combining functions using the two statistics with permutation framework in this work. We compare the proposed combining methods to nonparametric cluster mass method and partial inference methods. We also evaluate them with simulation study and apply them in real data analysis.

2.1 Introduction

Neuroimaging inferences are generally based on the extent of a cluster or the maximum voxel intensities within a cluster. Cluster extent inference method is sensitive for spatially extended signals [Friston *et al.*, 1996; Poline *et al.*, 1997], while methods based on intensity are sensitive when signal magnitudes are large [Friston *et al.*, 1991; Worsley *et al.*, 1992]. However, the two inference methods do not have much power for moderate extent and intensity [Hayasaka & Nichols, 2004].

Poline *et al* [Poline *et al.*, 1997] developed parametric combining method based on Gaussian Random Field Theory (RFT). They derived a joint distribution of cluster extent and suprathreshold peak height intensity. In their approach, the minimum P value from cluster extent and suprathreshold peak height intensity [Adler, 1981; Worsley *et al.*, 1992; Friston *et al.*, 1994] is used to find a critical region of a brain. Their approach is sensitive to those clusters with either large signal extent or high suprathreshold peak height intensity. In addition, because this method is based on RFT, there are some assumptions with the method. They assumed all images are smoothed and can be approximated with a stationary Gaussian field. They also assumed that the autocorrelation function between voxels can be measured with a Gaussian correlation function. In addition, they required that the threshold used to define clusters is relatively high. Furthermore, this parametric combining method is applicable to Gaussian field. A *t*-to-*Z* transformation has to be applied if the image is a *t* image, although the degrees of

freedom of this t image may be small.

Bullmore *et al.* [Bullmore *et al.*, 1999] proposed a nonparametric cluster mass inference method without those strict assumptions. Cluster mass is defined as the integral of suprathreshold intensities within a cluster [Holmes, 1994]. Their method is based on permutation framework. They recorded cluster mass statistic for each cluster in each realization to produce the distribution of cluster mass and find the corrected P value for a specific cluster. They also recorded the maximum cluster mass to generate the distribution of the maximum cluster mass and find the corrected P value for a specific cluster. The uncorrected and corrected P values are P values according to whether the multiple comparison issue is considered or not. The difference between them is explained in Appendix A.1

Hayasaka and Nichols developed some combining functions from cluster extent and suprathreshold peak height intensity [Hayasaka & Nichols, 2004]. They used P values from the above two statistics in combining functions, Tippet and Fisher functions. Their Tippet combining function used minimum P value approach, which is similar to Poline *et al.*'s method. They also developed weighted Tippet and Fisher functions according to signals of interest, either localized high intensity but small cluster extent or low intensity but large cluster extent.

In our previous work we found that the cluster mass was generally the most sensitive inference method [Hayasaka & Nichols, 2004]. As the cluster mass is the product of the cluster extent and the suprathreshold average intensity, cluster mass can be seen as a method that combines cluster extent and suprathreshold average intensity with the "product" combining function. We propose Tippet and Fisher combining functions using cluster extent and suprathreshold average intensity. The goal of this work is to show if combining P values from cluster extent P values and suprathreshold average intensity

P values are more sensitive than other combining statistics. Since we do not know the distribution of two combining functions, this work is also based on nonparametric permutation framework. We evaluate the two combining functions with simulation study. We also apply them to a second level data analysis because we can permute data labels.

2.2 Methods

2.2.1 Statistic Image

In a mass univariate data analysis, a general linear regression model (GLM)

$$(2.1) \qquad\qquad\qquad Y_i = X\beta_i + \varepsilon_i$$

is fit for each voxel $i = 1, \ldots, I$, where Y_i is a $N \times 1$ vector of responses, X is a common $N \times q$ design matrix of predictors, β_i is a $q \times 1$ vector of unknown parameters and ε_i is a $N \times 1$ vector of random errors. Typically, at each voxel, errors are assumed to be independent and identically distributed $N(0, \sigma_i^2)$ random variates, though dependent errors can be accommodated [Luo & Nichols, 2003]. The ordinary least squares estimator of β_i is,

$$\widehat{\beta_i} = (X^\mathrm{T} X)^{-1} X^\mathrm{T} Y_i$$

and σ_i^2 is

$$\widehat{\sigma_i^2} = e_i^\mathrm{T} e_i / \eta$$

where $e_i = Y_i - X\widehat{\beta_i}$ and where η is the error degrees of freedom. Then the Student's t-statistic at voxel i is

$$(2.2) \qquad\qquad\qquad T_i = \frac{c\widehat{\beta_i}}{\sqrt{c(X^\mathrm{T} X)^{-1} c^\mathrm{T} \widehat{\sigma_i^2}}}$$

where c is a contrast of interest (row vector). We write the t-statistic image as $\mathbf{T} = \{T_i\}_{i=1}^I$.

Given cluster-forming threshold $u_c > 0$, the set of suprathreshold statistics $\{T_i : T_i > u_c\}_{i=1}^{I}$ is used to define clusters. Contiguous clusters are defined by a neighborhood scheme, typically 18 connectivity scheme on a three dimensional image, i.e., in a $3\times3\times3$ cube, all 18 voxels are connected to the center except 8 voxels at corner.

2.2.2 Partial inference methods

After a t image is generated and clusters are defined, voxel intensity inference method will be used for each voxel of the t image, as well as the cluster extent inference method will be used for all defined clusters. We regard these two methods as partial inference methods since we use them separately without combining one result with the other. With voxel intensity inference method, if the degrees of freedom of a t image are known, then the P value for each voxel is calculated according to t distribution property. The cluster extent based method is built on RFT [Adler, 1981; Poline *et al.*, 1997]. The P value for each cluster is generated according to the distribution of cluster extent.

To perform cluster wise inference method, there are two types of P values. One is the P value from the peak height intensity within a cluster. The distribution of peak height intensity above a threshold is exponential [Adler, 1981; Friston *et al.*, 1994]. The other one is the P value from the cluster extent of that cluster. The distribution of cluster extent is also exponential [Friston *et al.*, 1994; Poline *et al.*, 1997]. Therefore, we have two types of P values for each cluster with two partial inference methods, a P value for the peak height intensity within a cluster and a P value for the cluster extent of this cluster. With permutation framework, those P values are all nonparametric P values.

With nonparametric permutation framework, P values for peak height intensity and cluster extent will be acquired using separate permutation tests. The only assumption for the permutation test is exchangeability [Nichols & Holmes, 2002]. Under the null hypothesis, data labels are randomly assigned without changing the distribution of the

test statistics. Normally, this method needs long compute times to generate an empirical distribution. The P value of a specific voxel (cluster) is measured as comparing the test statistic with the empirical distribution of that statistic. For example, the P value of peak height intensity of a cluster is the proportion of number of peak height intensity in its empirical distribution that is greater than or equal to the observed one. To perform cluster wise inference method, we record the peak height intensity and cluster extent for each cluster in every realization to calculate uncorrected P values for an observed cluster, respectively. We also record the maximum of peak height intensity and the maximum of cluster extent in each realization to calculate corrected P values for an observed cluster, respectively.

2.2.3 Multiple comparison issue

The number of voxels in a brain image is huge. Generally, there are over 100,000 voxels in a brain image. We have a null hypothesis for each voxel. The crucial point is that we have to test all hypotheses simultaneously. The use of largest value for a test statistic in each realization is a way to solve the multiple comparison problem among clusters. This method controls a family wise error (FWE) [Nichols & Hayasaka, 2003]. The corrected P value of a cluster with voxel intensity inference method is generated by comparing the peak height intensity of a cluster with the empirical distribution of the maximum T distribution, T_{max}. Similarly, the corrected P value of a cluster with cluster extent inference method is obtained by comparing the cluster extent of this cluster with the empirical distribution of maximum cluster extent distribution, S_{max}. The FWE correction method with permutation framework is explained in Nichols & Holmes in detail [Nichols & Holmes, 2002].

2.2.4 Combining functions

When combining P values we have a choice of combining either corrected or uncorrected P values. Let uP^t and uP^c be the uncorrected P values for suprathreshold average intensity (AvgT) and cluster extent, respectively; let cP^t and cP^c be the corresponding FWE corrected P values. For each cluster we compute Tippet and Fisher combining functions, or statistics, based on the uncorrected P values, respectively.

$$(2.3) \qquad W^T \quad = \quad 1 - \min(\log_{10} uP^t, \log_{10} uP^c)$$

$$(2.4) \qquad W^F \quad = \quad -2(\log_{10} uP^t + \log_{10} uP^c)$$

as well as the analogous calculations for corrected P values. The Tippet combining function is equivalent to picking the better of the two P values. Cluster mass W^M is computed as the sum of T values within a cluster above the cluster-defining threshold. Corrected P values are found with permutation by building the maximum distribution [Nichols & Holmes, 2002]; uncorrected P values are found by building the distribution of all clusters found in all permutations.

The combined P values (W^F & W^T) represent an arbitrary statistic, upon which either corrected or uncorrected inferences can be made. There are four types of results for each combining function. For example, corrected P values from Tippet combining function using uP^t, uP^c or cP^t, cP^c. We can not use uP^t with cP^c or cP^t with uP^c. If we want corrected P values from Tippet function but uP^t and uP^c are used, we record the maximum Tippet value in each realization to generate the empirical distribution of the maximum Tippet statistic. Similarly, uncorrected P values from the Tippet combining function can be also obtained with uP^t, uP^c or cP^t, cP^c. For example, if we want uncorrected P values from Tippet function but cP^t and cP^c are used, we record Tippet values for all clusters from all realization to generate the empirical distribution of the

Tippet statistic. Considering also the partial (non-combined) tests, we have five cluster test statistics to compare: AvgT, cluster extent, W^T, W^F and W^M.

2.2.5 Simulation

We use 2,000 realizations of a 15-subject dataset with Gaussian noise images in 3 dimensional space ($48 \times 48 \times 32$ voxels) and add a sphere-shaped signal with uniform intensity to each dataset, varying signal diameter and intensity. A cluster defining threshold is P=0.01 with 14 degrees of freedom of t distribution and nominal $\alpha = 0.05$ level is used and for each realization. A permutation test (1,000 perms) is performed. The rejection rate (power) of each type of test is also recorded, we only show results for corrected inferences based on (cP^t, cP^c). The results for uncorrected inferences based on (uP^t, uP^c) are similar to the results with (cP^t, cP^c).

2.2.6 Application

We apply the proposed combining methods on a second level fMRI working memory dataset, which has 12 subjects [Marshuetz *et al.*, 2000]. The analysis is based on contrast images for item recognition versus control. All of the statistics described above are computed, in addition to suprathreshold peak height intensity, i.e. maximum voxel intensity within a cluster (MaxT) and Tippet and Fisher combining functions with MaxT (to compare with previous work on MaxT). The total number of permutation we used is 4096 (2^{12}) and the cluster defining threshold is P=0.001 with 11 degrees of freedom of t distribution.

2.3 Results

2.3.1 Simulation

The rejection rates of the partial, and combined tests from the simulations are shown in Figure 2.1. It shows that when the diameter is small, for example, 6, the AvgT, Tippet and Fisher with AvgT statistic inference methods have the best rejection rate, and cluster extent inference method is very insensitive. For larger diameters cluster extent inference method is the most sensitive, but the combining methods are all nearly as good. In general, for small diameter and moderate intensity (0.5-.5), the AvgT, Tippet and Fisher using AvgT statistics have the best performance.
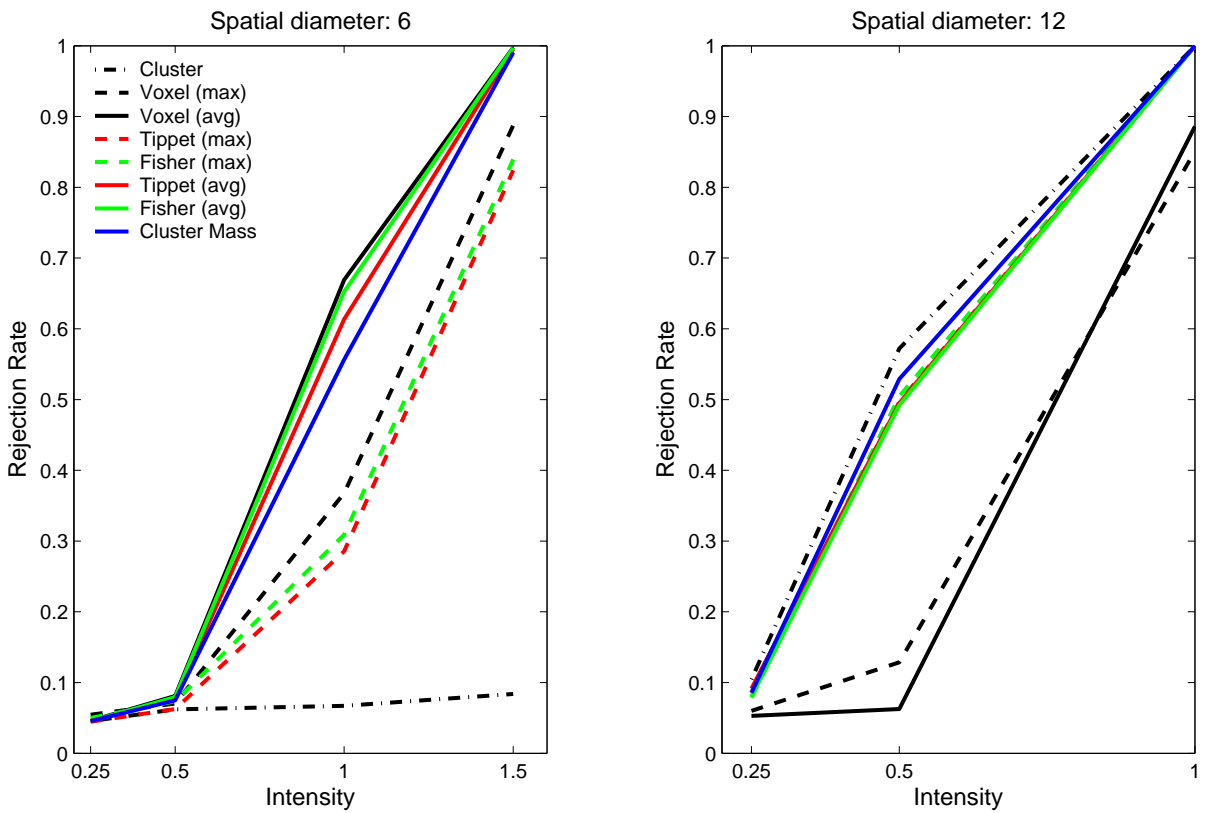
Figure 2.1: Rejection rates versus signal intensity

2.3.2 Real Data

The combined inferences based on (cP^t, cP^c) were consistently more sensitive than those based on (uP^t, uP^c). The results for the 5 largest clusters are shown in Figure 2.2, Table 2.1 and 2.2. Figure 2.3 shows the rejection regions implied by the different statistics considered. In lower right corner plot, if we use partial cluster extent inference method, if the cluster extent of a observed cluster is greater than 50, that cluster will be considered as a significant one. It is similar to use partial suprathreshold average intensity method, all clusters are regarded as significant if their AvgTs are above the horizontal line. If we apply cluster mass inference method, clusters located on the right side of blue dash curve are regarded as significant. For example, in upper left corner plot, i.e. the correct combining method using corrected P values plot, the 3$^{\text{rd}}$ and 5$^{\text{th}}$ cluster are significant with all method. The the 1$^{\text{st}}$, 2$^{\text{nd}}$ and 4$^{\text{th}}$ clusters are significant if we use cluster extent, cluster mass, Tippet and Fisher inference method but suprathreshold average voxel intensity inference method. Cluster 6-10 are not significant with any inference methods. The figure also shows that the 1$^{\text{st}}$ is a flat cluster with large cluster extent and small suprathreshold average intensity. Table 2.1 shows results using AvgT statistic and Table 2.2 shows results with MaxT statistic.

The statistic that has uniformly the worst results is the partial AvgT inference method based on real data analysis. However, the Fisher combining method with AvgT and cluster extent is often the most sensitive and never far from optimal. The MaxT methods appear to be similar or a more sensitive than the AvgT methods in this data.

2.4 Conclusion

Since cluster mass is defined as the integral of suprathreshold intensities within a cluster, cluster mass can also be considered as the "product" of cluster extent and
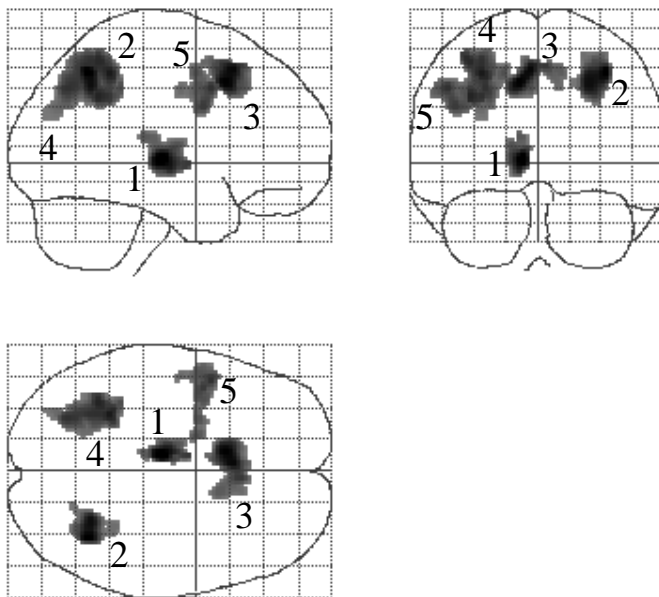
Figure 2.2: 5 largest clusters in working memory data

suprathreshold average intensity. We propose some combining functions with the two statistics to leverage the strength from cluster wise inference method and voxel wise inference method. The inference methods based on those combining functions are built on the nonparametric permutation framework due to less strict assumption about the image itself and unknown distribution of the combining statistics.

Our simulations and real data analysis show that no single method is optimal for all types of signals, however the partial methods (cluster extent specifically) vary considerably in their sensitivity, while our combining methods are usually not far from optional. The main result from this work is that combining with corrected P values appears to be more sensitive than combining with uncorrected P values. Combining based on AvgT, while slightly less optimal based on our real data, was found to be more sensitive in our simulations. In general, for small diameter and moderate intensity (0.5-.5), the AvgT, Tippet and Fisher using AvgT statistics have the best performance.
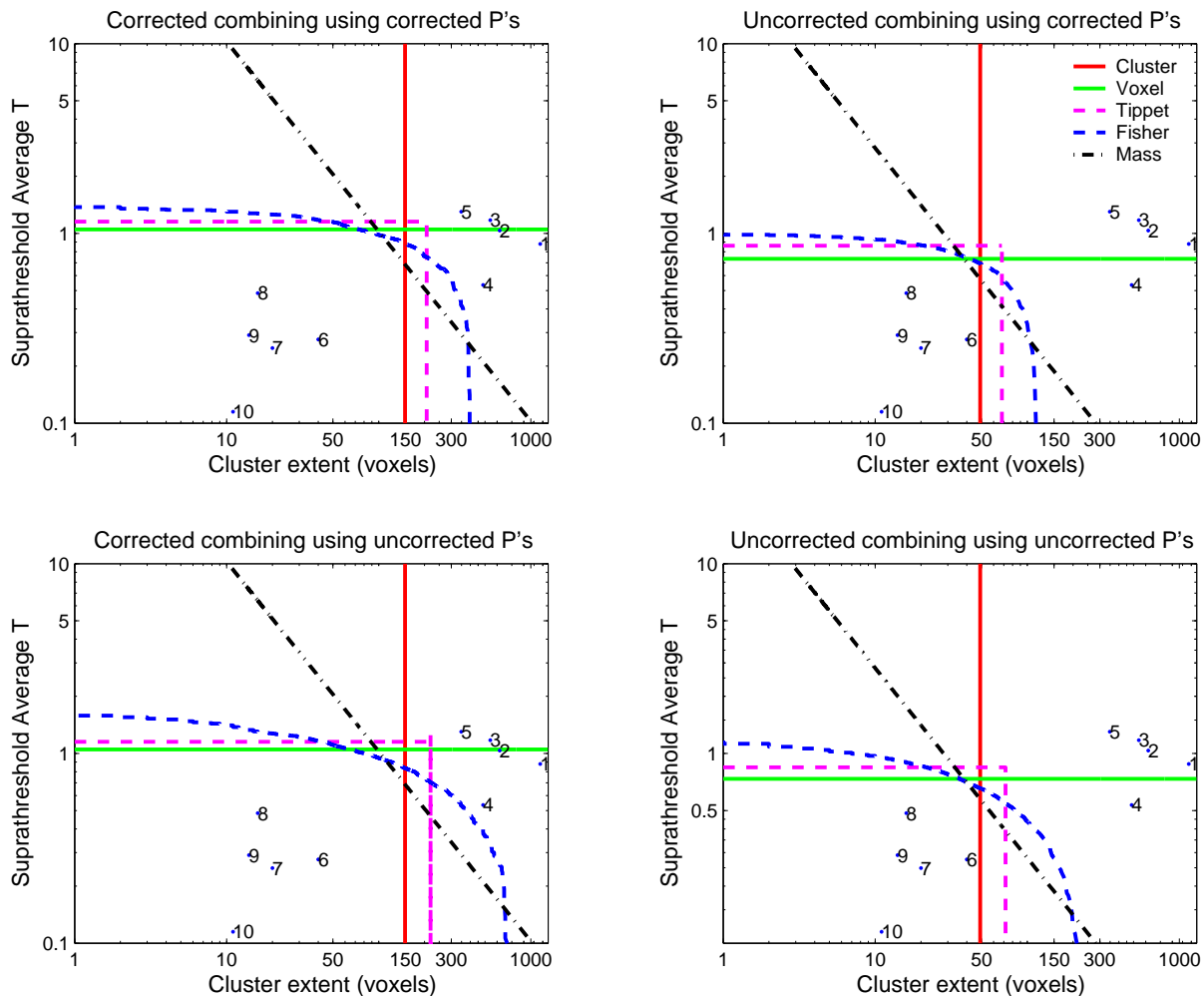
Figure 2.3: Corrected and uncorrected combining using corrected and uncorrected P's. The numbers (1-10) refer to clusters (from largest to smallest) and clusters 1-5 are further studied in Tables 1 & 2

Although there are less strict assumption involved in the permutation framework, there are some limitations with this method. Computational complexity is a concern of this method because of permutation property for large number of subjects in an experiment. For example, the number of permutation in the real data is 4096, which is 2 to the power 12 for 12 subjects. If there are 20 subjects, this number is huge. We can not use all possible permutations . We also can not use this method for a small group analysis, for example, 4 subjects in a group because all possible permutations is $2^4 = 16$. It is not reasonable to find an empirical distribution. In addition, this method

can not be applied for a single subject data analysis since it violates the exchangeability property. We only permute the data labels not voxels. There are not data labels in a single subject analysis. The voxels can not be permuted, otherwise, the spatial structure of brain will be broken.

Poline *et al.* [Poline *et al.*, 1997] developed parametric joint distribution of cluster extent and suprathreshold peak height intensity. Their method is via RFT, which is a way to solve multiple comparison problems, it is reasonable to develop a combining method with superthreshold average intensity and cluster extent via RFT. It is also rational to develop parametric cluster mass inference method since the cluster mass inference is generally the most sensitive one according to our previous study [Hayasaka & Nichols, 2004].

2.5   Acknowledgments

| ♯ i | Extent | AvgT | Corrected P′s | | | | | Uncorrected P′s | | | | |
| | | | Partial | | Combined | | | Partial | | Combined | | |
			Cluster	AvgT	W^T	W^F	W^M	Cluster	AvgT	W^T	W^F	W^M
1	1154	4.9052	0.0002	0.1362	0.0005	0.0002	0.0002	0.0000	0.0226	0.0001	0.0000	0.0000
2	625	5.0607	0.0037	0.0562	0.0078	0.0024	0.0012	0.0006	0.0087	0.0012	0.0005	0.0002
3	542	5.1987	0.0039	0.0249	0.0078	0.0010	0.0012	0.0008	0.0037	0.0014	0.0002	0.0002
4	485	4.5595	0.0051	0.5259	0.0103	0.0244	0.0107	0.0010	0.1425	0.0017	0.0041	0.0020
5	349	5.325	0.0098	0.0105	0.0193	0.0010	0.0027	0.0018	0.0015	0.0032	0.0002	0.0006

Table 2.1: Cluster extent & AvgT: combining with corrected P′s

♯	Extent	MaxT	Corrected P's					Uncorrected P's				
			Partial		Combined			Partial		Combined		
i			Cluster	MaxT	W^T	W^F	W^M	Cluster	MaxT	W^T	W^F	W^M
1	1154	7.3582	0.0002	0.0803	0.0002	0.0005	0.0002	0.0000	0.0131	0.0001	0.0001	0.0000
2	625	9.3708	0.0037	0.0117	0.0068	0.0017	0.0012	0.0006	0.0018	0.0012	0.0004	0.0002
3	542	10.1937	0.0039	0.0051	0.0073	0.0005	0.0012	0.0008	0.0008	0.0013	0.0001	0.0002
4	485	6.3076	0.0051	0.2305	0.0098	0.0220	0.0107	0.0010	0.0458	0.0017	0.0037	0.0020
5	349	13.1468	0.0098	0.0002	0.0002	0.0005	0.0027	0.0018	0.0000	0.0001	0.0001	0.0006

Table 2.2: Cluster extent & MaxT: combining with corrected P's

Chapter 3

Cluster Mass Inference via Random Field Theory

Cluster extent and voxel intensity are two widely used statistics in neuroimaging inference. Cluster extent is sensitive to spatially extended signals while voxel intensity is better for intense but focal signals. To leverage strength from both statistics, several nonparametric permutation methods have been proposed to combine the two methods. Simulation studies have shown that of the different cluster permutation methods, the cluster mass statistic is generally the best. However, to date, there is no parametric cluster mass inference available. In this paper, we propose a cluster mass inference method based on random field theory (RFT). We develop this method for Gaussian images, evaluate it on Gaussian and Gaussianized t-statistic images and investigate its statistical properties via simulation studies and real data. Simulation results show that the method is valid under the null hypothesis and demonstrate that it can be more powerful than the cluster extent inference method. Further, analyses with a single-subject and a group fMRI dataset demonstrate better power than traditional cluster size inference, and good accuracy relative to a gold-standard permutation test.

3.1   Introduction

Cluster extent and voxel intensity are two widely used statistics in neuroimaging inference. Cluster extent is sensitive to spatially extended signals [Friston *et al.*, 1996;

Poline *et al.*, 1997], while voxel intensity is sensitive to focal, intense signals [Friston *et al.*, 1991; Worsley *et al.*, 1992]. Both can suffer from a lack of power for signals of moderate extent and intensity [Hayasaka & Nichols, 2004]. Furthermore, one does not generally know, a priori, whether the generated signal is large in extent, intensity or both. While some practitioners simply select the statistic that gives the most statistically significant test, this embodies a multiple testing problem and will result in inflated false positive error rates. An ideal test statistic would combine spatial extent and peak height intensity and would be sensitive to both without increasing the number of tests considered.

Poline *et al.* [Poline *et al.*, 1997] (henceforth referred to as PWEF) developed a method which combines extent and intensity based on Gaussian random field theory (RFT). They derived the joint distribution of cluster extent and voxel-wise peak height intensity and made inference on minimum P value of a cluster extent test and a local maximum intensity test. However, their method is only applicable to Gaussian or approximately Gaussian images (e.g. a very large group analysis, or a single subject fMRI analysis).

Cluster mass, the integral of suprathreshold intensities within a cluster, naturally combines both signal extent and signal intensity. Initially suggested by Holmes [Holmes, 1994], Bullmore *et al.* [Bullmore *et al.*, 1999] used permutation to obtain cluster mass P values. Currently the cluster mass is the default test statistic in the BAMM[1] and CAMBA[2] software, and is implemented in FSL's randomise[3] tool and in the SnPM[4] toolbox for SPM[5].

Hayasaka & Nichols [Hayasaka & Nichols, 2004] studied the statistical properties of cluster mass along with a variety of other "combining methods" in the permutation testing framework. Among the combining methods they considered were Tippet's method [Lazar

---

[1]http://www-bmu.psychiatry.cam.ac.uk/BAMM
[2]http://www-bmu.psychiatry.cam.ac.uk/software/
[3]http://www.fmrib.ox.ac.uk/fsl/randomise
[4]http://www.sph.umich.edu/ni-stat/SnPM
[5]http://www.fil.ion.ucl.ac.uk/spm

et al., 2002; Pesarin, 2001] (minimum P values, used by PWEF) and Fisher's method (-2 × sum of ln P values). Through simulation studies and analyses of real data they concluded that the nonparametric cluster mass method is generally more powerful than other methods they investigated.

A strength of nonparametric inference methods is that they rely on fewer assumptions about the distributional form of the data. However, they require additional computational effort and are not very flexible. For example, the precise permutation scheme used depends on the experimental design and cannot be trivially determined from a design matrix. Nuisance covariates cannot be accommodated in general, as they induce null-hypothesis structure which violates exchangeability. Also, nonparametric methods cannot be used directly for single subject data analysis as a parametric autocorrelation model or wavelet transformation is needed to whiten the data. For all of these reasons, a parametric cluster mass inference method that can operate with a general linear model and deal with single subject analyses would be of great value.

In this paper we develop a theoretical distribution for the cluster mass statistic via Gaussian RFT. We generalize the work of PWEF, deriving the cluster mass statistic, extending the method to Gaussianized t data. We study the statistical size and power of our test on Gaussian and Gaussianized t image data through simulations and illustrate the method on two real data examples, a single subject fMRI dataset and a group level fMRI data analysis with low degrees of freedom.

3.2 Materials and Methods

3.2.1 Cluster mass test theory

In a mass univariate data analysis, a general linear regression model (GLM)

$$(3.1) \qquad Y_i = X\beta_i + \varepsilon_i$$

is fit for each voxel $i = 1, \ldots, I$, where Y_i is an $N \times 1$ vector of responses, X is a common $N \times q$ design matrix of predictors, β_i is a $q \times 1$ vector of unknown parameters and ε_i

is a $N \times 1$ vector of random errors. Typically, at each voxel, errors are assumed to be independent and identically distributed $N(0, \sigma_i^2)$ random variates, though dependent errors can be accommodated [Luo & Nichols, 2003]. The ordinary least squares estimator of β_i is $\widehat{\beta}_i = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y_i$, and of σ_i^2 is $\widehat{\sigma}_i^2 = e_i^{\mathrm{T}}e_i/\eta$, where $e_i = Y_i - X\widehat{\beta}_i$ and where η is the error degrees of freedom. Then the Student's t-statistic at voxel i is

$$(3.2) \qquad T_i = c\widehat{\beta}_i \left(c(X^{\mathrm{T}}X)^{-1}c^{\mathrm{T}}\widehat{\sigma}_i^2 \right)^{-1/2}$$

where c is a contrast of interest (row vector). We write the t-statistic image as $\mathbf{T} = \{T_i\}_{i=1}^{I}$.

Given cluster-forming threshold $u_c > 0$, the set of suprathreshold statistics $\{T_i : T_i > u_c\}_{i=1}^{I}$ is used to define clusters. Contiguous clusters are defined by a neighborhood scheme, typically 18 connectivity scheme on a three dimensional image.

Let L be the number of clusters found, with cluster ℓ having S_ℓ voxels (i.e. the cluster extent), $\ell = 1, 2, \ldots, L$. Further let I_ℓ be the set of voxel indices corresponding to cluster ℓ. The cluster mass, M_ℓ, of cluster ℓ is the summation of the suprathreshold intensities:

$$(3.3) \qquad M_\ell = \sum_{i \in I_\ell} H_i$$

where $H_i = T_i - u_c$. Note that $M_\ell = S_\ell \bar{H}_\ell$ where $\bar{H}_\ell = \sum_{i \in I_\ell} H_i/S_\ell$ is the average suprathreshold intensity of cluster ℓ, showing cluster mass to be the product of the cluster extent and the average suprathreshold intensity.

To use Random Field Theory results, we begin by assuming that the standardized error images, called the *component fields*, are discrete samplings of a continuous, smooth, stationary Gaussian random process. The component field for scan j is $\{\varepsilon_{ij}/\sigma_i\}_i$, where ε_{ij} is the error for scan j at voxel i. The component fields are assumed to follow a mean zero, unit variance multivariate Gaussian distribution. Stationarity implies that the spatial correlation is determined by an autocorrelation function that is homogeneous over space. The process is regarded as "smooth" if the autocorrelation function has two

derivatives at the origin. Based on these assumptions, the t image defined by (3.2) defines a Student's t random field.

While any univariate random variable can be transformed into a Gaussian variate, or *Gaussianized*, a Gaussianized t image may not resemble a realization of Gaussian random field. Randomness in $\hat{\sigma}_i^2$ reduces the smoothness of the statistic image relative to the component fields [Worsley *et al.*, 1992], as reviewed in Appendix A.2.7. However, Worsley et al [Worsley *et al.*, 1996] argues that when the t degrees of freedom exceed 120, the Gaussianized t-statistic can be regarded as a Gaussian Random Field. Hence we proceed by deriving results assuming $\mathbf{T}$ is a Gaussian image, but later return to the issue of Gaussianization.

The full derivation of our null distribution of the cluster mass statistic is given in Appendix A.2, but we sketch an overview of the result here. The derivation starts by approximating the statistic image about a local maximum as a paraboloid [Siegmund & Worsley, 1995; Yendiki & Fessler, 2007], which allows cluster mass to be obtained a function of cluster extent, S_ℓ, and suprathreshold peak intensity, $H_\ell = \max\{H_i : j \in I_\ell\}$,

$$(3.4) \qquad\qquad M_\ell \approx 2/(D+2) \times S_\ell \times H_\ell$$

where D is the dimension of the image. While this parabolic approximation is essential to the derivation of the null distribution of M_ℓ, note we do not actually fit paraboloids to the image, and the test statistic computed from the data is exactly as specified in Eq. (3.3).

By assuming that the autocorrelation function of the image is proportional to a Gaussian probability density function, the distribution of M_ℓ conditional on H_ℓ can be found. We follow PWEF, making a small excursion assumption that replaces peak height $u_c + H_\ell$ with u_c, creating what we denote the $\mathcal{U}$ result, but also repeat the derivation without this assumption, deriving the $\mathcal{Z}$ result.

Finding the joint distribution of (M_ℓ, H_ℓ) and integrating out H_ℓ yields the final result, an expression for $\mathrm{P}(M_\ell > m)$, the uncorrected P-value for an observed cluster

mass value of m. This requires two numerical integrations, one dependent on u_c, and one on m. In practice, for any given dataset, P-values for a grid m values can be pre-computed and interpolation used to find the P-value for an arbitrary value of m.

This theoretical approach also produces a new result for cluster extent S_ℓ, distinct from the original ([Friston *et al.*, 1991]) result, which we also evaluate for completeness.

As $P(M_\ell > m)$ is an uncorrected P-value which does not account for searching over all clusters in the image, it is only appropriate for a single cluster that can be pre-identified before observing the data [Friston, 1997], a situation that rarely arises in practice. As detailed in Appendix A.2, the uncorrected P-values can be transformed into familywise-error corrected P-values which accounts for the chance of one or more false positive clusters anywhere in the image.

Student's *t*-statistic image

As discussed above, when the degrees of freedom are small a Gaussian random field will not provide a good approximation for a Student's *t*-statistic image. In such cases we Gaussianize the t image via the probability integral transform. The transformed image, however, will be rougher than the component fields, and so the roughness parameter must be adjusted according to the degrees of freedom of the *t*-statistic image. Thus we can apply our method to Gaussianized t images with just a modification to the smoothness estimate, as described in Appendix A.2.7.

3.2.2   Simulations

To evaluate the accuracy of our cluster mass result, Equation (3.4), both 2D (256×256) and 3D ($64 \times 64 \times 30$) Gaussian noise images are simulated. To understand the influence of image roughness on the proposed statistic, each of the 10,000 independent Gaussian noise images are convolved with different isotropic Gaussian smoothing kernels. Kernel sizes 2, 4, 8, 10, and 12 voxels full width at half maximum (FWHM[6]) are

---

[6]Kernel standard deviation = FWHM / $\sqrt{8 \ln 2} \approx 0.4247$ FWHM

used, and these sizes then directly determine $|\Lambda|$, the image roughness parameter. Two cluster forming thresholds are investigated ($u_c = 2.326$ and $u_c = 3.090$, corresponding to uncorrected $P = 0.01$ and $P = 0.001$, respectively). A nominal significance level of 0.05 is used for all inferences.

To evaluate the method on Gaussianized t-statistic images, 15 Gaussian noise images are simulated, mean-centered and divided by the voxel-wise standard error to produce 14 degrees-of-freedom t images. A t-to-z transformation is then applied to generate Gaussianized t images with the necessary adjustment to the smoothness parameter (Appendix A.2.7).

To assess the power of our method, a spherically shaped signal (radius 1, 3, 5, 7, 10mm) with various uniform intensities (0.25, 0.5, 0.75, 1, 1.5, 2) is added to the center of Gaussian noise images. Power is measured as the probability of a true positive cluster, defined a significant cluster that contains one or more non-null voxels. The cluster extent inference methods are those from RFT [Adler, 1981] implemented in the Statistical Parametric Mapping (SPM) software.

One objective of the evaluations is to determine whether the $\mathcal{U}$ result, based on the small excursion approximation, or the $\mathcal{Z}$ result is more accurate. Since the derivation depends on the joint distribution of cluster mass and peak height, we examine the approximation accuracy of our results for this bivariate distribution with simulation. In addition to visualizing images of the predicted and simulated densities for the $\mathcal{Z}$ and $\mathcal{U}$ results, we compute the Kullback-Leibler divergences [Kullback & Leibler, 1951], a measure of distance between two distributions. This allows a quantitative comparison between the two results.

The ultimate accuracy of the method depends on the marginal distribution of cluster mass. We compare the specificity and validity of the mass test statistic for the $\mathcal{U}$ and $\mathcal{Z}$ results, as well as cluster size P-values found with our derived cluster extent distribution and cluster extent P-values produced by SPM. We present results for both uncorrected

and corrected P values to understand the performance of the method, though only the corrected P-values are of practical interest. The specificity and validity is gauged with plots of theory-based P-values versus Monte Carlo ("true") P-values, called P-P plots. When a method has exact specificity the theory will produce the same P-value as Monte Carlo simulation, and the plotted line will follow the identity. When a method is conservative the line will fall above the identity, and when anticonservative (fails to control Type I error rate) the line will fall below the identity.

3.2.3   Applications

We demonstrate our cluster mass inference method on two fMRI data sets, one single subject and one group dataset

FIAC data

The first example is the Functional Imaging Analysis Contest (FIAC)[7] example . The experiment uses a sentence listening task, considering effects of different or same speakers and different or same sentences. We only consider the sentence effect "Different Sentence vs. Same Sentence": In each block, six sentences are read; in the "Different" condition six different sentences are read, while in "Same" condition the same sentence is repeated six times. For complete details see [Madic & Group, 2005].

We use subject 3 ("func4"), block design data with 6mm FWHM smoothing, fit with a GLM which produces a t statistic image with 179 degrees-of-freedom. Here we can assume that the t image reasonably approximates a Gaussian image and use the method directly on the t image. The cluster forming threshold is $P = 0.001$ uncorrected.

---

[7]http://www.madic.org/fiac/

Working Memory Data

We also use a group level analysis with 12 subjects from a working memory experiment. Since the degrees of freedom are rather small (11), we perform a t-to-z transformation to generate a Gaussianized t image.

While the experiment considers different aspects of working memory, we only use the item recognition task. In the item recognition condition subjects are shown a set of five letters and, after a 2 second delay, shown a probe, to which respond "Y" if it was in the set, or "N" otherwise; in a control condition five "X"s are shown and the probe is just "Y" or "N" indicating the required response. For full details see Marshuetz *et al* [Marshuetz *et al.*, 2000].

A one-sample t-test is used to model the data. We use t-to-z transformation and a cluster defining threshold of $P = 0.01$ uncorrected ($t_{11} = 4.02$ or $z = 3.09$). The roughness parameter is adjusted by 1.3891 [Holmes, 1994; Worsley *et al.*, 1992] to account for increased roughness of the Gaussianized t statistic. In addition to parametric results in SPM, we also use SnPM to obtain nonparametric cluster extent and mass results (see Appendix A.1 for a summary of permutation cluster inference). With 12 subjects there are $2^{12} = 4096$ possible sign flips of the contrast data to create a permutation distribution.

3.3 Results

3.3.1 Simulations

For the simulation studies, we only show results for a smoothness parameter of FWHM = 8 voxels, as the results are similar to the other smoothness parameters.

Accuracy of derived joint distribution

The top row of Figure 3.1 shows the true (simulated) joint distribution of cluster mass and peak height intensity, the $\mathcal{Z}$ result and the $\mathcal{U}$ result for 3D Gaussian noise

images. The bottom row shows difference images of true and derived distributions for the $\mathcal{Z}$ and $\mathcal{U}$ results. The distributions are qualitatively similar, though for very small cluster masses and cluster height around 0.5 to 1.0, the two results tend to underestimate the truth; while for cluster mass between 0 and 50 and cluster heights between 0 and 0.5, the results can overestimate the truth. The Kullback-Leibler divergences are 1.285 for the $\mathcal{Z}$ result and 1.610 for the $\mathcal{U}$ result.

Figure 3.2 displays corresponding results for 3D Gaussianized t image. Again, there is little difference between the true distribution and the two results, and again the Kullback-Leibler divergence between the true distribution and the $\mathcal{Z}$ result is smaller than that between the true distribution and the $\mathcal{U}$ result (1.701 vs. 2.338). Thus, for both Gaussian images and Gaussianized images, the $\mathcal{Z}$ result appears to be superior to the $\mathcal{U}$ result.

Accuracy of derived cluster mass null distribution

Figure 3.3 shows the P-P plots for 3D Gaussian null simulated data and Figure 3.4 3D Gaussianized t-statistic null simulated data. Both cluster mass (dot-dashed lines) and cluster size results (solid lines) are shown. For all of our derived methods, the $\mathcal{U}$ results are more conservative (the null will be rejected less often than nominal) than the $\mathcal{Z}$ results. The SPM cluster size results are also more conservative than the $\mathcal{Z}$ results for Gaussian null simulated data and the $\mathcal{U}$ results for Gaussianized t-statistic null simulated data. While our $\mathcal{Z}$ result for cluster size exhibits some anticonservativeness, overall the $\mathcal{Z}$ result of cluster mass is the least conservative method, while maintaining validity over most of the range of probabilities included in this simulation study.

Figure 3.5 shows the Type I error rates for a 3D Gaussianized t image with 14 degrees of freedom with various smoothness parameters (FWHM) and cluster defining thresholds. The figure shows that the $\mathcal{Z}$ cluster mass result provides better results for high thresholds and large FWHM than for low threshold and low FWHM. For corrected P values, this result is valid for all levels of smoothing studied, whereas the $\mathcal{Z}$ result
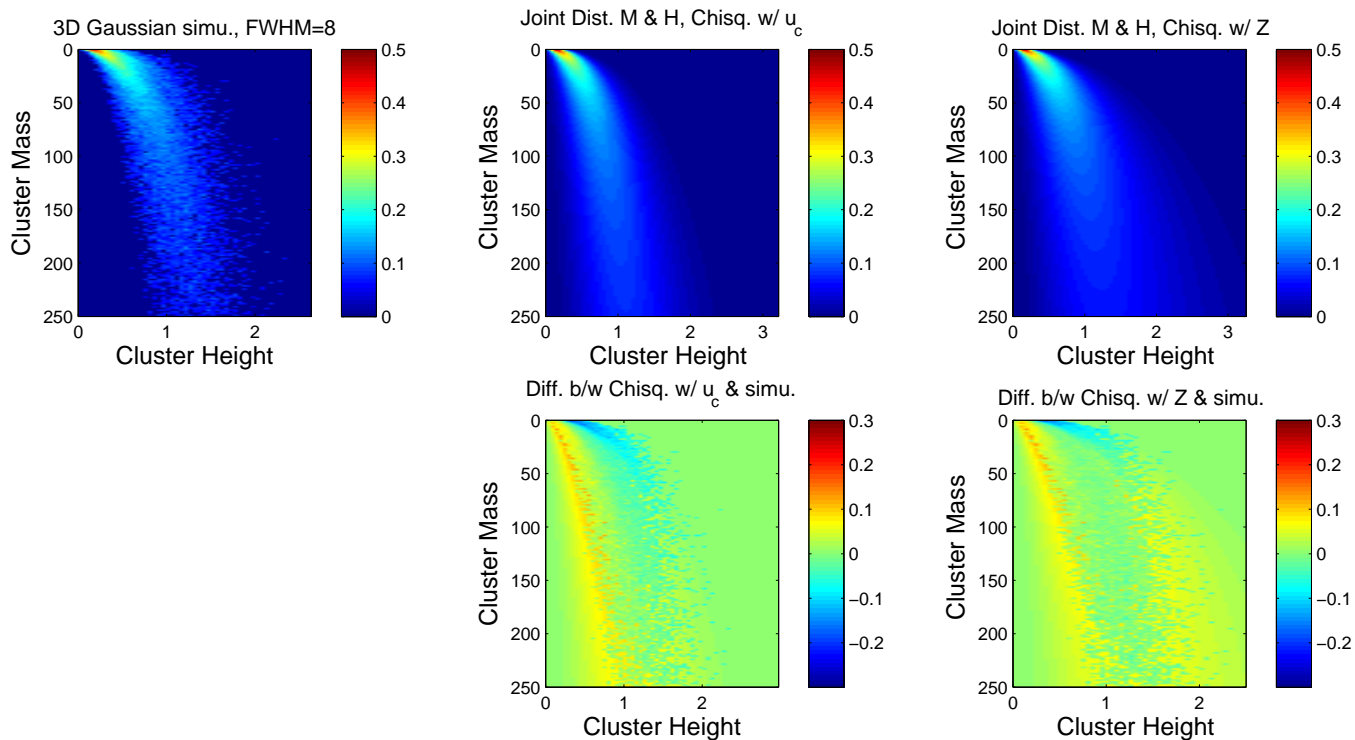
Figure 3.1: Comparison of true and theoretical joint distributions of cluster mass and peak height intensity, for Gaussian images. On top left is the true distribution obtained from simulation, on the top middle is the $\mathcal{U}$ result and on the top right is the $\mathcal{Z}$ result. Below each of the theoretical results is the true minus estimated distributions. While only an intermediate result, the agreement is reasonable, with better performance obtained with the $\mathcal{Z}$ result. All distributions are transformed by the fourth root to improve visualization. Unless otherwise noted, simulation settings used in the figures are: $u_c = 2.3263$ (p=0.01), $64 \times 64 \times 30$ image at FWHM 8 voxels.

of cluster extent is, by and large, invalid. Furthermore, the $\mathcal{Z}$ cluster mass corrected P-values—those that are used in practice—are always closer to the nominal significance level when correcting for multiple comparisons.

Power comparisons

Having found our own cluster extent result to be invalid, we compare the power of our $\mathcal{Z}$ cluster mass result only to SPM's cluster extent result. Figure 3.6 lists simulated power for the cluster extent (SPM) and cluster mass ($\mathcal{Z}$). As expected, for a given intensity, the power increases with signal intensity, and, for a given radius, power increases as
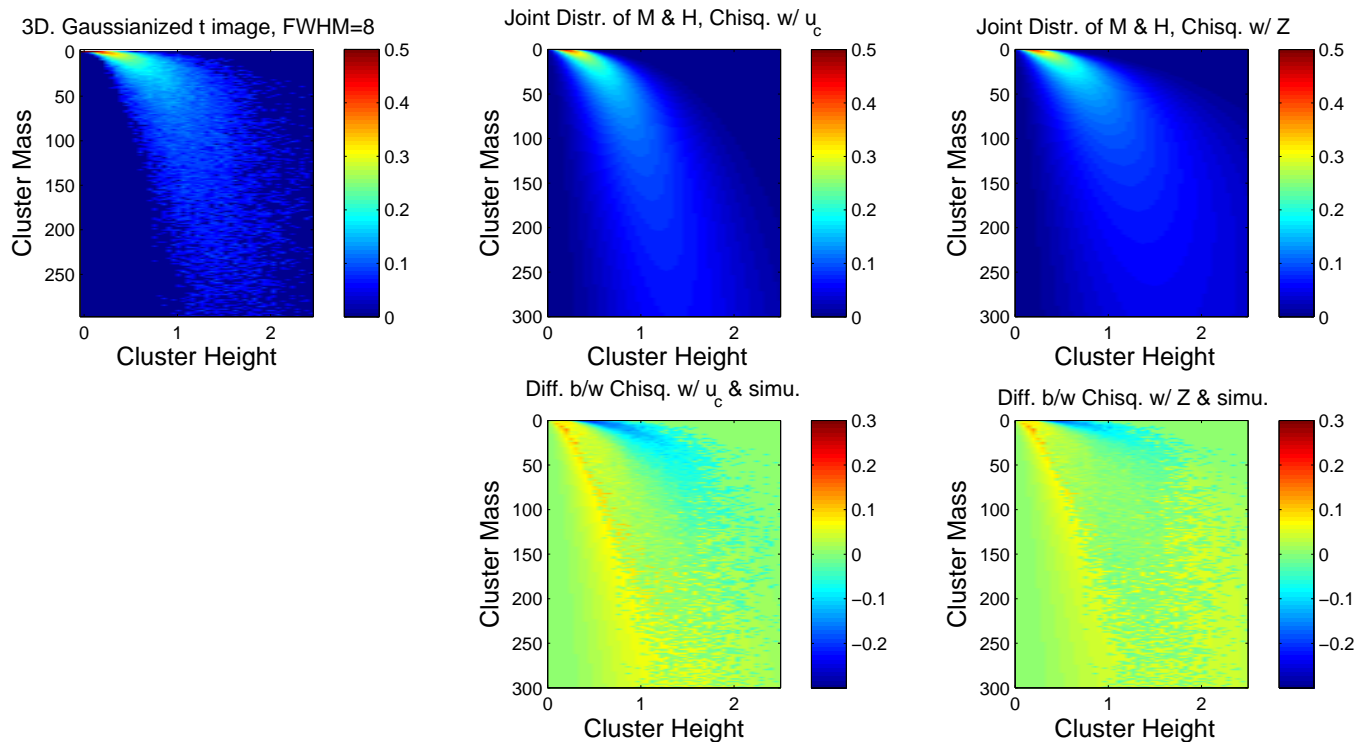
Figure 3.2: Comparison of true and theoretical joint distributions of cluster mass and peak height intensity, for Gaussianized t_{14} images. Same format as in Figure 3.1. Again the agreement between simulated truth and derived theoretical result is good, with a closer match seen with the $\mathcal{Z}$ result.

the signal intensity increases. When the image smoothness is low (FWHM ≤ 4 voxels), SPM cluster extent generally provides better power than the $\mathcal{Z}$ mass result. However, for greater smoothness (FWHM ≥ 8 voxels), the $\mathcal{Z}$ result is more powerful than SPM, regardless of signal extent or signal intensity.

3.3.2 Real Data Evaluations

The FIAC data results show the method's performance at high degrees-of-freedom, while the working memory data assess the method using Gaussianization of the t image.

FIAC data

The estimated smoothness of the component fields based on the residuals is [2.4964 2.3599 1.7525] voxel FWHM with 27,862 $3.0 \times 3.0 \times 4.0 \text{mm}^3$ voxels. Figure 3.7 shows
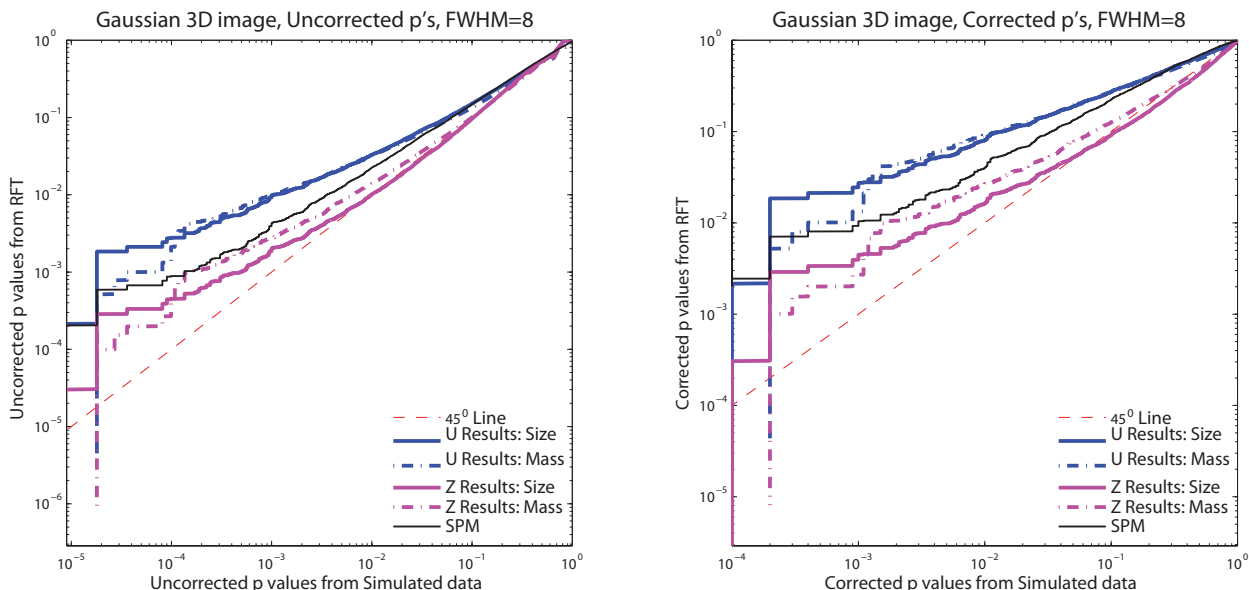
Figure 3.3: Monte Carlo simulation P-values versus theoretical P-values for uncorrected and corrected P-values with Gaussian images. Values in the plot above the identity indicate conservative performance, below the identity invalid performance. Our $\mathcal{Z}$ cluster mass method exhibits slightly conservative performance, but much less conservative than the other methods.

the maximum intensity projection of the all clusters found with a $P = 0.001$ threshold, the three most prominent being a pair of bilateral activations in inferior frontal gyri and one in the frontal pole. Note that the primary auditory cortex effect did not survive $P = 0.001$ threshold, and inspection of the unthresholded statistic image suggests the frontal pole cluster is a false positive activation due to susceptibility artifacts. However, the general shape and size of the clusters are still representative of true positive signals and are useful for evaluating our method.

Table 3.1 provides the values of cluster extent, suprathreshold peak height intensity and cluster mass for each cluster, as well as the P-values, all sorted by peak height. The first three clusters have corrected significance with cluster mass, while peak height and cluster extent only find one cluster significant each. The uncorrected significances show
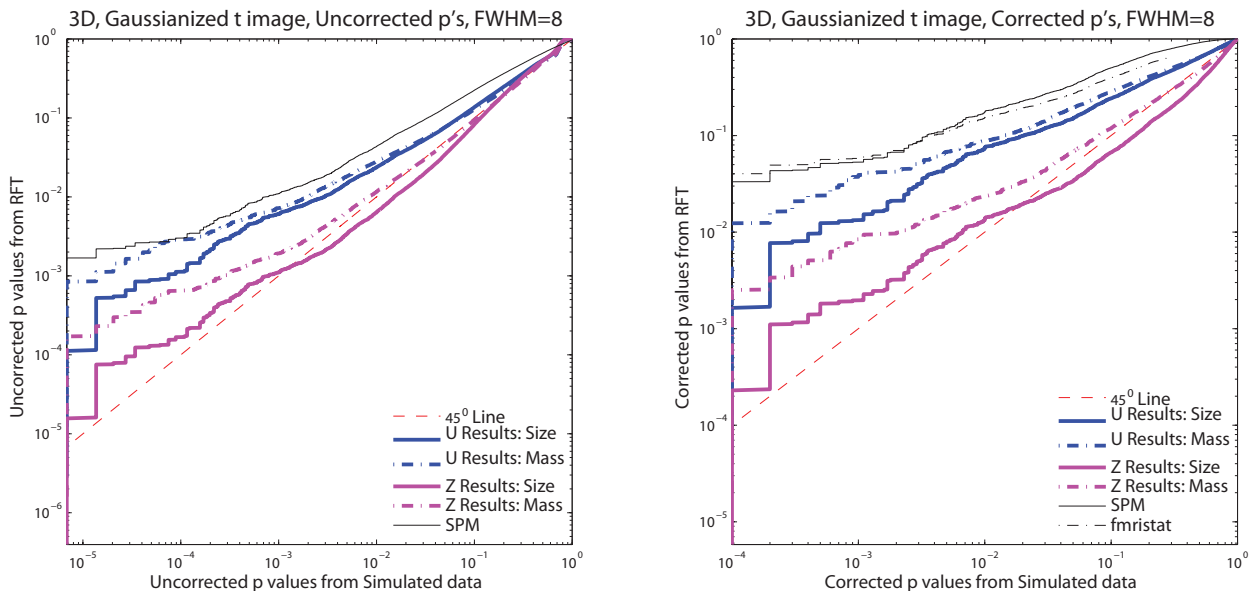
Figure 3.4: Monte Carlo simulation P-values versus theoretical P-values for uncorrected and corrected P-values with Gaussianized t_{14} images. Despite Gaussianization, our $\mathcal{Z}$ cluster mass method provides close to exact performance, and less conservative performance than other methods.

that if a cluster is significant by any of the three methods, it is significant by cluster mass. Again, while we do not advocate use of uncorrected inferences, this demonstrates the relative sensitivity of the method.

Working Memory Data

The estimated smoothness is [4.8611 6.4326 6.6156] voxel FWHM with 122,659 $2.0 \times 2.0 \times 2.0$ voxels. Figure 3.8 shows the all of the clusters found with a $P = 0.001$ cluster-forming threshold. Table 3.2 compares our RFT cluster mass results to an equivalent permutation method. Our RFT method finds the five largest clusters significant, as does the RFT cluster size statistic. Notable is the close correspondence between the RFT P-values and the permutation P-values.

3.4   Discussion and conclusion

Although cluster mass inference with nonparametric permutation has been found to be a quite sensitive inference method for neuroimaging data [Hayasaka & Nichols, 2004], permutation is computationally intensive, not a very flexible modeling framework. We propose a new theoretical cluster mass inference method for Gaussian images and Student's t-statistic images, based on Gaussian RFT. Our simulation studies show that our derived null distribution is accurate, and performs well not only for Gaussian images, but also for Student's t-statistic image. Like other RFT methods, our results depend only on the smoothness and the volume of the image. While we did not find closed form results for the P-value for an arbitrary mass value, the P-value can be quickly found based on interpolation of a pre-computed look-up table.

Our evaluations of the test's specificity reveal that the proposed cluster mass inference method works best when the image is sufficiently smoothed, at least 4 voxel FWHM, and ideally for larger smoothness parameters (FWHM $\geq$ 8 voxels). We stress that this is a substantial magnitude of smoothness (typical estimated smoothness is FWHM 2-4 voxels). However, our real data evaluations found our method to perform as good or better than parametric cluster size inference, even though image smoothness was only about 2 voxels FWHM in the single subject dataset. Hence, even with slightly conservative P-values, the mass statistic appears very sensitive to real data signals.

Consistent with findings using the nonparametric cluster mass inference method, our theoretical cluster mass inference statistic generally has better power than either the cluster extent inference statistic or the voxel intensity statistic, alone. This is especially true when the cluster extent and the suprathreshold peak height intensity are moderately sized. More remarkable, is that despite a large number of assumptions and a sequence of approximations, our RFT cluster mass P-values are very close to the permutation results which have very few assumptions.

The Gaussianization of t images is a shortcoming of the method, but it is not an uncom-

mon strategy. The FSL [8] software has always (as of version 4.0) used Gaussianization of t and F images. While the SPM software has abandoned Gaussianization for voxel-wise inference ever since SPM99, its cluster extent inference has always (as of SPM5) used Gaussian and not t random field results cluster extent P-values and currently neglects the smoothness adjustment described in Appendix A.2.7.

Although the proposed cluster mass inference method has many good statistical properties, it has its limitations. When we derive the formulas for the marginal distribution of cluster mass, we assume that the shape of a cluster above a certain threshold is approximated by a paraboloid. This assumption is rational for a Gaussian image that has been convolved with a Gaussian smoothing kernel. However, for real data, this assumption may be too strong, even after smoothing the data. For example, we may have a large flat cluster with only one voxel of high intensity. The activated regions may also have other shapes that are not well approximated by a paraboloid. In addition, we use a Gaussian shaped correlation function to simplify the variance in the derivation. We also assume that we have stationary fields, though an extension to accommodate local variation in smoothness [Hayasaka *et al.*, 2004] may be possible.

While we have only attempted to derive Gaussian results, a reviewer notes that [Worley, 1994] derived the Hessian of a t field which, when simplified by conditioning and combined with results from [Cao, 1999], could provide a means to derive t cluster mass statistic.

Finally we note that, while both real data examples were fMRI, the method makes no assumptions about the modality and should operate well with PET and other types of imaging data. To this end, an extension to SPM will be available soon to allow use of our results; check the SPM Extensions website[9] for a link.

---

[8]http://www.fmrib.ox.ac.uk/fsl/
[9]http://www.fil.ion.ucl.ac.uk/spm/ext

3.5   Acknowledgments
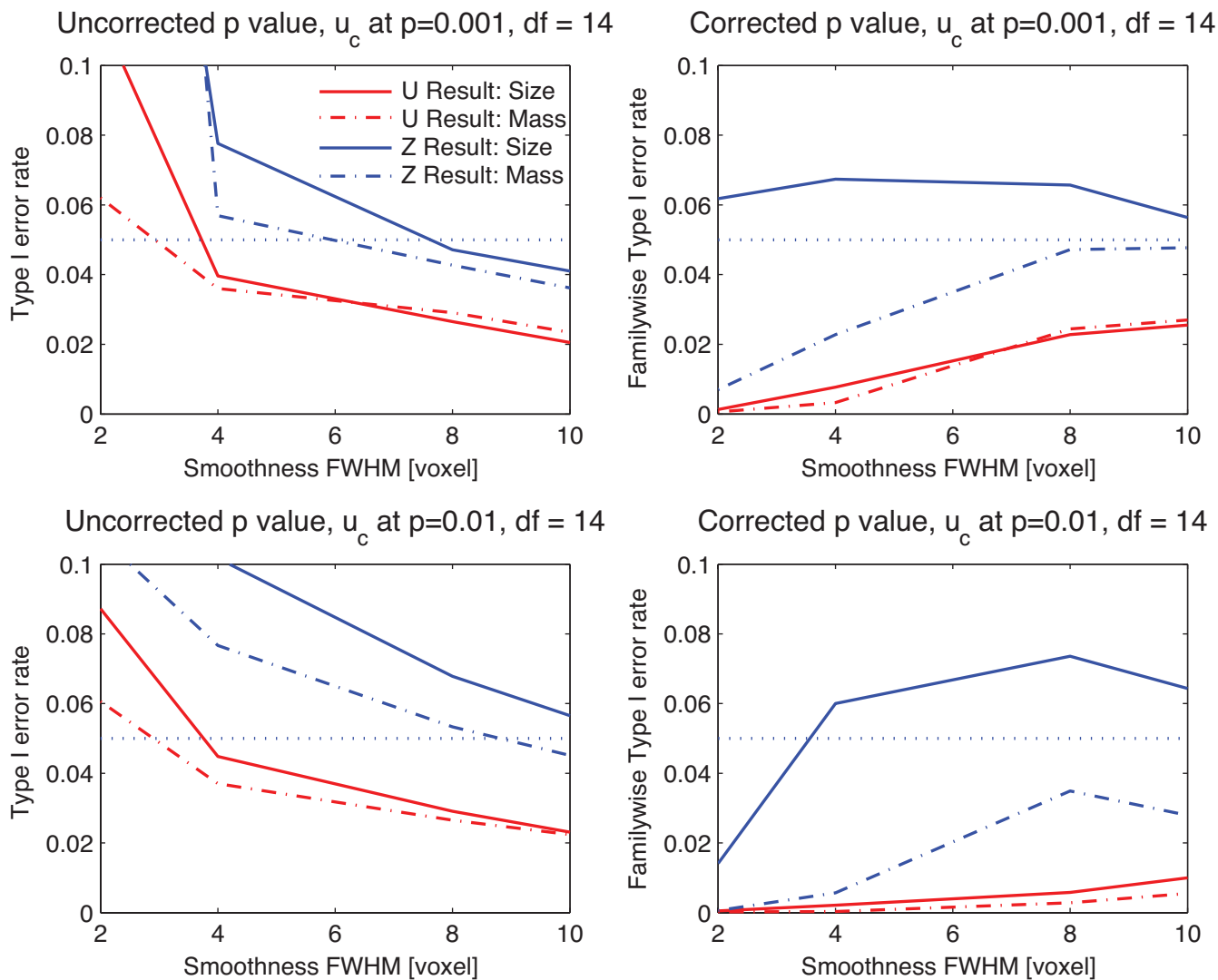
Figure 3.5: Type I error rate for Gaussianized t images, for both $P = 0.01$ and $P = 0.001$ cluster-forming thresholds, with different smoothness. While uncorrected P-values perform poorly under low smoothness, our $\mathcal{Z}$ cluster mass method has the corrected P-values that are closest to the nominal $\alpha = 0.05$ level without being invalid.
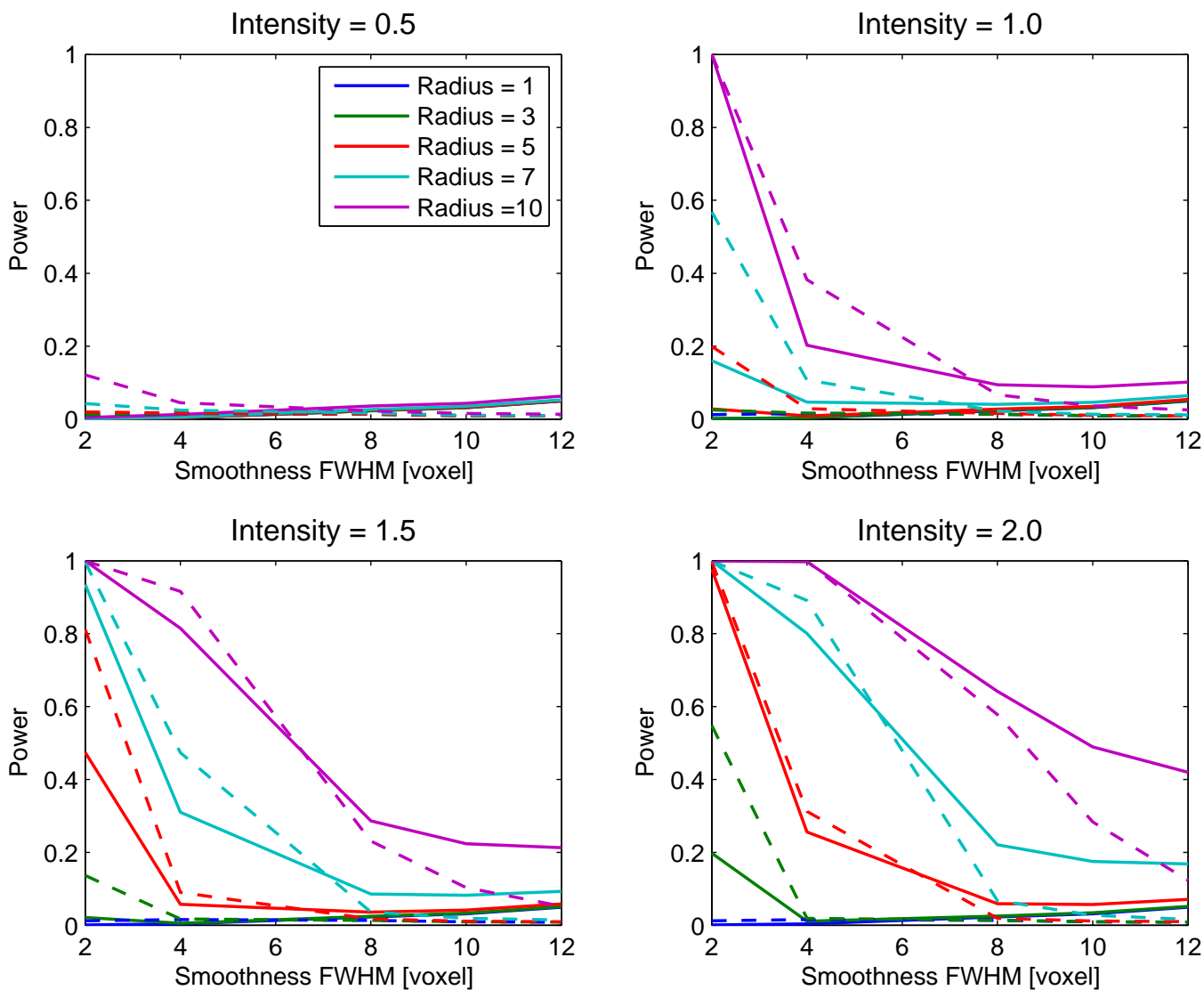
Figure 3.6: Power of our proposed cluster mass inference method (solid lines), compared with standard cluster extent inference method implemented in SPM (dashed lines), for different cluster sizes and signal intensities. Gaussian images were used with a cluster defining threshold of 2.3263 (p=0.01).
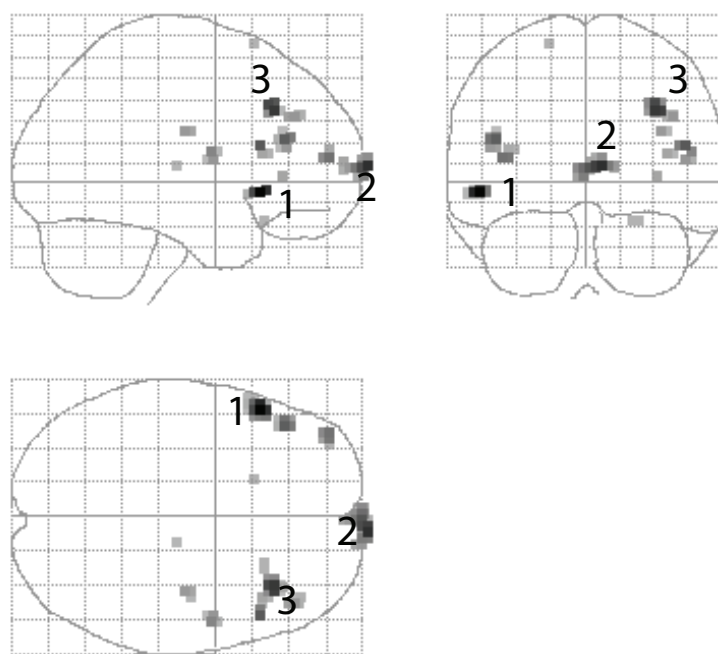
Figure 3.7: Results for "sentence" effect in FIAC single subject data.

| Cluster | Cluster | | | Uncorrected P values | | | Corrected P values | | | Location |
No	Extent	Height	Mass	Extent	Height	Mass	Extent	Height	Mass	(x,y,z mm)
1	13	5.09	9.35	0.0069	0.0008	0.0011	0.1606	0.0192	0.0279	(-52,22,-5)
2	24	4.52	12.54	0.0009	0.0092	0.0004	0.0238	0.2096	0.0106	(8,75,8)
3	13	4.45	7.97	0.0069	0.0122	0.0018	0.1606	0.2665	0.0451	(34,29,35)
4	5	4.10	2.09	0.0633	0.0463	0.0404	0.7999	0.6920	0.6425	(49,22,18)
5	10	4.08	3.60	0.0140	0.0508	0.0138	0.2992	0.7251	0.2959	(-44,34,21)
6	6	3.87	2.60	0.0446	0.1056	0.0269	0.6782	0.9319	0.4960	(-41,56,12)
7	5	3.65	1.22	0.0633	0.2134	0.0967	0.7999	0.9956	0.9145	(52,-2,15)
8	5	3.48	0.98	0.0633	0.3492	0.1334	0.7999	0.9999	0.9664	(73,39,32)
9	3	3.43	0.64	0.1447	0.4013	0.2324	0.9764	1.0000	0.9973	(37,-15,25)
10	1	3.34	0.25	1.0000	0.5261	0.6816	1.0000	1.0000	1.0000	(35,33,3)
11	2	3.21	0.22	0.2433	0.7304	0.7648	0.9979	1.0000	1.0000	(23,24,-19)
12	1	3.18	0.09	1.0000	0.7924	1.0000	1.0000	1.0000	1.0000	(-18,19,68)
13	1	3.16	0.07	1.0000	0.8429	1.0000	1.0000	1.0000	1.0000	(13,-19,8)

Table 3.1: Real data results for FIAC single subject data analysis, comparing extent, peak height and mass statistics for cluster inference. The cluster mass has good sensitivity, and, in particular, when any of the three inference methods are significant, cluster mass is usually significant,
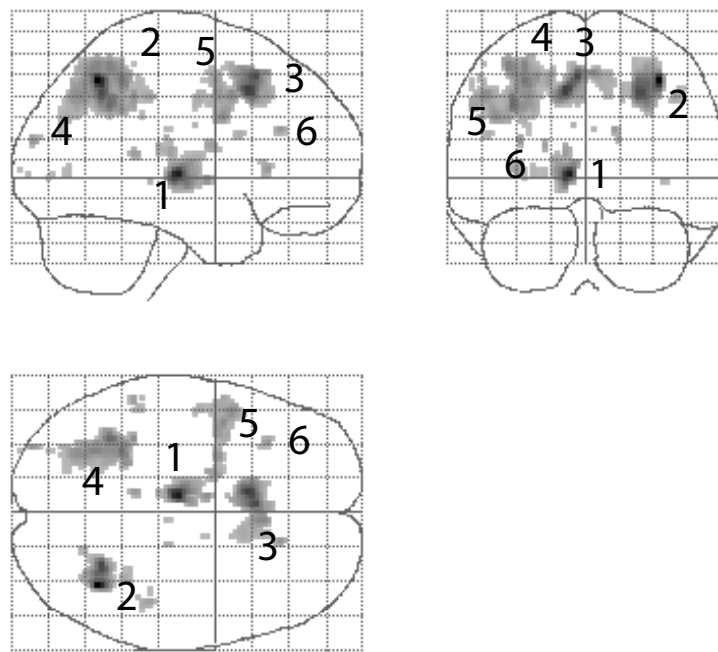
Figure 3.8: Results from item recognition effect in the working memory data.

Cluster	Random Field Theory Cluster Mass Inference									
Cluster		Cluster			Uncorrected p-values			Corrected p-values			Location
No	Extent	Height	Mass	Extent	Height	Mass	Extent	Height	Mass	(x,y,z mm)
1	347	5.47	182.19	0.0005	0.0001	0.0002	0.0043	0.0011	0.0018	(-8,-18,2)
2	540	4.99	262.29	0.0001	0.0012	0.0001	0.0007	0.0111	0.0004	(36,-58,48)
3	620	4.82	272.05	0.0000	0.0026	0.0001	0.0004	0.0231	0.0004	(-10,16,44)
4	1150	4.34	448.15	0.0000	0.0192	0.0000	0.0000	0.1602	0.0000	(-30,-46,48)
5	481	4.02	119.41	0.0001	0.0621	0.0008	0.0012	0.4313	0.0076	(-48,8,40)
6	40	3.43	5.26	0.1012	0.4110	0.1684	0.6014	0.9761	0.7836	(-34,24,4)

Cluster	Permutation-based Cluster Mass Inference									
Cluster		Cluster			Uncorrected P values			Corrected P values			Location
No ‡	Extent	Height	Mass	Extent	Height	Mass	Extent	Height	Mass	(x,y,z mm)
1	347	5.47	182.19	0.0018	0.0000	0.0007	0.0098	0.0002	0.0034	(-8,-18,2)
2	540	4.99	262.29	0.0008	0.0008	0.0003	0.0039	0.0051	0.0015	(36,-58,48)
3	620	4.82	272.05	0.0006	0.0018	0.0002	0.0037	0.0117	0.0012	(-10,16,44)
4	1150	4.34	448.15	0.0000	0.0132	0.0000	0.0002	0.0803	0.0002	(-30,-46,48)
5	481	4.02	119.41	0.0010	0.0461	0.0018	0.0049	0.2305	0.0093	(-48,8,40)
6	40	3.43	5.26	0.0658	0.3327	0.1202	0.2759	0.7515	0.4312	(-34,24,4)

Table 3.2: Real data results for the small group fMRI data, comparing RFT parametric and permutation nonparametric inferences. Note the similarity between the RFT P-values and permutation P-values, even though the RFT method depends on many assumptions and approximations.

Chapter 4

Optimizing Kernel Size for the Smoothed Variance t-test

Previous study showed that, for small group studies with 20 or fewer subjects, the smoothed variance t-test has been found to increase detection sensitivity and to be a powerful alternative to the usual t-test [Nichols & Holmes, 2002; Hayasaka & Nichols, 2003]. The reason is that the effective degrees of freedom (EDF) of a smoothed sample variance image will be larger than the degrees of freedom (DF) of the original sample variance image. However, smoothing a sample variance image will induce bias because a voxel based sample variance is an unbiased estimator of the true variance. Moreover, the smoothing kernel size applied to sample variance is another concern since we do not know how much is enough. The concern is that with a large smoothing kernel, high variance "hot spot" may be smoothed out. This smoothing procedure will cause too many false positive results. If variance images are smoothed with a small size smoothing kernel, the EDF do not change appreciably. One purpose of this study is to increase EDF in order to increase detection sensitivity while avoiding too much bias. In this work, we study the relationship between the smoothing kernel size and EDF, mean square error (MSE), bias and variance (VAR) of smoothed sample variance images, while accounting for the data smoothness and true variance variance smoothness. The final goal is to find the optimal smoothing kernel size for smoothed sample variance images.

4.1   Introduction

Generally, in two level fMRI data analysis, the degrees of freedom (DF) at the first level study are large enough to capture the significant signals because there are over 100 scans for the whole experiment. The t statistic images can be regarded as Gaussian image because of large DF. However, in the second level analysis, it is not that case. For small group studies with 20 or fewer subjects, the threshold is relatively high compared with large group studies having over 100 subjects. Since the expense for a large experiment with many subjects is large, the sample size is often small. However, the results from a small sample have lower detection sensitivity because of large critical value. For example, if we use voxel wise inference method, the critical value is 2.8214 (p=0.01) for a t distribution with 9 DF but 2.3642 when the DF is 100. Given a t statistic image, the lower the valid critical value, the greater the power. If the effective DF (EDF) can be increased, the detection sensitivity will increase because of relative lower critical value.

Our previous studies has found that the smoothed variance t-test is a powerful alternative to the usual t-test [Nichols & Holmes, 2002; Hayasaka & Nichols, 2003]. The power is greater than the usual one because smoothed variance images effectively increases the error DF. Lukic *et.al* had also shown that pooled variance smoothing would have better detection results than those from the usual methods [Lukic *et al.*, 2002]. Worsley *et. al* [Worsley *et al.*, 2002] considered to smooth the ratio of the random effects variance to the fixed effects variance for group data analysis when the fixed effects variance were assumed to be fixed and same in different subjects. They also used the Satterthwaite approximation to develop the relationship between the EDF and the size of smoothing kernels, which were applied for the ratio images.

Since an unsmoothed sample variance image is unbiased estimator of the true variance image, any smoothing will induce bias. The severity of bias depends on the size of smoothing kernel used for true variance images. In addition, if the true mixed-effects variance image for the second level data analysis is not proportional to the fixed effects

variance, Worsley *et. al*'s ratio smoothing approach will not succeed. In this work, we develop a theoretical framework to obtain bias, mean squared error (MSE), EDF and variance (VAR) of the smoothed sample variance image, as a function of (1) data smoothness, (2) true variance image smoothness and (3) applied variance smoothing. One novel aspect is we use of χ^2 random fields to model the true variance images. We compare our theoretical results with empirical ones derived from 10-subject samples of a 150-subject dataset. We also compare the theoretical results with simulated ones to evaluate the proposed relationship between different smoothness parameter and bias, MSE, EDF and VAR of the smoothed variance images.

4.2 Methods

4.2.1 Models

Generally, there are over 100 scans for each subject in fMRI data analysis. In two level data analysis, it is not necessary to smooth the sample variance at the first level since the DF are large enough that t statistic images at this level can be regarded as normal images. In this study, we only focus on the second level analysis in a two level fMRI data analysis. We use a random effects model for the two level study.

Suppose we have a model,

The first level,

For the vth voxel of the kth subject,

$$Y_k(v) = X_k(v)\beta_k(v) + \epsilon_k(v)$$

where, $Y_k(v)$ is a T time series vector of fMRI response, $X_k(v)$ is a $T \times p$ design matrix, $\beta_k(v)$ is a p vector, and $\epsilon(v) \sim N(0, \sigma^2(v))$

The second level,

$$\beta_k(v) = Z_k(v)\beta(v) + \eta_k(v)$$

where, $Z_k(v)$ is the inter subject parameter for the kth subject, $\beta(v)$ is the group-level parameter at the vth voxel, and $\eta_k(v) \sim N(0, \mathbf{\Sigma}(v))$.

Let $\hat{\beta}_k(v)$ denotes the estimate from the first level, and $\tilde{\beta}_k(v)$ denotes the estimate from the second level. We also assume there are N subjects in an experiment. Then, the usual sample covariance estimate is,

$$\hat{\Sigma}(v) = \frac{1}{N-1}\sum(\hat{\beta}_k(v) - \tilde{\beta}_k(v))^t(\hat{\beta}_k(v) - \tilde{\beta}_k(v))$$

If we assume the second level parameter $\beta(v)$ is a scalar, this will be the usual univariate sample variance of the first level $\hat{\beta}_k$. We then smooth the sample variance in order to have large EDF yet avoid too much bias. We estimate MSE, VAR of the smoothed variance image and EDF for different Σ cases separately. We develop methods for a Σ that is constant, heterogeneous but fixed and heterogeneous random three cases, respectively.

Spatially homogeneous variance images

We first consider a special case where the true variance image Σ is known as homogeneous. $\Sigma = \sigma_0^2$, where $\Sigma = \{\sigma_1^2, \ldots, \sigma_{N_p}^2\}$, N_p is the number of voxels within an image, and $\sigma_0^2 = \{\sigma_v^2 = \sigma_0^2, v = 1, 2, \ldots, N_p\}$. In this case, according to the derivation B.1, there is no bias due to variance smoothing. The EDF, VAR and MSE are,

$$(4.1) \qquad 1/\text{EDF} \approx \frac{1}{\eta}\cdot\left[1 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2}$$

$$(4.2) \qquad \text{MSE} = \text{VAR} = \frac{2\sigma_0^4}{\eta}\cdot\left[1 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2}$$

where F_V is full width half maximum (FWHM) of the smoothing kernel used for a sample variance image, F_D is FWHM of the smoothing kernel used for the original data, which are assumed to be independent between any two voxels before smoothing, η is the DF before the sample variance image is smoothed, and D is the dimension of the data (e.g. D = 3 for a space). Eq. (4.1 & 4.2) show that the EDF, VAR are related to the DF of the unsmoothed sample variance. They are also related to the smoothing kernels for smoothing a sample variance image and the data. If the sample variance image is not smoothed, the EDF is same as the DF of the unsmoothed variance image.

If $F_V = F_D$, then EDF $= 5.2\eta$, a large change in DF. If we know the $\eta = 14$ and we want EDF $= 100$, then $F_V = 1.16\ F_D$ is sufficient. In addition, there is no bias due to variance smoothing in this case, and the MSE is proportional to 1/EDF. We note that Eq. (4.1) is consistent with the result derived by Worsley *et.al* [Worsley *et al.*, 2002]. They smoothed the ratio of random effects variance to fixed effects variance, while we directly smoothed random effects variance. The derivation for EDF, MSE and VAR of the smoothed variance images are summarized in Appendix B.1 in detail.

Spatially heterogeneous but fixed variance images

We now consider the case where the true variance image Σ is spatially heterogeneous but fixed. $\Sigma \neq \sigma_{\mathbf{0}}^{\mathbf{2}}$, and for any two voxels, j, k in Σ, σ_j^2, σ_k^2 are fixed, where j, $k \in \mathcal{I} = \{1, 2, \ldots, N_p\}$. For a specific j^{th} voxel, V_j is the sample variance at the jth voxel. We assume the covariance between two voxels j, k is $\mathrm{Cov}(X_j, X_k | \sigma_j^2, \sigma_k^2) = \rho_{jk}\sigma_j\sigma_k$. We also assume the corresponding smoothed variance at the jth voxel is,

$$\hat{V}_j = \sum_{k=1}^{N_p} w_{jk}V_k \text{, and } \sum_{k=1}^{N_p} w_{jk} = 1$$

According to the derivation in Appendix B.2, the EDF, VAR and MSE are

$$(4.3) \qquad \mathrm{EDF} \approx \frac{(\sum_{k=1}^{N_p} w_{jk}\sigma_k^2)^2}{\frac{1}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p} w_{jk}w_{jl}\rho_{kl}^2\sigma_k^2\sigma_l^2}$$

$$(4.4) \qquad \mathrm{VAR} = \mathrm{Var}(\hat{V}_j | \sigma_1^2, \ldots, \sigma_{N_p}^2) = \frac{2}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p} w_{jk}w_{jl}\rho_{kl}^2\sigma_k^2\sigma_l^2$$

$$(4.5) \qquad \mathrm{Bias}^2 = (\mathrm{bias}_j | \sigma_1^2, \ldots, \sigma_{N_p}^2)^2 = (\sum_{k=1}^{N_p} w_{jk}\sigma_k^2 - \sigma_j^2)^2$$

And MSE $=$ VAR $+$ Bias^2. The above three equations are very complicated since the variance image is heterogeneous and fixed. We do not have closed forms for those formulas in this case.

Spatially heterogeneous and random variance images

Generally, the true variance image Σ is not always constant or fixed but random and unknown in reality, $\Sigma \neq \sigma_0^2$. We propose modeling the true variance as a scale of a χ^2 random field. We plot the empirical probability density function (PDF) and theoretical PDF of the true variance image generated with 153 subjects when $F_D = 4$mm (see section 4.2.3 below). If σ^2 denotes the true variance, we assume $\dfrac{\nu\sigma^2}{\sigma_0^2} \sim \chi_\nu^2$. We estimate $\nu = 4.498$ using the Satterthwaite approximation method and let $\sigma_0^2 = 0.3895$ be the mean of the sample variance.
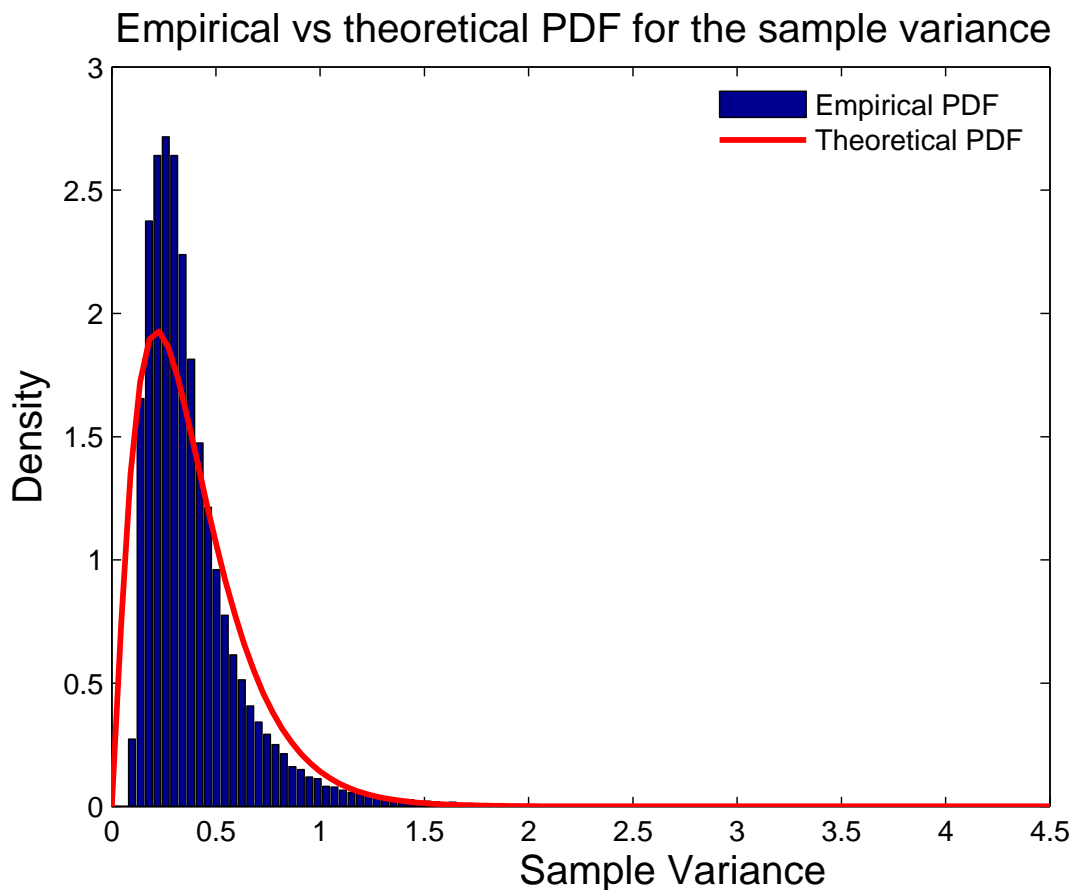
Figure 4.1: The empirical and theoretical PDF of the sample variance image generated with 153 subjects

Figure 4.1 shows that the empirical distribution of the sample variance image have a heavy right tail similar to the χ^2 distribution with certain degrees of freedom. Though

not a perfect fit, we proceed to use χ^2 random field to model the true variance image $\mathbf{\Sigma}$. For any voxels i, j, $k \in \mathcal{I} = \{1, 2, \ldots, N_p\}$, we assume that the true variance σ_i^2, σ_j^2, $\sigma_k^2 \in \mathbf{\Sigma}$ have properties,

$$\frac{\nu \cdot \sigma_i^2}{\sigma_0^2} \sim \chi_\nu^2$$

$$\text{Correlation} \left(\frac{\nu \cdot \sigma_j^2}{\sigma_0^2}, \frac{\nu \cdot \sigma_k^2}{\sigma_0^2} \right) \neq 0$$

where ν is the degrees of freedom of the χ^2 distribution. Obviously, σ_i^2 can be measured as $\dfrac{\chi_v^2 \cdot \sigma_0^2}{\nu}$. The EDF, VAR, MSE and bias^2 for this case are

$$(4.6) \quad 1/\text{EDF} \approx \frac{1}{\eta} \left[1 + 2 \left(\frac{F_V}{F_D} \right)^2 \right]^{-D/2} + \frac{2}{\eta \nu} \left[1 + 2 \left(\frac{F_V}{F_{\sigma^2}} \right)^2 + 2 \left(\frac{F_V}{F_D} \right)^2 \right]^{-D/2},$$

$$(4.7) \quad \text{VAR} = \frac{2\sigma_0^4}{\eta \nu} \left\{ \nu \left[1 + 2 \left(\frac{F_V}{F_D} \right)^2 \right]^{-D/2} + 2 \left[1 + 2 \left(\frac{F_V}{F_{\sigma^2}} \right)^2 + 2 \left(\frac{F_V}{F_D} \right)^2 \right]^{-D/2} \right\},$$

$$(4.8) \quad \text{bias}^2 = \frac{2\sigma_0^4}{\nu} \left\{ \left[1 + 2 \left(\frac{F_V}{F_{\sigma^2}} \right)^2 \right]^{-D/2} - 2 \left[1 + \left(\frac{F_V}{F_{\sigma^2}} \right)^2 \right]^{-D/2} + 1 \right\}.$$

$\text{MSE} = \text{VAR} + \text{bias}^2$, where

- F_V, the FWHM of smoothing kernels used for smoothing the sample variance images

- F_D, the FWHM smoothness of data

- F_{σ^2}, the FWHM smoothness of the true variance image $\mathbf{\Sigma}$

Eq. (4.6) shows that spatially homogeneous variance images is a specific case of spatially heterogeneous variance images since Eq. (4.1) is a part of Eq. (4.6). When $F_V = 0$, no smoothing used for sample variance images, EDF is same as the original DF and there is no bias involved. Given other parameters constant, the larger F_V used for the sample variance, the larger EDF and bias, but smaller VAR becomes. In addition, EDF is also

related to the degrees of freedom of the χ^2 random field, the data smoothness F_D, and the true variance smoothness F_{σ^2}. Eq. (4.8) shows that bias does not depend on the data smoothness nor the original DF. It also shows that bias is bounded, for example, $\text{bias}^2 \in [0 \, 2\sigma_0^2/\nu)$ when D $= 2$. Appendix B.3 provides all derivation in detail.

4.2.2 Simulation

Because of the stocastic model for the true variance, the simulation requires three levels: Realization, Experiment and Subject. We generate one true variance image in each realization, which includes multiple experiments, within which are several subjects. The simulation scheme consists of the following steps:

1. For Realization $k = 1, \ldots, n_R$, the true variance image $\boldsymbol{\Sigma_k}$ is generated by squaring, summing ν standard smoothed Gaussian images, and scaling the summation with σ_0^2/ν. Those smoothed Gaussian images are produced by Gaussian noise images smoothed with F_{σ^2} and scaled to have unit variance after smoothing. The true variance image $\boldsymbol{\Sigma_k}$ has mean σ_0^2, i.e. $\boldsymbol{\Sigma_k} \approx \dfrac{\chi_\nu^2 \cdot \sigma_0^2}{\nu}$. In this step, we also set the true variance image $\boldsymbol{\Sigma_k}$ to be constant $\boldsymbol{\sigma_0^2}$ in order to evaluate the results on the derivation for spatially homogeneous variance images case.

2. For Realization k, Experiment $j = 1, \ldots, n_E$ are created. In Experiment j, a set of n_S subjects are created by generating standard Gaussian noise images $N(0, 1)$, then smoothed with F_D and scaling each by $\sqrt{\boldsymbol{\Sigma_k}}$.

3. A one-sample model is applied to each Experiment, creating a sample variance image V_{jk}. The sample variance image is smoothed with a smoothing kernel F_V, producing the smoothed variance image $\hat{V}_{jk}$

4. A smoothed variance t statistic image is generated with the smoothed variance image $\hat{V}_{jk}$

The image dimensions are [64 64] voxels in 2D. Ideally we would make 3D images but the simulations are very memory intensive. The whole simulation starts with padded in each dimension by $2 \times F_{\sigma^2}$ to prevent kernel truncation artifacts. For example, padded image dimensions are [176 176] voxels when $F_{\sigma^2} = 28$. Those specific parameters are $n_R = 1000$, $n_E = 15$ and $n_S = 10$. The number of experiments in a realization and the number of subjects within an experiment are same as the those in real data analysis, respectively. The range of F_V is set to [0:2:24], where voxel size is [1 1]mm. $F_V = 0$ is used to check the consistency of simulation results and theory results in no smoothing status. F_D is set to 4, 6, 8, and 10 voxels, which are matched to real data analyses. F_{σ^2} are different with different F_D. For example, $F_{\sigma^2} = 20$ for $F_D = 4$ and $F_{\sigma^2} = 28$ for $F_D = 10$. The selection strategy is motivated by the results from both real data analysis and theory derivation. $\sigma_0^2 = 0.3895$, which is the average over all voxels in the sample variance image, which is generated from 153 subjects. To access the power of our method, a spherically shaped signal (radius $= 2$ voxels) with various uniform intensities (0.25, 0.5, 1) is added to the center of those standard smoothed Gaussian images. The smoothed sample variance t threshold for each α, $\alpha = 0.05$, 0.01 and 0.001, is determined by simulation when the true variance is constant and no signal is added to those standard smoothed Gaussian images, for each different combination of F_D and F_V.

This simulation framework allows Monte Carlo estimation of VAR (variance of $\hat{V}_{jk}$), bias^2 (average of $(\hat{V}_{jk} - \Sigma_{\mathbf{k}})^2$) possible. We compute bias^2 intead of bias, because we do not want a positive bias in one part of the image to cancel with a negative bias in another part. We can compare our theoretical results with those simulation results.

4.2.3 Application to real data

With a large fMRI dataset, we have $n_R = 1$ Realization from which we can generate multiple Experiments and subjects. From a 153 subjects odd-ball fMRI dataset, we use resampling without replacement method to create several sets of $n_E = 15$ Experiment,

within each has $n_S = 10$ subjects. This strategy is to mimic the Realization level in the simulation. The original data were smoothed with FWHM = 4mm. In order to compare the results for low F_D, moderate F_D and high F_D, we smooth the data to make F_D be 6, 8 and 10 mm, respectively. For each F_D, we smooth the sample variance image with F_V = [0:2:24]mm for each Experiment. Variance, bias and MSE of the smoothed variance images and EDF are measured and compared with the simulation results.

4.3   Result

4.3.1   Theory and simulation results

We only show those results for constant variance images and spatially heterogeneous random variance images cases, respectively, from simulation and derivation. In the constant variance image case, the true variance $\mathbf{\Sigma}$ is set to be constant spatially, while in the spatially random field variance image case, the true variance image is a scaled χ^2 random field.

Spatially homogeneous variance images

In this case, the true variance image is set to be a constant, $\mathbf{\Sigma} = \sigma_{\mathbf{0}}^{\mathbf{2}}$, i.e. values for all voxels are same as σ_0^2, which is obtained by averaging over all voxels in the sample variance image, which is generated with whole subjects from real data. We compare the simulation results with theoretical results to evaluate the accuracy of our derivation. Figure 4.2 and 4.3 show that the simulation and theory agree with MSE, VAR. Figure 4.2 displays the results when $F_D = 4$ voxels, while Figure 4.3 displays the results when $F_D = 6$ voxels. The two figures show that MSE drops faster at the very beginning and the trend becomes slow when F_V/F_D turns to be larger, for example, $F_V/F_D > 2$. Finally, MSE goes toward zero. In addition, the simulation and derivation also agree with bias, which is zero both in simulation and derivation results since MSE = VAR. The higher the F_V, the less MSE is. The two figures also show that the EDF from the simulation is similar to that from theory.

Figure 4.2: MSE, VAR, Bias^2 vs F_V/F_D for sample variance smoothing from simulation results and theoretical derivation for $F_D = 4$ voxel. The solid lines represent the results from simulation, while the dash lines represent the results from theory. The vertical line means the smoothing kernel size for data ($F_D = 4$ voxels) is same as the smoothing kernel size for the sample variance, while the horizontal line illustrates the degrees of freedom of the original sample variance. For example, DF $= 9$ in the simulation setup.
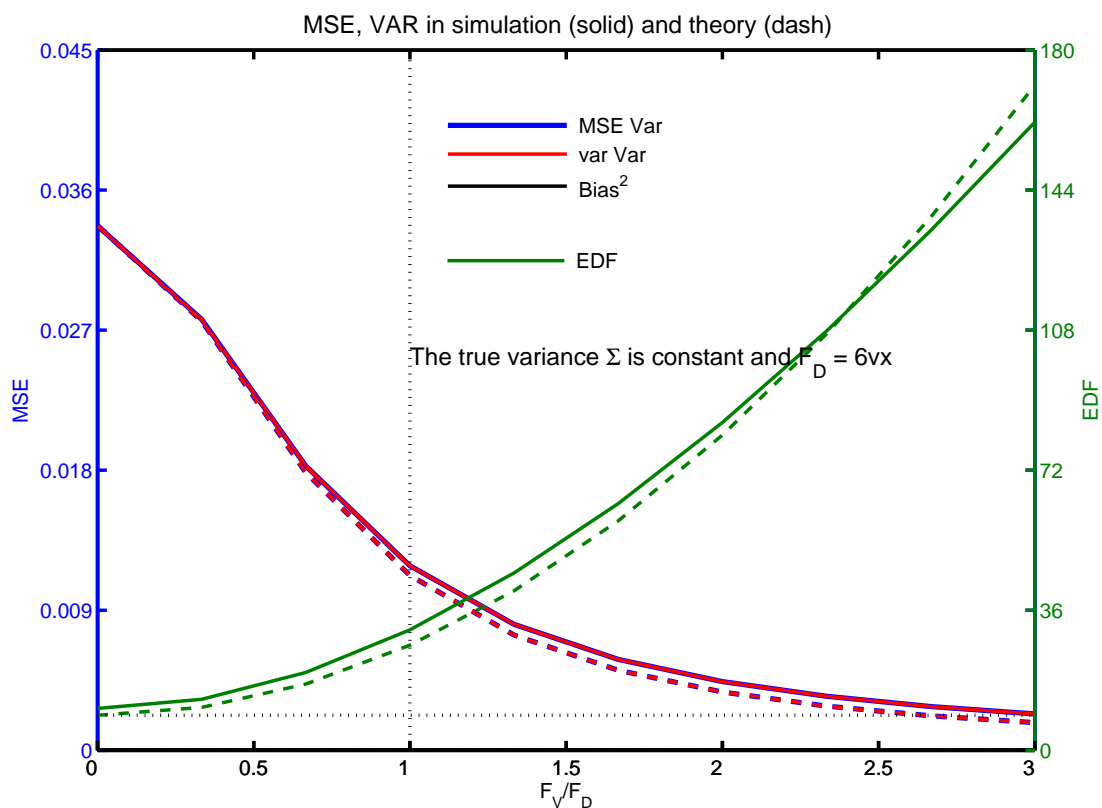
Figure 4.3: MSE, VAR, Bias^2 vs F_V/F_D for sample variance smoothing from simulation results and theoretical derivation for $F_D = 6$ voxel. The solid lines represent the results from simulation, while the dash lines represent the results from theory. The vertical line means the smoothing kernel size for data ($F_D = 6$ voxels) is same as the smoothing kernel size for the sample variance, while the horizontal line illustrates the degrees of freedom of the original sample variance. For example, DF $= 9$ in the simulation setup.

Spatially heterogeneous and random variance images

In order to compare the results from simulations and theory in this case, we set $\nu = 10$. Figure 4.4, 4.5, 4.6, 4.7, 4.8 show the MSE, VAR, Bias^2 and EDF from simulation results and theoretical derivation. The F_{σ^2} is set to be different according to different F_D since the ratio of these two plays an important role in the results. Figure 4.4 and 4.5 show that $F_{\sigma^2} = 14$ and 20, respectively, for $F_D = 4$ in order to compare different F_{σ^2}. Figure 4.6 and Figure 4.7 are for $F_D = 6$ and 8, respectively. The F_{σ^2} is set to be same for $F_D = 6$ and 8 in order to compare the difference because of different F_D. For smaller F_D, $F_D = 4$, the F_{σ^2} is generally smaller than that for larger F_D, for example, $F_D = 10$. Figure 4.8 is the results for $F_D = 10$.

Figure 4.4 and 4.5 are for $F_D = 4$ with different F_V, respectively. The curvature of MSE turns to be smaller with larger F_{σ^2}, although the starting points are same in both figures. The optimal F_V for MSE obtaining its minimum is 8 ($F_V/F_D = 2$) when $F_{\sigma^2} = 14$, but 10 ($F_V/F_D = 2.5$) when $F_{\sigma^2} = 20$. This result shows that the optimal F_V depends on the F_{σ^2} given F_D. However, there is no big difference in EDF with different F_{σ^2}.

Figure 4.6 and 4.7 show that, given F_{σ^2} and F_V, EDF is generally smaller when $F_D = 8$ than that when $F_D = 6$, although EDF turns to be larger for larger F_V/F_D. The trend for MSE goes down to a minimum point then goes up in the two figures. The optimal F_V to make MSE attain its minimum is 12 ($F_V/F_D = 2$) when $F_D = 6$, while $F_V = 14$ ($F_V/F_D = 1.75$) when $F_D = 8$, for both simulation and theoretical results. The trend for VAR always goes down and is opposite for Bias^2. This is consistent with the intuition because larger F_V will reduce variability as smoothing does, but discards information which increases bias. However, the curvature of MSE for $F_D = 6$ is slightly smaller than that for $F_D = 8$. It also shows that the starting points of VAR are always same for all F_D when $F_V = 0$ since the start number is independent of F_D when $F_V = 0$ according to Eq. (4.7). The two figures provide similar information in comparing the
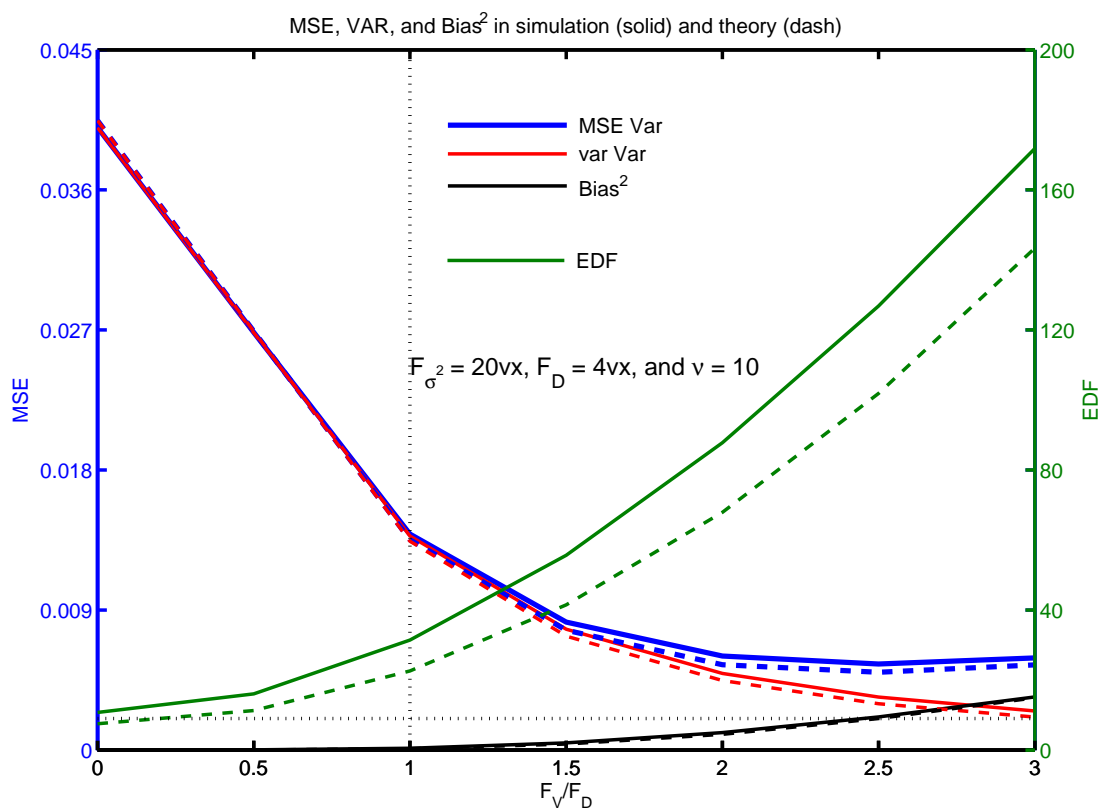
Figure 4.4: MSE, VAR, Bias^2 and EDF vs F_V/F_D for sample variance smoothing from simulation results and theoretical derivation for $F_D = 4$ voxels, $F_{\sigma^2} = 20$ voxels. The solid lines represent the results from simulation, while the dash lines represent the results from theory. The vertical line means the smoothing kernel size for data ($F_D = 4$ voxels) is same as the smoothing kernel size for the sample variance, while the horizontal line illustrates the degrees of freedom of the original sample variance. For example, DF $= 9$ in the simulation setup.

Figure 4.5: MSE, VAR, Bias^2 and EDF vs F_V/F_D for sample variance smoothing from simulation results and theoretical derivation for $F_D = 4$ voxels, $F_{\sigma^2} = 14$ voxels. The solid lines represent the results from simulation, while the dash lines represent the results from theory. The vertical line means the smoothing kernel size for data ($F_D = 4$ voxels) is same as the smoothing kernel size for the sample variance, while the horizontal line illustrates the degrees of freedom of the original sample variance. For example, DF $= 9$ in the simulation setup.

simulation results to the theoretical results, although there is slightly difference due to Monte Carlo simulation error.
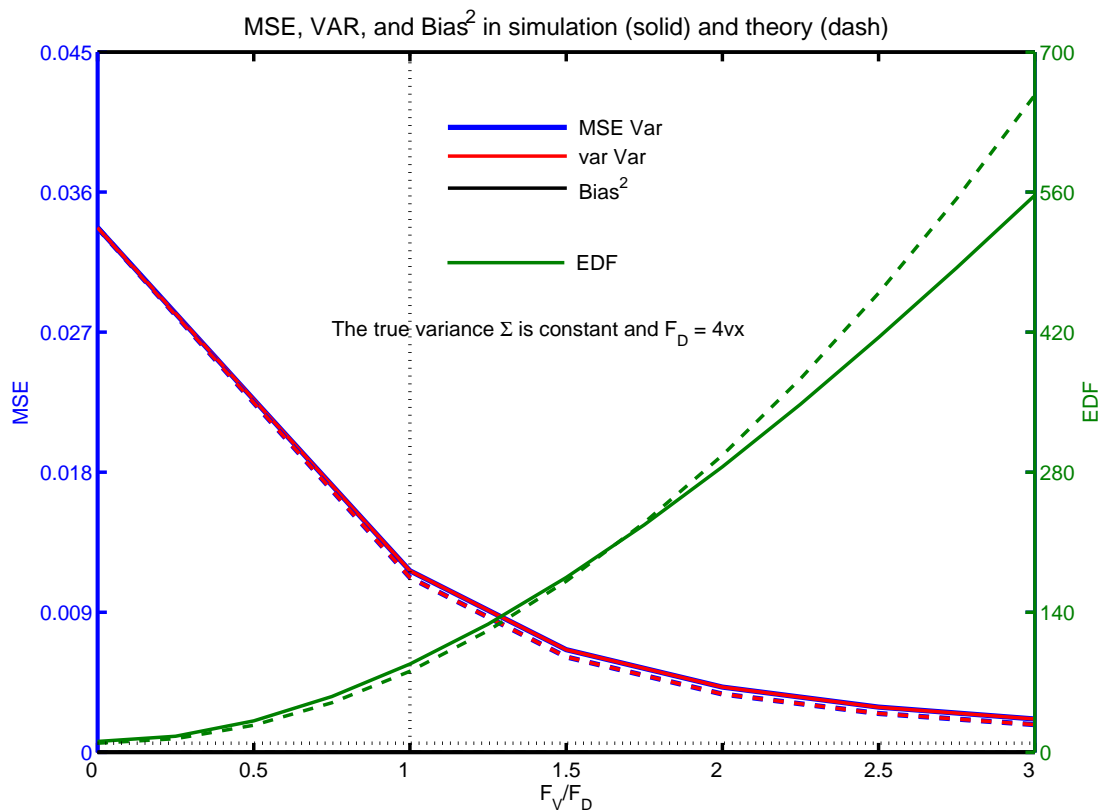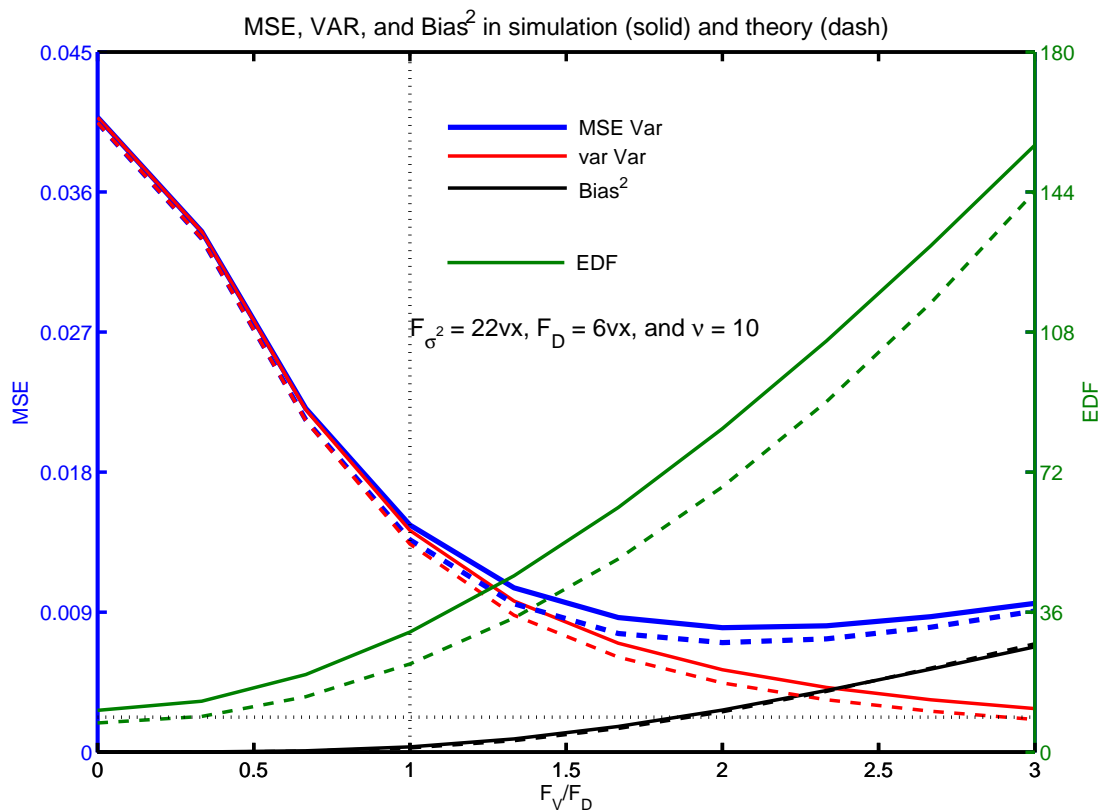

Figure 4.6: MSE, VAR, Bias^2 and EDF vs F_V/F_D for sample variance smoothing from simulation results and theoretical derivation for $F_D = 6$, $F_{\sigma^2} = 22$. The solid lines represent the results from simulation, while the dash lines represent the results from theory. The vertical line means the smoothing kernel size for data ($F_D = 6$) is same as the smoothing kernel size for the sample variance, while the horizontal line illustrates the degrees of freedom of the original sample variance. For example, DF $= 9$ in the simulation setup.

Figure 4.8 shows that EDF turns to be smaller for $F_D = 10$ than those for $F_D = 4, 6$, and 8. This phenomenon also happens for the curvature of MSE. The optimal F_V to make MSE attains its minimum in both simulation and theory is 18, which also illustrates the agreement between the simulation and theoretical results. The difference between them turns to larger for larger F_V.

We also plot the ratio of MSE from our simulation to MSE from the theory. Figure
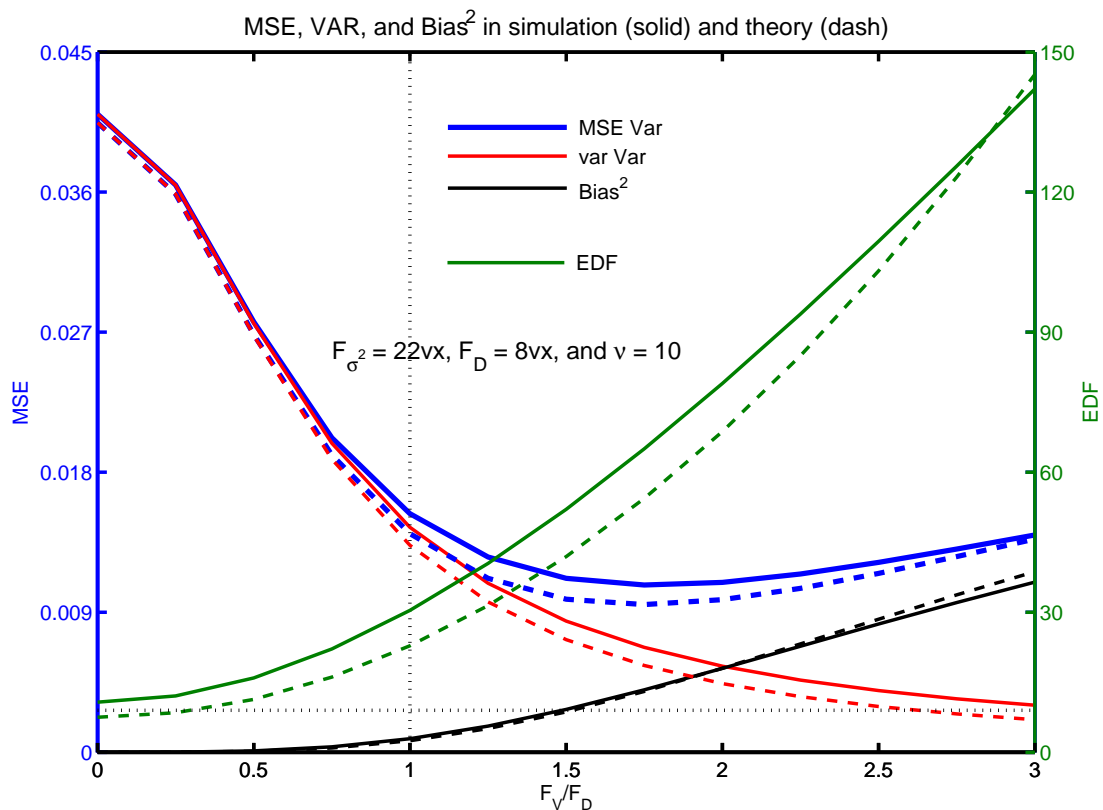
Figure 4.7: MSE, VAR, Bias^2 and EDF vs F_V/F_D for sample variance smoothing from simulation results and theoretical derivation for $F_D = 8$, $F_{\sigma^2} = 22$. The solid lines represent the results from simulation, while the dash lines represent the results from theory. The vertical line means the smoothing kernel size for data ($F_D = 8$) is same as the smoothing kernel size for the sample variance, while the horizontal line illustrates the degrees of freedom of the original sample variance. For example, DF = 9 in the simulation setup.
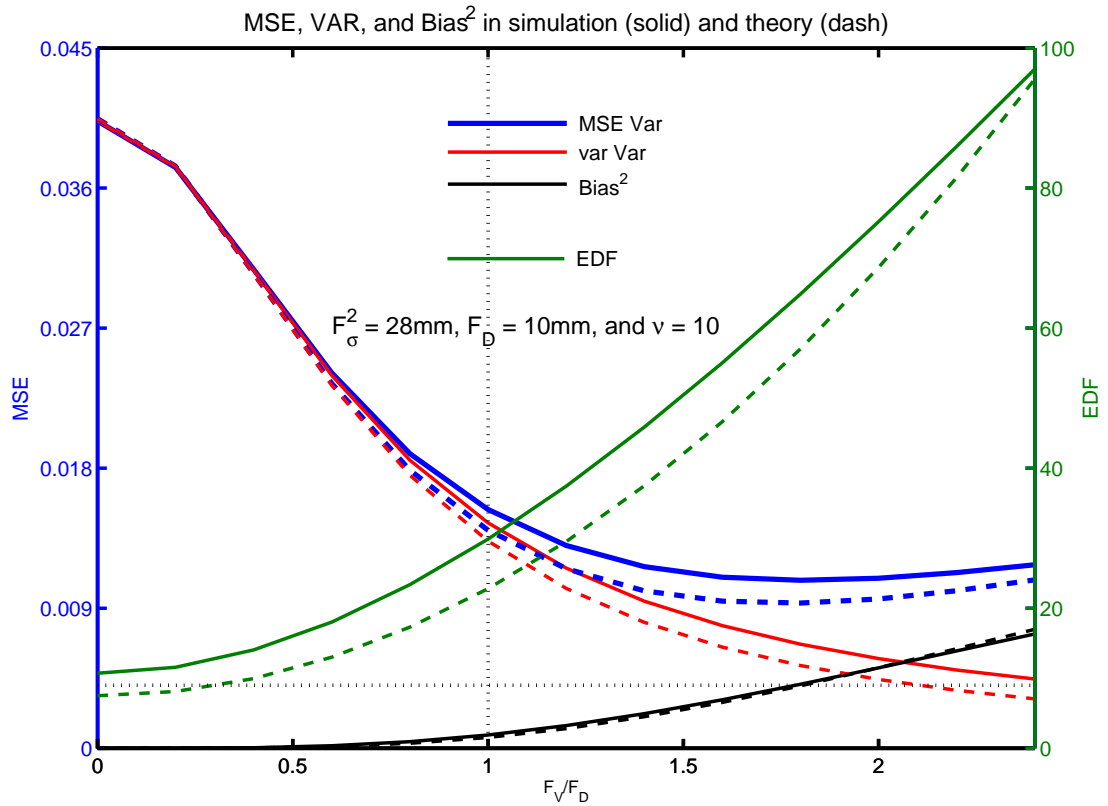
Figure 4.8: MSE, VAR, Bias^2 and EDF vs F_V/F_D for sample variance smoothing from simulation results and theoretical derivation for $F_D = 10$, $F_{\sigma^2} = 28$. The solid lines represent the results from simulation, while the dash lines represent the results from theory. The vertical line means the smoothing kernel size for data ($F_D = 8$) is same as the smoothing kernel size for the sample variance, while the horizontal line illustrates the degrees of freedom of the original sample variance. For example, DF $= 9$ in the simulation setup.

4.9 shows that the maximum of ratio is smaller with small smoothing kernel size than that with large smoothing kernel, whatever kernel size for data, the true variance image and the sample variance. In addition, the maximum shifts to right with large smoothing parameters. It is also shown that the F_V to make ratio obtain its maximum is close to the optimal F_V for MSE.
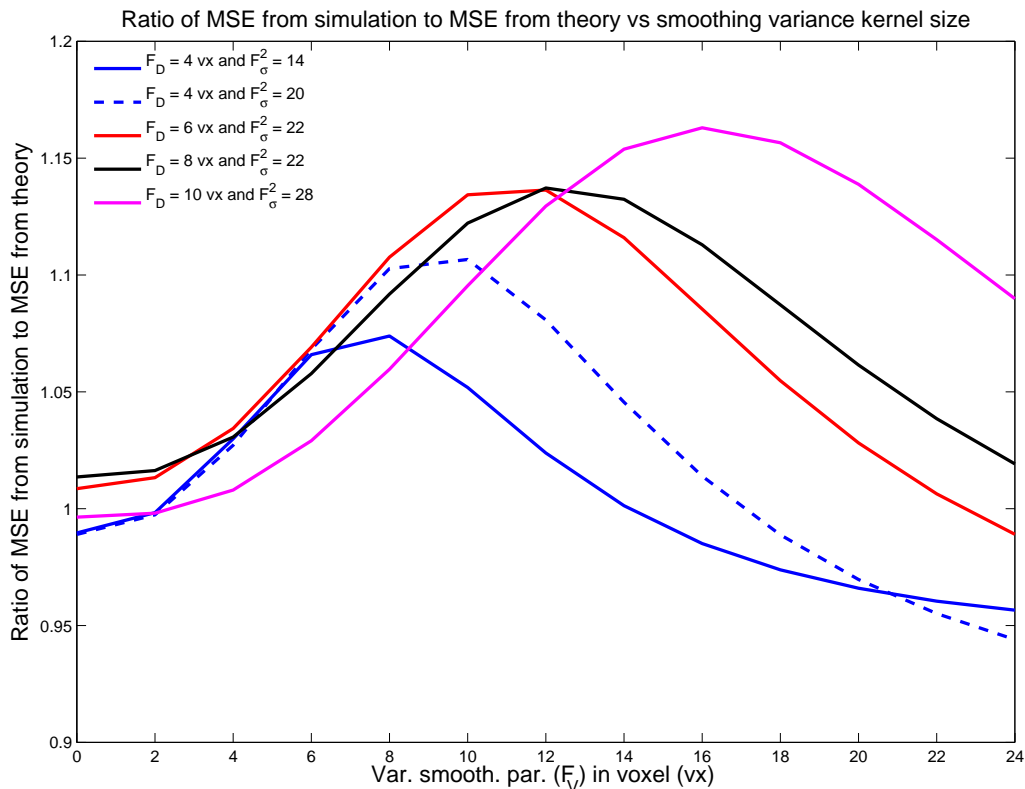
Figure 4.9: Ratio of MSE from simulation to MSE from theory vs variance smoothing kernel size

4.3.2   Application to real data

Figure 4.10, 4.11, 4.12 and 4.13 show the empirical relationship between MSE, VAR, Bias^2 and EDF of estimated variance as a function of variance smoothness parameter F_V, for $F_D = 4$, 6, 8 and 10 mm. In general, the shape of the curves is similar to the theoretical and simulation results, except at the smallest variance smoothing kernel size. For $F_D = 4$ mm, the $F_V = 10$ mm ($F_V/F_D = 2.5$) gives minimum MSE, while for

$F_D = 6$ mm, the F_V is 12 mm ($F_V/F_D = 2$) and for $F_D = 8$ mm, the F_V is 14 mm ($F_V/F_D \in (2 \quad 3)$). This result is consistent with that from simulation and theory results. Note for $F_V = 0$, EDF should be $\eta = 9$. However, the approximation is EDF $= 7.3$. This shows the inaccuracies in the Satterthwaite approximation in estimating EDF when $F_V = 0$. Moreover, the EDF from real data is much smaller than that from simulation and derivation since there is no information about the underneath true variance image of this experiment. It is unreasonable to estimate the ν, F_{σ^2} for the real data analysis.
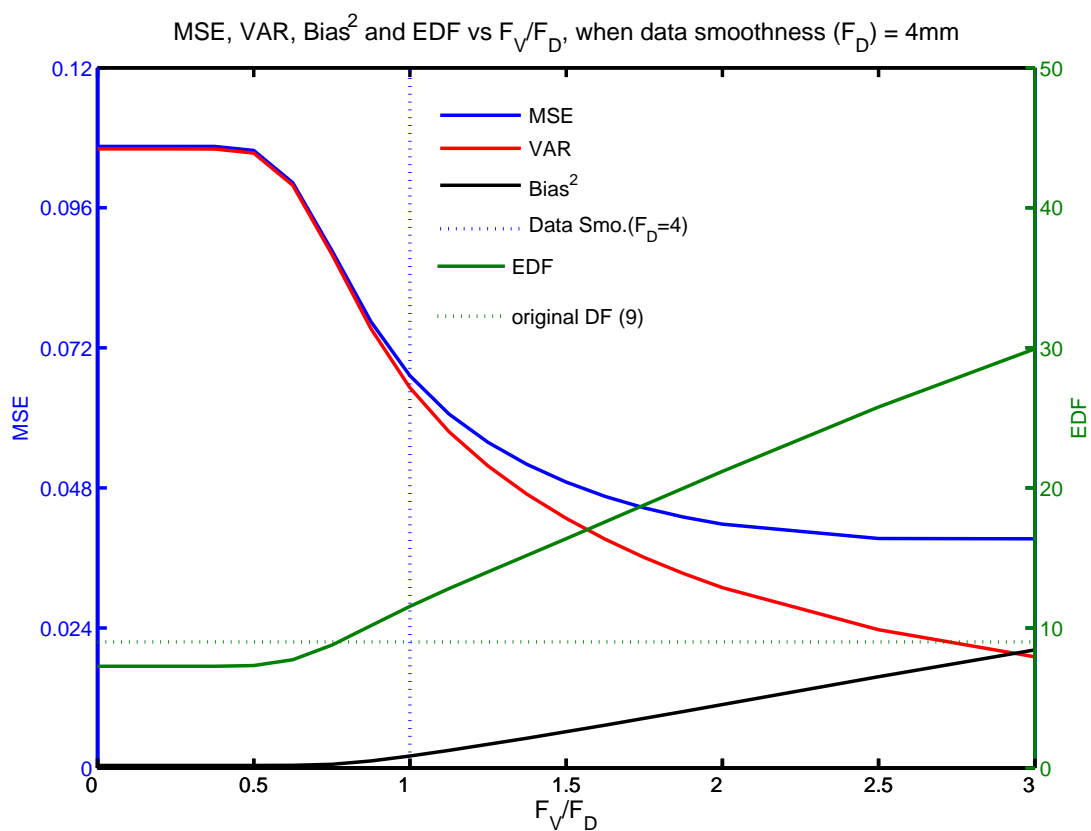
Figure 4.10: MSE, VAR, Bias^2 and EDF vs F_V/F_D for real data analysis when $F_D = 4$ mm. The vertical line means the smoothing kernel size for data ($F_D = 4$) is same as the smoothing kernel size for the sample variance, while the horizontal line illustrates the degrees of freedom of the original sample variance. For example, DF $= 9$ in real data analysis.
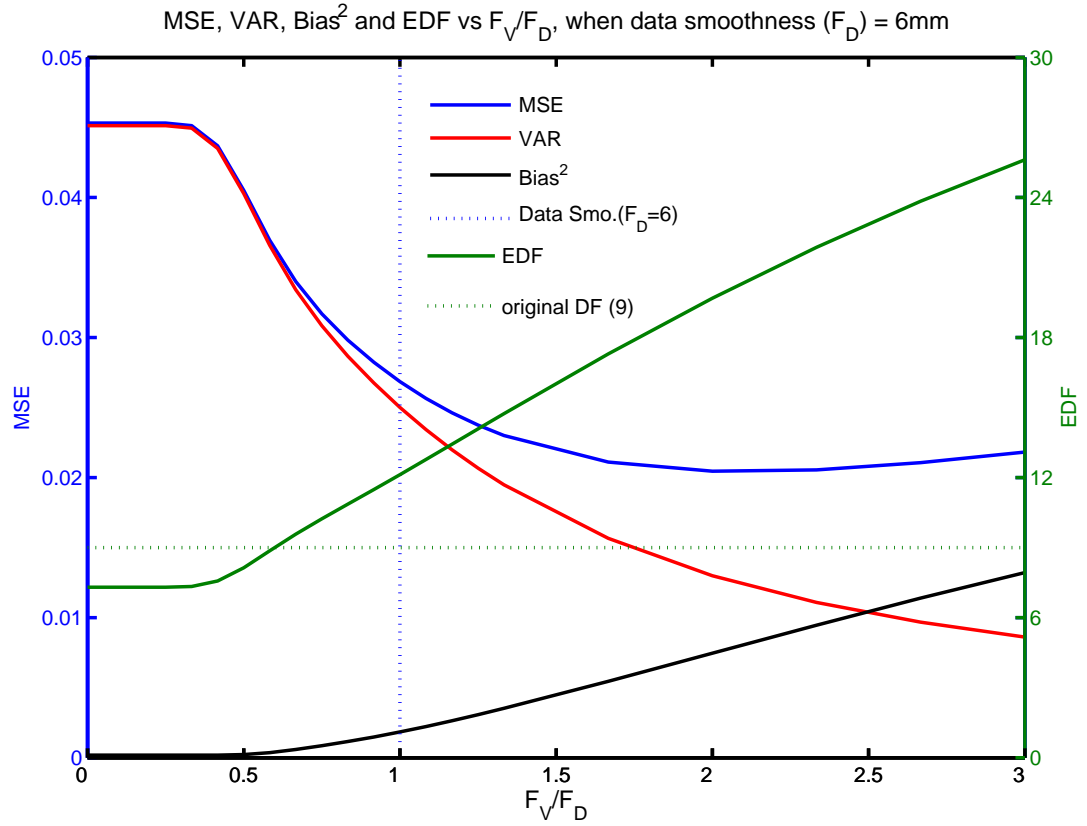
Figure 4.11: MSE, VAR, Bias^2 and EDF vs F_V/F_D for real data analysis when $F_D =$ 6 mm. The vertical line means the smoothing kernel size for data (F_D = 6) is same as the smoothing kernel size for the sample variance, while the horizontal line illustrates the degrees of freedom of the original sample variance. For example, DF = 9 in real data analysis.
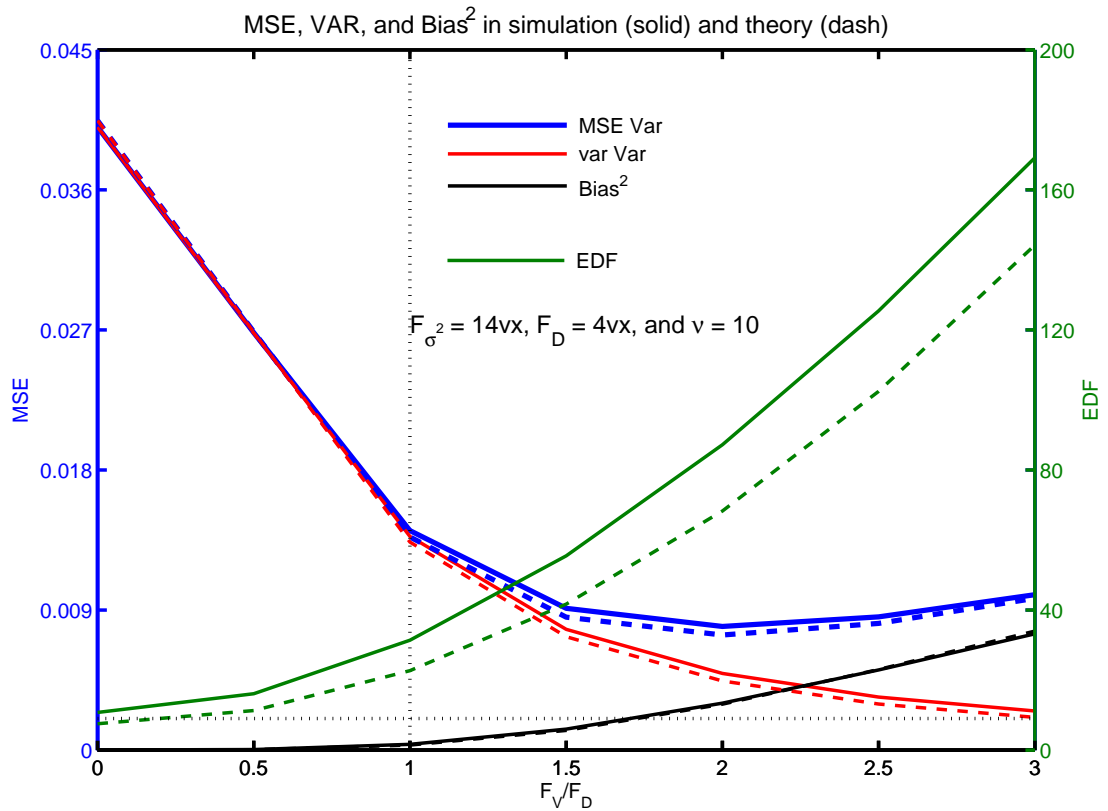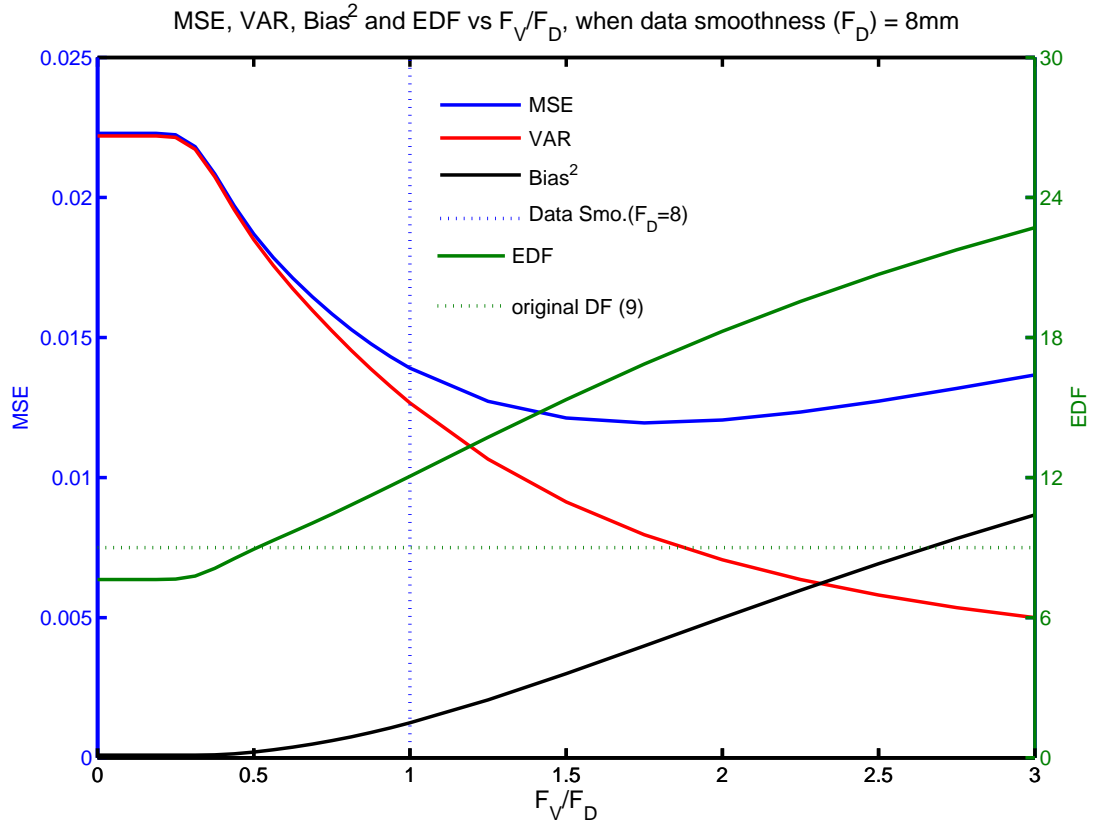
Figure 4.12: MSE, VAR, Bias^2 and EDF vs F_V/F_D for real data analysis when $F_D =$ 8 mm. The vertical line means the smoothing kernel size for data (F_D = 8) is same as the smoothing kernel size for the sample variance, while the horizontal line illustrates the degrees of freedom of the original sample variance. For example, DF = 9 in real data analysis.
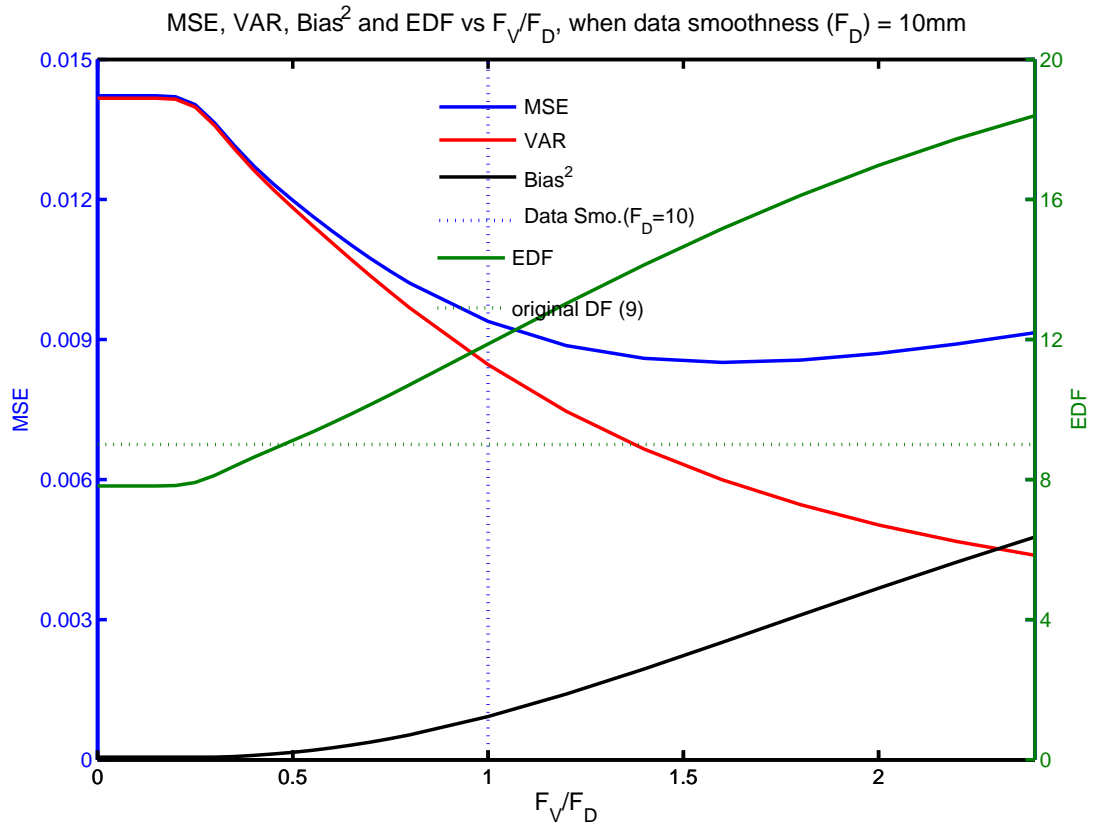
Figure 4.13: MSE, VAR, Bias^2 and EDF vs F_V/F_D for real data analysis when $F_D = 10$ mm. The vertical line means the smoothing kernel size for data ($F_D = 10$) is same as the smoothing kernel size for the sample variance, while the horizontal line illustrates the degrees of freedom of the original sample variance. For example, DF $= 9$ in real data analysis.

4.3.3   Detection sensitivity

Table 4.1 lists power changes with F_V and signal intensity when $F_D = 6$. It shows that power increases with increasing F_V, although there are slight changes when signal intensity is 0.25, $F_V = 8$ voxels and $\alpha = 0.05$ and 0.01. It also shows power increases when intensity increases, which is consistent with our expectation.

4.4   Discussion and Conclusion

Worsley *et. al.* proposed to smooth the ratio of the random effects variance to the fixed effects variance for group data analysis when the fixed effects variance has been assumed to be fixed and same in different groups [Worsley *et al.*, 2002]. Raz and Fessler also considered to smooth the variance images to increase the degrees of freedom in 1999, though their work was never published. However, they did not provide closed forms for the relationship about all kinds of smoothness parameters. Although they considered the true variance image is heterogeneous, they did not discuss whether the true variance image is fixed or random. We develop results to relate variance smoothing kernel size to MSE, VAR, bias and EDF of those smoothed variance images. We find qualitatively similar results in real data for EDF. Our simulation studies show that our derived methods are accurate under null distribution. It performs well both in constant variance case and random variance case. Our method includes the smoothness parameters for data, true variance image and the sample variance images. Although the theoretical and simulation results have similar shape curves as real data analysis, the exact values are significant different as we were not able to estimate ν and F_{σ^2}.

The theory shows that if the true variance image is constant, the EDF is related to DF, F_D, F_V and spatial dimension D. The simulation and theory also show MSE always decreases towards zero with increasing F_{σ^2}. There is no bias for the smoothed variance image in this case since MSE is always same as VAR both in simulation results and theoretical derivation when $F_V = 0$. The increase trend for EDF is smaller than that

α		0.05				0.01				0.001		
					Uniform spherical shaped signal intensity							
F_V	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0	0.0502	0.3529	0.7752	0.9921	0.0108	0.1368	0.5047	0.9494	0.0016	0.0326	0.2223	0.7889
2	0.0499	0.3598	0.7857	0.9931	0.0105	0.1463	0.5348	0.9603	0.0017	0.0412	0.2754	0.7889
4	0.0493	0.3694	0.8000	0.9945	0.0102	0.1583	0.5719	0.9722	0.0019	0.0511	0.3364	0.9048
6	0.0487	0.3746	0.8085	0.9954	0.0100	0.1635	0.5927	0.9782	0.0019	0.0558	0.3682	0.9298
8	0.0484	0.3758	0.8134	0.9958	0.0098	0.1646	0.6052	0.9813	0.0018	0.0569	0.3835	0.9423
10	0.0478	0.3752	0.8167	0.9961	0.0097	0.1633	0.6113	0.9833	0.0019	0.0570	0.3908	0.9494
12	0.0475	0.3740	0.8191	0.9964	0.0098	0.1614	0.6148	0.9846	0.0020	0.0562	0.3928	0.9544
14	0.0472	0.3722	0.8210	0.9967	0.0099	0.1592	0.6168	0.9857	0.0021	0.0554	0.3928	0.9579
16	0.0469	0.3701	0.8222	0.9969	0.0100	0.1575	0.6175	0.9867	0.0022	0.0549	0.3911	0.9608
18	0.0468	0.3684	0.8233	0.9971	0.0101	0.1562	0.6174	0.9875	0.0023	0.0546	0.3890	0.9631
20	0.0466	0.3667	0.8239	0.9972	0.0102	0.1547	0.6167	0.9882	0.0025	0.0541	0.3862	0.9651
22	0.0465	0.3652	0.8245	0.9973	0.0105	0.1532	0.6164	0.9889	0.0025	0.0541	0.3843	0.9668
24	0.0465	0.3654	0.8250	0.9974	0.0107	0.1525	0.6158	0.9894	0.0026	0.0537	0.3821	0.9682

Table 4.1: Power comparison when $F_D = 6$, $F_{\sigma^2} = 22$

when the data smoothness (F_D) turns to larger, for example, Figure 4.2 and 4.3.

When the true variance image is a random field, the EDF is related to not only DF, F_D, F_V and D, but also F_{σ^2}, ν. There is optimal F_{σ^2} to let MSE attain its minimum for small ν for random true variance image. The χ^2 random field with larger ν is much closer to a constant image, for example, $\nu = 1,000$. It is a challenge for us to find the optimal F_{σ^2} for MSE for large ν. In real data analysis, we do not have any information for ν either. However, it is reasonable to assume ν is small based on the real data analysis result in this study, for example $\nu = 10$.

Given a F_{σ^2}, EDF when F_D is large is always less than that when F_D is small, for example, EDF is less for $F_D = 8$ than that for $F_D = 6$. Given a F_D, the optimal F_V to make MSE attain its minimum when $F_{\sigma^2} = 14$ is less than that when $F_{\sigma^2} = 20$. Fortunately, there is no big difference for EDF for different F_{σ^2} with same ν. The important point is that F_{σ^2} turns to be larger for larger F_D, for example, $F_{\sigma^2} = 20$ when $F_D = 4$, and $F_{\sigma^2} = 28$ when $F_D = 10$. In addition, it is easily to show bias is bounded for D = 2, although bias always increases with larger F_V whatever how large F_D and F_{σ^2} are.

There are some assumptions for deriving those formulas. We assume that the correlation between any two voxels are expressed with Gaussian correlation function for data and the true variance. We also assume that the random true variance is a stationary χ^2 random field. Moreover, we assume all data are continuous. However, there are some limitations with the method. Although we do not know ν and F_{σ^2} before we do any data analysis, we can also use the sample variance image to estimate ν and F_{σ^2} for the true variance image since the sample variance image is an unbiased estimator of the true variance image. We use Satterthwaite method to estimate ν. When we use the sample variance image generated with 153 subjects, $\hat{\nu}$ is 4.498. Figure 4.1 shows that the empirical distribution of the sample variance image is similar to the shape of the theoretical distribution, which is from χ^2 distribution with ν degrees of freedom. In

addition, the sample variance is unbiased estimator of the true variance. Kiebel et al., Forman et al., Jenkinson and Hayasaka et al. used the smoothness estimation based on standard residual images for real data [Kiebel *et al.*, 1999; Forman *et al.*, 1995; Jenkinson, 2000; Hayasaka & Nichols, 2003]. In order to estimate F_{σ^2} for the real data, we propose to use the following strategy. Let $X = \dfrac{\hat{\nu}\hat{\sigma}^2}{\sigma_0^2} \sim \chi_{\hat{\nu}}^2$ and then standardize with $X_Z = \dfrac{\left(\frac{\hat{\nu}\hat{\sigma}^2}{\sigma_0^2} - \hat{\nu}\right)}{\sqrt{2\hat{\nu}}}$. We estimate the smoothness parameter, $\hat{F}_{\chi^2}$ with X_Z with smoothness estimator function in the SPM2 package, modified to work with a single standardized image.

As $F_{\sigma^2} = \sqrt{2}F_{\chi^2}$ according to Appendix B.3.1, we can then obtain an estimate of F_{σ^2}. Further work is needed to create estimators for F_{σ^2} in detail, and compute P value bias.

4.5 Acknowledgments

Chapter 5

Conclusion & Future Work

In this dissertation, we introduced new modeling and inference methods for fMRI data. The first topic was to propose a new combining method with cluster extent and suprathreshold average voxel intensity statistics with the permutation testing framework. This method was used to leverage the strength from the cluster wise inference method and the voxel wise inference method. However, there were some limitations with this nonparametric permutation framework. We proposed a parametric cluster mass inference method as a complementary method to nonparametric competitors. This parametric method showed increased sensitivity for sufficiently smoothed images. Lastly, we introduced a method to increase the effective degrees of freedom of variance images for smoothed variance t-test while avoiding accessive bias. We used a χ^2 distribution to model the true variance images for small group data analysis.

Since cluster mass is defined as the integral of suprathreshold intensities within a cluster, cluster mass can also be considered as the "product" of cluster extent and suprathreshold average intensity. We proposed some combining functions with the two statistics. Our simulations and real data analysis showed that no single method was optimal for all types of signals, however the partial methods (cluster extent specifically) varied considerably in their sensitivity, while the combining methods were usually not far from optional.

There were some limitations with this method. Computational complexity was a concern of this method because of permutation property for large number of subjects in

an experiment. This method could not be applied for single subject data analysis since it violated the exchangeability property. The voxels could not be permuted, otherwise, the spatial structure of brain would be affected.

The proposed parametric cluster mass inference method was a complementary method to its nonparametric competitor and all kinds of permutation based combining methods. This proposed method was based on Gaussian RFT. Our simulation studies showed that our derived null distribution was accurate, and performed well not only for Gaussian images, but also for Student's t statistic images. Like other RFT methods, our results depended only on the smoothness and the volume of the image. This method assumed that the shape of a cluster above a certain threshold was approximated by a paraboloid. A Gaussian shaped correlation function was assumed to simplify the variance in the derivation. In addition, we also assumed that all fields were stationary.

Finally, we had developed results to relate variance smoothing kernel size to variance estimator MSE, bias^2 and DF, finding qualitatively similar results in real data for EDF. Our simulation studies showed that our derived methods were accurate under the null hypothesis. It performed well both in the constant variance case and in the random variance case. The effective degrees of freedom increased as smoothness parameter of the sample variance images and increased the detection sensitivity. We could also find the optimal smoothing kernel size for the sample variance image that avoids too much bias. However, it was hard to find the optimal smoothness parameter for the sample variance images when the true variance image was constant. The ideal method is to use the pooled variance image.

We assumed that the correlation between any two voxels are expressed by a Gaussian correlation function for data and the true variance image. We also assume that all images are stationary and continuous. However, we did not know the degrees of freedom of the χ^2 distribution and the smoothness parameter of the true variance images. We only used 2D data (64×64) because if excess memory requirements needed in the simulation. In

addition, we considered a hot spot case in the derivation. However, this case was not included in the simulation.

We only used sphere-shaped signals with uniform intensity in the simulation in evaluating the characteristics in this dissertation. Therefore, we are unable to validate the methods with different shaped signals. In addition, we assumed the images were stationary based on RFT. Currently, there is no method to examine whether an image is stationary or not. It is intuitive that a nonparametric method is applied avoid stationarity assumptions in developing inference methods.

Although we developed a method for optimizing the smoothing kernel size for the sample variance image, it was hard to find the degrees of freedom of the true variance image. It was also hard to find the smoothing kernel size for the true variance image. We proposed to use Satterthwaite method to the sample variance to estimate the degrees of freedom of the true variance image. We also proposed to use smoothness estimation for the component field to approximate the smoothness kernel size for the true variance image. However, more accurate estimation for these two parameters is needed in the future.

Appendices

Appendix A

Derivation for the parametric cluster mass distribution

A.1 Cluster P values

We use nonparametric permutation to obtain uncorrected and FWE corrected cluster mass P-values on real data to provide a comparison for our proposed parametric mass statistic. As most neuroimaging permutation literature focuses on voxel-wise inference, we briefly review nonparametric cluster inference.

A nonparametric uncorrected P-value for a single voxel is trivial, as it is just the direct application of a univariate permutation test. Defining an uncorrected P-value for clusters, however, is difficult as there is no unique way to define equivalent clusters after permutation of the data. If there are L cluster in the original statistic image, in a permuted-data statistic image there will rarely be L clusters and there will almost never be a cluster in exactly the same location. Instead of matching clusters between permutations, an assumption of stationarity is made. The distribution of cluster statistics (e.g. size, mass, local peak height, etc) does not vary across space. With such a stationarity assumption, cluster statistics can be pooled over space, and a pooled permutation distribution created. While permutation distributions typically containing K elements, where K is the number of permutations, the uncorrected cluster permutation distribution will contain $\sum_{k=1}^{K} L_k$ elements, where L_k is the number of clusters found in

permuting k's statistic image. The uncorrected P-value is the proportion of the $\sum_{k=1}^{K} L_k$ elements that are as large or larger than the observed cluster statistic.

FWE corrected cluster P-values are more straightforward, and only require creating the maximal cluster statistic distribution. Because the search over the image is for the maximal statistic, no assumption of stationarity is required. Even when some regions of the image that are smoother (or, by chance, give rise to larger cluster statistics) the maximum operator naturally accounts for such variation. (Nonstationarity *is* a problem for parametric cluster inference, though see [Hayasaka *et al.*, 2004]). For each permutation the maximal cluster statistic is recorded, and the corrected P-value is the proportion of the K maximal elements that are as large or larger than the observed cluster statistic.

Lastly, we note that if cluster statistics are marked as significant only when FWE-significant at 0.05, there is then 95% confidence of no false positive clusters anywhere in the image. For more on FWE see the paper of Nichols and Hayasaka (2003) [Nichols & Hayasaka, 2003].

A.2 Derivation of Null Distribution of Cluster Mass

Our derivation of the distribution of cluster mass follows that of Poline *et al.* [Poline *et al.*, 1997] (PWEF) with several departures. A rough outline of the derivation is as follows:

1. A second order Taylor series approximates the statistic image at a local maximum as a paraboloid, determined by peak height and curvature about the maximum.

2. The geometry of a paraboloid gives cluster extent and mass as a function of peak height and the curvature (Jacobian determinant).

3. Distribution of the curvature, conditional on peak height, is found using an assumption of a Gaussian autocorrelation function.

4. Combining the two previous results relates extent and mass, conditional on peak height, to a χ^2 distribution. A bias correction is made using the expected Euler characteristic.

5. At this point PWEF used a small excursion assumption; we produce a pair of results, with and without this assumption.

6. The joint distribution of mass and height are found and marginalized to produce final mass result.

A.2.1 Notation & Preliminaries

Let $\mathbf{Z}(\mathbf{x})$ be a D-dimensional Gaussian image, with

$$
\begin{aligned}
\mathrm{E}(\mathbf{Z}(\mathbf{x})) &= 0, \\
\mathrm{Cov}(\mathbf{Z}(\mathbf{x})) &= 1, \\
\mathrm{Cov}(\nabla \mathbf{Z}(\mathbf{x})) &= \Lambda
\end{aligned}
$$

for all $\mathbf{x} \in \Omega \subset \Re^D$ in the image volume, where ∇ is the gradient operator and Λ is the $D \times D$ matrix which parameterizes roughness. We assume the process is smooth, in that $\nabla^2 \rho(\mathbf{0})$ exists, where $\rho(\cdot)$ is the autocorrelation function and ∇^2 is the Hessian operator.

Without loss of generality, suppose there exists a local maximum at $\mathbf{x} = \mathbf{0}$, and consider the approximating paraboloid from a second order Taylor series about $\mathbf{x} = \mathbf{0}$

$$
\mathbf{W}(\mathbf{x}) = \mathbf{Z}(\mathbf{0}) + \mathbf{x}^{\mathbf{T}}(\nabla^2 \mathbf{Z}(\mathbf{0}))\mathbf{x}/2
$$

Suppressing the spatial index, let $Z \equiv \mathbf{Z}(\mathbf{0})$, and denote $J = |-\nabla^2 \mathbf{Z}(\mathbf{0})|$ the negative Jacobian determinant.

For a cluster-defining threshold u_c, let $H = Z - u_c$ be the suprathreshold magnitude (note that we suppress the ℓ subscript used in the body of this chapter). Then the

geometry of the approximating paraboloid gives cluster extent as

$$(A.1) \qquad\qquad S = a2^{D/2}H^{D/2}J^{-1/2}$$

where $a = \pi^{D/2}/\Gamma(D/2+1)$ is the volume of the unit sphere, and mass as

$$(A.2) \qquad\qquad M = 2SH/(D+2).$$

A.2.2 Distribution of $S|H$

Conditional on H, PWEF showed that another Taylor series yields

$$(A.3) \qquad\qquad \ln J|H \approx \ln|\Lambda| + D\ln(H+u_c) + \eta,$$

where η is mean zero Gaussian with variance[1]

$$\mathrm{Cov}(\eta|H) = \left[\mathrm{tr}\left((\Lambda^{-1}\otimes\Lambda^{-1})\rho^{(4)}(\mathbf{0})\right) - D^2\right]/(H+u_c).$$

While this expression is quite involved, if we assume that ρ is proportional to a Gaussian probability density function (PDF), it simplifies to $\mathrm{Cov}(\eta|Z) = 2D/(H+u_c)^2$. Subsequently we will need $J^{-1/2}$, and so write the exponentiated and powered equation (A.3) as $J^{-1/2} \approx |\Lambda|^{-1/2}(H+u_c)^{-D/2}\exp(\eta/2)^{-1}$. However, as in PWEF, we find that numerical evaluations of the final result are poor when η is assumed to be Gaussian (results not shown). We instead linearize the exponential,

$$(A.4) \qquad\qquad J^{-1/2}|H \;\approx\; |\Lambda|^{-1/2}(H+u_c)^{-D/2}(1+\eta/2)^{-1}$$

and approximate $1+\eta/2$ with η', where $\nu\eta'$ is χ^2_ν variate. Matching the second moments of $1+\eta/2$ and η' gives $\nu = 4(H+u_c)^2/D$. Combining with Equations (A.4) and (A.1) yields

$$(A.5) \qquad\qquad S|H \approx a2^{D/2}|\Lambda|^{-1/2}(H+u_c)^{-D/2}H^{D/2}\eta'^{-1}.$$

---

[1]Note there is a typo in the PWEF paper's equation (8), where $2Z$ should in fact be just Z, or $H+u_c$ as we have written.

A.2.3 $\mathcal{U}$ Result for M

PWEF proceeded by using a small excursion approximation, that H is small relative to u_c, replacing $H + u_c$ with u_c. With this change, and marginalizing out H, the expected cluster extent can be found as

$$(A.6) \qquad\qquad \mathrm{E}_\mathcal{U}(S) = (2\pi)^{D/2}\, |\Lambda|^{-1/2}\, u_c^{-D}.$$

However, accurate results using the expected Euler Characteristic [Adler, 1981] give

$$(A.7) \qquad \mathrm{E}_{\mathrm{EC}}(S) = (2\pi)^{D/2}\, |\Lambda|^{-1/2}\, u_c^{-(D-1)}\, (1 - \Phi(u_c))/\phi(u_c)$$

where Φ is the standard Gaussian CDF and ϕ is the standard Gaussian PDF. Hence, the approximation for $S|H$ is scaled by

$$(A.8) \qquad\qquad c_\mathcal{U} = \frac{\mathrm{E}_{\mathrm{EC}}(S)}{\mathrm{E}_\mathcal{U}(S)} = u_c(1 - \Phi(u_c))/\phi(u_c).$$

As a side note, this is Mill's ratio [Gordon, 1941] scaled by u_c, which will have $c_\mathcal{U}$ converging to 1 from below for large u_c.

The bias-adjusted result is

$$(A.9) \qquad M|H \approx a c_\mathcal{U} 2^{D/2+1}(D+2)^{-1}\, |\Lambda|^{-1/2}\, u_c^{-D/2}\, H^{D/2+1}\, \eta'^{-1},$$

which is a scaled inverse χ^2 random variable with $\nu_\mathcal{U} = 4u_c^2/D$ degrees of freedom and scale parameter

$$q_\mathcal{U}(H) = a c_\mathcal{U} 2^{D/2+1}(D+2)^{-1}\, |\Lambda|^{-1/2}\, u_c^{-D/2}\, H^{D/2+1}$$

The marginal distribution of H is approximately exponential with mean $1/u_c$ [Adler, 1981], and thus the joint PDF of M and H is

$$(A.10) \qquad f_\mathcal{U}(M, H) \approx \frac{(q_\mathcal{U}(H)\nu_\mathcal{U}/2)^{\nu_\mathcal{U}/2}}{\Gamma(\nu/2)} \frac{\exp\left[(q_\mathcal{U}(H)\nu_\mathcal{U}/2)/M\right]}{M^{\nu_\mathcal{U}/2+1}} \exp[-u_c H] u_c$$

for $M, H > 0$. The uncorrected P-value for cluster mass is then found with

$$P_\mathcal{U}(M > m) \approx \int_m^\infty \int_0^\infty f_\mathcal{U}(M, H)\, dH\, dM$$

using numerical integration over a fine grid.

A.2.4 $\mathcal{Z}$ Result for M

We repeat the preceding without the small excursion approximation. We call this the $\mathcal{Z}$ result, since $Z = H + u_c$ is left as is. Returning to (A.5) and marginalizing out H we get

(A.11) $$\mathrm{E}_{\mathcal{Z}}(S) = a2^{D/2} \, |\Lambda|^{-1/2} \, \mathrm{E}\{[H/(H + u_c)]^{D/2}\}$$

where the final term must be found numerically for a particular u_c. This provides the bias adjustment term

(A.12) $$c_{\mathcal{Z}} = \mathrm{E}_{\mathrm{EC}}(S)/\mathrm{E}_{\mathcal{Z}}(S).$$

This provides an approximation for $M|H$ as a scaled inverse χ^2 random variable with ν degrees of freedom and scale parameter

$$q_{\mathcal{Z}}(H) = ac_{\mathcal{Z}}2^{D/2+1}(D + 2)^{-1} \, |\Lambda|^{-1/2} \, (H + u_c)^{-D/2} \, H^{D/2+1}$$

and joint PDF of M and H of

(A.13) $$f_{\mathcal{Z}}(M, H) \approx \frac{(q_{\mathcal{Z}}(H)\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \frac{\exp\left[(q_{\mathcal{Z}}(H)\nu/2)/M\right]}{M^{\nu/2+1}} \exp[-u_c H] u_c.$$

As before, the uncorrected P-value for cluster mass is then found with

$$P_{\mathcal{Z}}(M > m) \approx \int_m^\infty \int_0^\infty f_{\mathcal{Z}}(M, H) \, dH \, dM$$

using numerical integration over a fine grid.

A.2.5   Corrected P-values

The uncorrected P-values can be transformed into family-wise error (FWE) corrected P-values with either a Bonferroni correction for the expected number of clusters or the Poisson clumping heuristic [Adler, 1981; Cao & Worsley, 2001; Hayasaka & Nichols, 2003]. We opt for the later, as it provides a continuous transformation between uncorrected and corrected P-values.

A FWE corrected P-value accounts for the chance of the maximal statistic exceeding that actually observed. Assuming the clusters arise as a Poisson process, this P-value is found as

$$\text{(A.14)} \qquad P(\max_{\ell} M_{\ell} > m) \approx 1 - \exp\{-E(L) \cdot P(M_{\ell} > m)\},$$

where $E(L)$ is the expected number of clusters in the image. For moderate thresholds u_c the Euler characteristic will count the number of clusters, and hence we approximate $E(L)$ by $E_{EC}(L)$. The most accurate results for $E_{EC}(L)$ depends on the dimension and the topology of the search region [Worsley $et\ al.$, 1996]. For a 3D, approximately spherical search region

$$\text{(A.15)} \qquad E_{EC}(L) = \lambda(\Omega)|\Lambda|^{1/2}(2\pi)^{-2}(u_c^2 - 1)\exp[-u_c^2/2];$$

where $\lambda(\Omega)$ is the volume of the search region. In addition, for a high threshold u_c, the number of clusters above the threshold will be approximated by [Adler, 1981; Poline $et\ al.$, 1997]

$$E(L) = \lambda(\Omega)|\Lambda|^{1/2}(2\pi)^{-2}u_c^2\exp[-u_c^2/2]$$

A.2.6   Smoothness Estimation & Λ

The preceding results depend on the roughness of the component random fields, as parameterized by $|\Lambda|$. Worsley et al. [Worsley $et\ al.$, 1992] proposed re-expressing this as the FWHM Gaussian kernel required to smooth an independent random field into one with roughness Λ. Assuming the smoothing is aligned with the major axes of the image, this relationship is

$$|\Lambda|^{1/2} = \frac{(4\ln 2)^{D/2}}{\prod_d \text{FWHM}_d}$$

where FWHM_d is the smoothness in the d-th dimension. If the smoothness is not known, $|\Lambda|^{1/2}$ can be estimated from the residual images of a general linear model [Kiebel $et\ al.$, 1999]

A.2.7 Student's *t-image*

Worsley et al. and Holmes [Kiebel *et al.*, 1999] showed that if the roughness of the Gaussian component fields is Λ, the roughness for a Student's t-statistic image can be approximated by $\Lambda_T = \lambda_n \Lambda$, where $n > 4$ is the number of scans used to generate the t image and λ_n is the correction factor [Holmes, 1994; Worsley *et al.*, 1992]. When applying our method to Gaussianized data we adjust Λ accordingly.

Appendix B

Derivation for effective degrees of freedom

B.1 Spatially homogeneous variance images

For the spatial homogeneous variance image case, we assume the true variance image Σ is constant and fixed, i.e. $\sigma_j^2 = \sigma_0^2$, and independent across the entire image, where $j = 1, \cdots, N_p$ is the index of the jth voxel, N_p is the number of voxels, and $\Sigma = (\sigma_1^2, \sigma_2^2, \cdots \sigma_{N_p}^2)$. At the jth voxel, conditional on σ_0^2, we denote data by X and assume,

$$X_{ji}|\sigma_0^2 \quad \sim \quad N(\mu_j, \sigma_0^2)$$

where,

$j = 1, 2, \cdots, N_p$ indexes voxels of an image

$i = 1, 2, \cdots, \eta$ indexes subjects in an experiment

Under the null hypothesis, $\mu_j = 0$, the data conditional on σ_0^2 are also considered as white noise convolved with a Gaussian kernel $f(\mathbf{u})$ with mean zero and variance $s_d^2 = F_D^2/(8 \log 2)$, F_D is the Full Width Half Maximum of smoothing kernel from the data, or the smoothness of the standardized subject-level conditional on Σ i.e.

$$f(\mathbf{u}) = \phi(\mathbf{u}; s_d^2)$$

where,

$$\phi(\mathbf{u}; s_d^2) = (2\pi s_d^2)^{-D/2} e^{-\|\mathbf{u}\|^2/(2s_d^2)}$$

is a Gaussian density function with mean zero and variance s_d^2 in D dimension.

Hence the spatial correlation is

$$c(\mathbf{u}) = \frac{(f \otimes f)(\mathbf{u})}{(f \otimes f)(\mathbf{0})} = e^{-\|\mathbf{u}\|^2/(4s_d^2)} = (2\pi 2s_d^2)^{D/2}\phi(\mathbf{u}; 2s_d^2)$$

We can also write correlation function between the kth and l^{th} voxels as

$$\rho_{kl} = g(\mathbf{u}_{kl}, s_d^2) = \exp\{-\|\mathbf{u}_{kl}\|_2^2/(4s_d^2)\} = \exp\{-\|\mathbf{P}_k - \mathbf{P}_l\|_2^2/4s_d^2\}^1$$

$$\mathbf{P}_k = (x_k, y_k, z_k)^t \text{ denotes the coordinate of the } k^{th} \text{ voxel,}$$

Hence,

$$\text{Cov}(X_{km}, X_{ln}|\sigma_0^2) = \rho_{kl}\sigma_0^2\delta_{m-n}$$

$$\delta_{m-n} = \begin{cases} 1 & m = n \text{ , same subject} \\ 0 & \text{otherwise} \end{cases}$$

The sample variance for the jth voxel is

$$V_j = \frac{1}{\eta}\sum_{i=1}^{\eta}(X_{ji})^2$$

Then,

$$
\begin{aligned}
E(V_j|\sigma_0^2) &= E(\frac{1}{\eta}\sum_{i=1}^{\eta}(X_{ji})^2|\sigma_0^2) \\
&= \frac{1}{\eta}\sum_{i=1}^{\eta}E\left[X_{ji}^2|\sigma_0^2\right] \\
&= \frac{1}{\eta}\sum_{i=1}^{\eta}\sigma_0^2 \\
&= \sigma_0^2
\end{aligned}
$$

---

[1] $\|X\|_2 = \|X\|$ is l_2 norm, and $\|X\|_2^2$ denotes the distance between two voxels

$$
\begin{aligned}
\text{Var}(V_j|\sigma_0^2) &= E(V_j^2|\sigma_0^2) - \left[E(V_j|\sigma_0^2)\right]^2 \\
&= \frac{1}{\eta^2}E\left(\sum_{m=1}^{\eta}X_{jm}^2\sum_{n=1}^{\eta}X_{jn}^2 \mid \sigma_0^2\right) - (\sigma_0^2)^2 \\
&= \frac{1}{\eta^2}\{\sum_{m=1}^{\eta}E[X_{jm}^4|\sigma_0^2] + \sum_{m\neq n;m,n=1}^{\eta}E[X_{jm}^2X_{jn}^2|\sigma_0^2]\} - (\sigma_0^2)^2 \\
&= \frac{1}{\eta^2}(\eta\cdot 3\sigma_0^4 + \eta(\eta-1)\sigma_0^4) - \sigma_0^4 \\
&= \frac{2\sigma_0^4}{\eta}
\end{aligned}
$$

$$
\begin{aligned}
\text{Cov}(V_k, V_l|\sigma_0^2) &= \frac{1}{\eta^2}\sum_{m=1}^{\eta}\sum_{n=1}^{\eta}\text{Cov}(X_{km}^2, X_{ln}^2|\sigma_0^2) \\
&= \frac{1}{\eta^2}\sum_{m=1}^{\eta}[E(X_{km}^2X_{lm}^2|\sigma_0^2) - \sigma_0^4]
\end{aligned}
$$

Because

$$
X_{km}|\sigma_0^2 \sim N(0, \sigma_0^2)
$$
$$
X_{lm}|\sigma_0^2 \sim N(0, \sigma_0^2)
$$

and $\quad X_{km}|\sigma_0^2, \ X_{lm}|\sigma_0^2$ follow bivariate normal distribution, which means,

$$
X_{km}|X_{lm} \sim N(\rho_{kl}X_{lm}, \sigma_0^2(1 - \rho_{kl}^2))
$$

Hence,

$$
\begin{aligned}
\text{Cov}(V_k, V_l|\sigma_0^2) &= \frac{1}{\eta^2}\sum_{m=1}^{\eta}(E\{E[X_{km}^2X_{lm}^2|X_{lm}]|\sigma_0^2\} - \sigma_0^4) \\
&= \frac{1}{\eta^2}\sum_{m=1}^{\eta}\{E[\rho_{kl}^2X_{lm}^4 + \sigma_0^2(1 - \rho_{kl}^2)X_{lm}^2|\sigma_0^2] - \sigma_0^4\} \\
&= \frac{1}{\eta^2}\sum_{m=1}^{\eta}[3\sigma_0^4\rho_{kl}^2 + (1 - \rho_{kl}^2)\sigma_0^4] - \sigma_0^4] \\
&= \frac{2\sigma_0^4\rho_{kl}^2}{\eta}
\end{aligned}
$$

Therefore,

$$
E(V_j) = E(E(V_j|\sigma_0^2)) = \sigma_0^2
$$

$$\text{Var}(V_j) = E(\text{Var}(V_j|\sigma_0^2)) + \text{Var}(E(V_j|\sigma_0^2))$$

$$= E\left[\frac{2\sigma_0^4}{\eta}\right] + \text{Var}(\sigma_0^2)$$

$$= \frac{2\sigma_0^4}{\eta}$$

$$\text{Cov}(V_k, V_l) = \frac{1}{\eta^2}\sum_{m=1}^{\eta}\sum_{n=1}^{\eta}\text{Cov}[X_{km}^2, X_{lm}^2]$$

$$= \frac{1}{\eta^2}\sum_{m=1}^{\eta}\sum_{n=1}^{\eta}\{E[X_{km}^2 X_{ln}^2] - (E(X_{km}))^2(E(X_{ln}))^2\}$$

$$= \frac{2\sigma_0^4\rho_{kl}^2}{\eta}$$

B.1.1 Smoothed variance images

General method

The degrees of freedom can be derived as follows.

For a specific j^{th} voxel V_j, the corresponding smoothed variance at the jth voxel is,

$$\hat{V}_j = \sum_{k=1}^{N_p} w_{jk}V_k \text{ , and } \sum_{k=1}^{N_p} w_{jk} = 1.$$

Then,

$$E(\hat{V}_j|\sigma_0^2) = \sigma_0^2$$

$$\text{Var}(\hat{V}_j|\sigma_0^2) = \sum_{k=1}^{N_p}\sum_{l=1}^{N_p} w_{jk}w_{jl}\text{Cov}(V_k, V_l|\sigma_0^2)$$

$$= \sum_{k=1}^{N_p}\sum_{l=1}^{N_p} w_{jk}w_{jl}\frac{2\sigma_0^4\rho_{kl}^2}{\eta}$$

$$= \frac{2\sigma_0^4}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p} w_{jk}w_{jl}\rho_{kl}^2.$$

Since $\hat{V}_j$ is a linear function of those elements from the sample variance image V (Worsley, 1996), it is rational to assume that

$$Q_j = \frac{d\hat{V}_j}{\tau^2} \sim \chi_d^2$$

where, d is the degrees of freedom for the smoothed variance image $\hat{V}$, and τ^2 is the true variance of $\hat{V}_j$. Then, according to the Satterthwaite approximation (Worsley, 2002),

$$\text{EDF} \approx d = \frac{2E(\hat{V}_j)^2}{\text{Var}(\hat{V}_j)} = \frac{2\sigma_0^4}{\frac{2\sigma_0^4}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2}$$

$$= \frac{\eta}{\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2}.$$

Convolution method

We also view the smoothed variance as,

$$\hat{V} = f \otimes V = wV,$$

where f is the smoothing kernel used for the variance images.

Also

$$w = \begin{pmatrix} w_{11} & w_{12} & \ldots & w_{1N_p} \\ w_{21} & w_{22} & \ldots & w_{2N_p} \\ \ldots & \ldots & \ldots & \ldots \\ w_{N_p1} & w_{N_p2} & \ldots & w_{N_pN_p} \end{pmatrix},$$

$$V = \begin{pmatrix} V_1 \\ V_2 \\ \ldots \\ V_{N_p} \end{pmatrix},$$

and

$$\sum_{k=1}^{N_p} w_{jk} = 1 \quad (j = 1, 2, \ldots, N_p).$$

Then,

$$E(\hat{V}|\sigma_0^2) = wE(V|\sigma_0^2) = \begin{pmatrix} w_{11} & w_{12} & \ldots & w_{1N_p} \\ w_{21} & w_{22} & \ldots & w_{2N_p} \\ \ldots & \ldots & \ldots & \ldots \\ w_{N_p1} & w_{N_p2} & \ldots & w_{N_pN_p} \end{pmatrix} \begin{pmatrix} \sigma_0^2 \\ \sigma_0^2 \\ \ldots \\ \sigma_0^2 \end{pmatrix} = \begin{pmatrix} \sigma_0^2 \\ \sigma_0^2 \\ \ldots \\ \sigma_0^2 \end{pmatrix}.$$

$$E(\hat{V}_j|\sigma_0^2) = \sigma_0^2,$$

$$\text{Var}(\hat{V}|\sigma_0^2) = w\text{Var}(V|\sigma_0^2)w^T = \frac{2\sigma_0^4}{\eta} \cdot w \begin{pmatrix} 1 & \rho_{12}^2 & \cdots & \rho_{1N_p}^2 \\ \rho_{21}^2 & 1 & \cdots & \rho_{2N_p}^2 \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{N_p1}^2 & \rho_{N_p2}^2 & \cdots & 1 \end{pmatrix} w^T,$$

$$\text{Var}(\hat{V}_j|\sigma_0^2) = \frac{2\sigma_0^4}{\eta} \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} w_{jk}w_{jl}\rho_{kl}^2.$$

If we assume that the covariance smoothing kernel is a Gaussian function, $f(\mathbf{u})$, with mean zero and variance $v^2 = F_V^2/(8\ln2)$,

$$f(\mathbf{u}) = (2\pi\, v^2)^{-D/2} \exp\{-\|\mathbf{u}\|^2/(2v^2)\} = \phi(\mathbf{u}, v^2).$$

Then, if $N_p \longrightarrow \infty$

$$
\begin{aligned}
\text{Var}(\hat{V}_j|\sigma_0^2) &= \frac{2\sigma_0^4}{\eta} \int \int (2\pi\, v^2)^{-D} \exp\{-\frac{t^2}{2v^2}\} \exp\{-\frac{s^2}{2v^2}\} \exp\{-\frac{(t-s)^2}{2s_d^2}\} dt ds \\
&= \frac{2\sigma_0^4}{\eta} \left(\frac{F_D^2}{F_D^2 + 2F_V^2}\right)^{D/2} \\
&= \frac{2\sigma_0^4}{\eta} \left[1 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2}.
\end{aligned}
$$

Similarly, according to the Satterthwaite approximation we have

$$
\begin{aligned}
\text{EDF} &\approx \frac{2E(\hat{V}_j|\sigma_0^2)^2}{\text{Var}(\hat{V}_j|\sigma_0^2)} = \eta \cdot \left(\frac{F_D^2 + 2F_V^2}{F_D^2}\right)^{D/2} \\
&= \eta \cdot \left[1 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{D/2}.
\end{aligned}
$$

B.1.2 Bias estimation

Since

$$\text{MSE}(\hat{V}_j) = \text{bias}_j{}^2 + \text{Var}(\hat{V}_j), \text{ and } \sigma_0^2 \text{ is a constant,}$$

$$
\begin{aligned}
(\text{bias}_j|\sigma_0^2)^2 &= \text{MSE}(\hat{V}_j|\sigma_0^2) - \text{Var}(\hat{V}_j|\sigma_0^2) \\
&= (E(\hat{V}_j|\sigma_0^2) - \sigma_0^2)^2 \\
&= 0.
\end{aligned}
$$

Therefore, $\text{MSE}(\hat{V}_j) = \text{Var}(\hat{V}_j) = \dfrac{2\sigma_0^4}{\eta} \left[1 + 2 \left(\dfrac{F_V}{F_D} \right)^2 \right]^{-D/2}.$

B.1.3 Summary

In summary, when the true variance image Σ is spatially homogeneous, the EDF, MSE and Variance (VAR) of the smoothed variance image for each voxel are

$$
\begin{aligned}
\text{EDF} &= \eta \cdot \left[1 + 2 \left(\frac{F_V}{F_D} \right)^2 \right]^{D/2}, \\
\text{VAR} &= \frac{2\sigma_0^4}{\eta} \left[1 + 2 \left(\frac{F_V}{F_D} \right)^2 \right]^{-D/2}, \\
\text{MSE} &= \text{VAR} = \frac{2\sigma_0^4}{\eta} \left[1 + 2 \left(\frac{F_V}{F_D} \right)^2 \right]^{-D/2},
\end{aligned}
$$

where, F_V, smoothness parameter (FWHM) for variance smoothing and F_D, smoothness parameter (FWHM) for data smoothing.

B.2 Spatially heterogeneous but fixed variance images

In this section, we assume that the variance image Σ is nonconstant but fixed. We want to know the degrees of freedom of the smoothed variance. Assume

$$\sigma_j^2 \neq \sigma_l^2,$$

where, σ_j^2, σ_l^2 are constant.

$$X_{jm}|\sigma_j^2 \sim N(\mu_j, \sigma_j^2),$$

$$\text{Cov}(X_{km}, X_{ln}|\sigma_j^2, \sigma_l^2) = \rho_{kl}\sigma_k\sigma_l\delta_{m-n},$$

where $j\,k, l = 1, 2, \cdots, N_p$ index voxels of an image, and $m\,, n = 1, 2, \cdots, \eta$ index subjects in an experiment.

Similar to B.1

$$\rho_{kl} = g(\mathbf{u}_{kl}, s_d^2) = \exp\{-\|\mathbf{u}_{kl}\|_2^2/(4s_d^2)\} = \exp\{-\|\mathbf{P}_k - \mathbf{P}_l\|_2^2/(4s_d^2)\}, \ s_d^2 = F_D^2/8\ln 2,$$

here, F_D is the FWHM from the data conditional on Σ, $\mathbf{P}_k = (x_k, y_k, z_k)^t$ denotes the coordinate of the kth voxel, and $\delta_{m-n} = 1$ if $m = n$, 0 otherwise.

Under null hypothesis, $\mu_j = 0$.

The sample variance for the jth voxel is

$$V_j = \frac{1}{\eta}\sum_{i=1}^{\eta} X_{ji}^2.$$

Then,

$$
\begin{aligned}
E(V_j|\sigma_j^2) &= E(\frac{1}{\eta}\sum_{i=1}^{\eta} X_{ji}^2|\sigma_j^2) \\
&= \frac{1}{\eta}\sum_{i=1}^{\eta} E\left[X_{ji}^2|\sigma_j^2\right] \\
&= \frac{1}{\eta}\sum_{i=1}^{\eta} \sigma_j^2 \\
&= \sigma_j^2,
\end{aligned}
$$

$$\begin{aligned}
\text{Var}(V_j|\sigma_j^2) &= E(V_j^2|\sigma_j^2) - \left[E(V_j|\sigma_j^2)\right]^2 \\
&= \frac{1}{\eta^2}E\left(\sum_{m=1}^{\eta}X_{jm}^2\sum_{n=1}^{\eta}X_{jn}^2 \mid \sigma_j^2\right) - \sigma_j^4 \\
&= \frac{1}{\eta^2}\{\sum_{m=1}^{\eta}E\left[X_{jm}^4|\sigma_j^2\right] + \sum_{m\neq n;m,n=1}^{\eta}E\left[X_{jm}^2X_{jn}^2|\sigma_j^2\right]\} - \sigma_j^4 \\
&= \frac{1}{\eta^2}(\eta\cdot 3\sigma_j^4 + \eta(\eta-1)\sigma_j^4) - \sigma_j^4 \\
&= \frac{2\sigma_j^4}{\eta},
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(V_k, V_l|\sigma_k^2, \sigma_l^2) &= \frac{1}{\eta^2}\sum_{m=1}^{\eta}\sum_{n=1}^{\eta}\text{Cov}(X_{km}^2, X_{ln}^2|\sigma_k^2, \sigma_l^2) \\
&= \frac{1}{\eta^2}\sum_{m=1}^{\eta}\{E\left(X_{km}^2X_{lm}^2|\sigma_k^2, \sigma_l^2\right) - E[X_{km}|\sigma_k^2]E[X_{ln}|\sigma_l^2]\}.
\end{aligned}$$

Because

$$X_{km}|\sigma_k^2 \sim N(0, \sigma_k^2),$$
$$X_{lm}|\sigma_l^2 \sim N(0, \sigma_l^2)$$

and $X_{km}|\sigma_k^2$, $X_{lm}|\sigma_l^2$ follow bivariate normal distribution, it follows that,

$$X_{km}|X_{lm} \sim N(\rho_{kl}X_{lm}\sigma_k/\sigma_l, \sigma_k^2(1-\rho_{kl}^2)).$$

Hence,

$$\begin{aligned}
\text{Cov}(V_k, V_l|\sigma_k^2, \sigma_l^2) &= \frac{1}{\eta^2}\sum_{m=1}^{\eta}(E\{E[X_{km}^2X_{lm}^2|X_{lm}]|\sigma_k^2, \sigma_l^2\} - \sigma_k^2\sigma_l^2) \\
&= \frac{1}{\eta^2}\sum_{m=1}^{\eta}\{E[\rho_{kl}^2\sigma_k^2X_{lm}^4/\sigma_l^2 + \sigma_k^2(1-\rho_{kl}^2)X_{lm}^2|\sigma_k^2, \sigma_l^2] - \sigma_k^2\sigma_l^2\} \\
&= \frac{1}{\eta^2}\sum_{m=1}^{\eta}[3\sigma_k^2\sigma_l^2\rho_{kl}^2 + (1-\rho_{kl}^2)\sigma_k^2\sigma_l^2) - \sigma_k^2\sigma_l^2] \\
&= \frac{2\sigma_k^2\sigma_l^2\rho_{kl}^2}{\eta}.
\end{aligned}$$

Therefore, the unconditional mean and variance in this case are

$$E(V_j) = E(E(V_j|\sigma_j^2)) = \sigma_j^2,$$

$$
\begin{aligned}
\text{Var}(V_j) &= E(\text{Var}(V_j|\sigma_j^2)) + \text{Var}(E(V_j|\sigma_j^2)) \\
&= E\left(\frac{2\sigma_j^4}{\eta}\right) + \text{Var}(\sigma_j^2) \\
&= \frac{2\sigma_j^4}{\eta},
\end{aligned}
$$

$$
\begin{aligned}
\text{Cov}(V_k, V_l) &= \frac{1}{\eta^2}\sum_{m=1}^{\eta}\sum_{n=1}^{\eta}\text{Cov}[X_{km}^2, X_{lm}^2] \\
&= \frac{1}{\eta^2}\sum_{m=1}^{\eta}\sum_{n=1}^{\eta}\{E[X_{km}^2 X_{ln}^2] - E(X_{km})^2 E(X_{ln})^2\} \\
&= \frac{2\sigma_k^2\sigma_l^2\rho_{kl}^2}{\eta}.
\end{aligned}
$$

B.2.1 Smoothed variance images

General method

For a specific j^{th} voxel V_j, the corresponding smoothed variance at the jth voxel is,

$$\hat{V}_j = \sum_{k=1}^{N_p} w_{jk}V_k \text{ , and } \sum_{k=1}^{N_p} w_{jk} = 1.$$

Then, conditional mean, variance for the smoothed variance are,

$$E(\hat{V}_j|\sigma_1^2, \ldots, \sigma_{N_p}^2) = \sum_{k=1}^{N_p} w_{jk}\sigma_k^2,$$

$$
\begin{aligned}
\text{Var}(\hat{V}_j|\sigma_1^2, \ldots, \sigma_{N_p}^2) &= \sum_{k=1}^{N_p}\sum_{l=1}^{N_p} w_{jk}w_{jl}\text{Cov}(V_k, V_l|\sigma_k^2, \sigma_l^2) \\
&= \sum_{k=1}^{N_p}\sum_{l=1}^{N_p} w_{jk}w_{jl}\frac{2\sigma_0^4\rho_{kl}^2}{\eta} \\
&= \frac{2}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p} w_{jk}w_{jl}\rho_{kl}^2\sigma_k^2\sigma_l^2.
\end{aligned}
$$

Since $\hat{V}_j$ is a linear function of those elements from the sample variance imge V (Worsley, 1996), it is rational to assume that

$$Q_j = \frac{d\hat{V}_j}{\tau^2} \sim \chi_d^2,$$

where, d is the degrees of freedom for the smoothed variance image $\hat{V}$, and τ^2 is the true variance of $\hat{V}_j$. Then, according to the Satterthwaite approximation (Worsley, 2002),

$$(B.1) \qquad \text{EDF} \approx d = \frac{2E(\hat{V}_j|\sigma_1^2, \ldots, \sigma_{N_p}^2)^2}{\text{Var}(\hat{V}_j|\sigma_1^2, \ldots, \sigma_{N_p}^2)} = \frac{(\sum_{k=1}^{N_p} w_{jk}\sigma_k^2)^2}{\frac{1}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p} w_{jk}w_{jl}\rho_{kl}^2\sigma_k^2\sigma_l^2}.$$

Convolution method

We also view the smoothed variance as

$$\hat{V} = f \otimes V = wV,$$

where f is the smoothing kernel used for the variance images.

Also,

$$w = \begin{pmatrix} w_{11} & w_{12} & \ldots & w_{1N_p} \\ w_{21} & w_{22} & \ldots & w_{2N_p} \\ \ldots & \ldots & \ldots & \ldots \\ w_{N_p1} & w_{N_p2} & \ldots & w_{N_pN_p} \end{pmatrix},$$

$$V = \begin{pmatrix} V_1 \\ V_2 \\ \ldots \\ V_{N_p} \end{pmatrix},$$

and

$$\sum_{k=1}^{N_p} w_{jk} = 1 \quad (j = 1, 2, \ldots, N_p).$$

Then,

$$E(\hat{V}|\sigma_1^2,\ldots,\sigma_{N_p}^2) = wE(V|\sigma_1^2,\ldots,\sigma_{N_p}^2)$$

$$= \begin{pmatrix} w_{11} & w_{12} & \ldots & w_{1N_p} \\ w_{21} & w_{22} & \ldots & w_{2N_p} \\ \ldots & \ldots & \ldots & \ldots \\ w_{N_p1} & w_{N_p2} & \ldots & w_{N_pN_p} \end{pmatrix} \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \ldots \\ \sigma_{N_p}^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{k=1}^{N_p} w_{1k}\sigma_k^2 \\ \sum_{k=1}^{N_p} w_{2k}\sigma_k^2 \\ \ldots \\ \sum_{k=1}^{N_p} w_{N_pk}\sigma_k^2 \end{pmatrix},$$

$$E(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2) = \sum_{k=1}^{N_p} w_{jk}\sigma_k^2,$$

$$\text{Var}(\hat{V}|\sigma_1^2,\ldots,\sigma_{N_p}^2) = w\text{Var}(V|\sigma_1^2,\ldots,\sigma_{N_p}^2)w^T$$

$$= \frac{2}{\eta}\cdot w \begin{pmatrix} \sigma_1^4 & \sigma_1^2\sigma_2^2\rho_{12}^2 & \ldots & \sigma_1^2\sigma_{N_p}^2\rho_{1N_p}^2 \\ \sigma_2^2\sigma_1^2\rho_{21}^2 & \sigma_2^4 & \ldots & \sigma_2^2\sigma_{N_p}^2\rho_{2N_p}^2 \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{N_p}^2\sigma_1^2\rho_{N_p1}^2 & \sigma_{N_p}^2\sigma_2^2\rho_{N_p2}^2 & \ldots & \sigma_{N_p}^4 \end{pmatrix} w^T,$$

$$\text{Var}(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2) = \frac{2}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p} w_{jk}w_{jl}\rho_{kl}^2\sigma_k^2\sigma_l^2.$$

We have a similar form as Eq. (B.1) for the smoothed variance degrees of freedom with the convolution method. However, since the true variance Σ is fixed, we do not have a clear form for this case.

B.2.2 Bias estimation

Since $\text{MSE}(\hat{V}_j | \sigma_1^2, \ldots, \sigma_{N_p}^2) = (\text{bias}_j | \sigma_1^2, \ldots, \sigma_{N_p}^2)^2 + \text{Var}(\hat{V}_j | \sigma_1^2, \ldots, \sigma_{N_p}^2)$, we have

$$
\begin{aligned}
(\text{bias}_j | \sigma_1^2, \ldots, \sigma_{N_p}^2)^2 &= \text{MSE}(\hat{V}_j | \sigma_1^2, \ldots, \sigma_{N_p}^2) - \text{Var}(\hat{V}_j | \sigma_1^2, \ldots, \sigma_{N_p}^2) \\
&= (E(\hat{V}_j | \sigma_1^2, \ldots, \sigma_{N_p}^2) - \sigma_j^2)^2 \\
&= (\sum_{k=1}^{N_p} w_{jk} \sigma_k^2 - \sigma_j^2)^2.
\end{aligned}
$$

There is no clear form for bias^2 in this case.

B.3 Spatially heterogeneous and random variance images

In this section, we assume that the true variance image Σ is a random field, which means the true variance image is not independent. We want to derive the degrees of freedom of the smoothed variance.

B.3.1 Correlation of any two voxels in a random field variance image

We assume a Gaussian random field $\mathcal{Z}$ has the following properties,

1. $Z(i)|\sigma_i^2 \sim N(0, \sigma_i^2)$, where, i indexes for voxels and $i = 1, \ldots, N_p$.

2.

$$
r(i,j) = \begin{cases} r_{ij} & \text{correlation between the } i^{th} \text{ and } j^{th} \text{ voxels within a subject} \\ & \text{in Gaussian random field } \mathcal{Z} \\ 0 & \text{correlation between subjects} \end{cases}
$$

3. the corresponding χ^2 random field is generated by ν independent and identically distributed Gaussian random field.

4. $\frac{\nu \cdot \sigma_i^2}{\sigma_0^2} \sim \chi_\nu^2$, we consider ν as ν subjects, and σ_0^2 is a constant.

The true variance image comes from several Gaussian noise images and each Gaussian noise image is convolved with a Gaussian kernel $h(\mathbf{u})$ with mean zero and variance $s_v^2 = F_{\sigma^2}^2/(8 \log 2)$, F_{σ^2} is the FWHM of smoothing kernel for those Gaussian noise images which are used to generate the true variance image, i.e.

$$
h(\mathbf{u}) = \phi(\mathbf{u}; s_v^2).
$$

Hence, the spatial correlation of each Gaussian random field is

$$
cc(\mathbf{u}) = e^{-\|\mathbf{u}\|^2/(4s_v^2)} = (2\pi 2 s_v^2)^{D/2} \phi(\mathbf{u}; 2s_v^2).
$$

We can also write the correlation function between the kth and l^{th} voxel in one Gaussian image as

$$r_{kl} = g(\mathbf{u}_{kl}, s_v^2) = \exp\{-\|\mathbf{u}_{kl}\|_2^2/(4s_v^2)\} = \exp\{-\|\mathbf{P}_k - \mathbf{P}_l\|_2^2/(4s_v^2)\}.$$

The assumption turns to be the following:

1. $Z_1(i), Z_2(i), \ldots, Z_n(i)$ are i.i.d $N(0, \sigma_i^2)$.

2. in each Gaussian image, the correlation $r(\mathbf{u}) = \exp\{-\|\mathbf{u}_{kl}\|_2^2/(4s_v^2)\}$.

3. $U_i|\sigma_i^2 = \sum_{k=1}^{\nu} \left(\frac{Z_k(i)}{\sigma_i}\right)^2 \sim \chi_\nu^2.$

We want to find the correlation between U_i and U_j in the χ^2 random field

$$
\begin{aligned}
\text{Cov}(U_i, U_j) &= \text{Cov}\left[\sum_{k=1}^{\nu}\left(\frac{Z_k(i)}{\sigma_i}\right)^2, \sum_{l=1}^{\nu}\left(\frac{Z_l(j)}{\sigma_j}\right)^2\right] \\
&= \sum_{k=1}^{\nu}\sum_{l=1}^{\nu}\text{Cov}\left[\left(\frac{Z_k(i)}{\sigma_i}\right)^2, \left(\frac{Z_l(j)}{\sigma_j}\right)^2\right] \\
&= \sum_{k=1}^{\nu}\text{Cov}\left[\left(\frac{Z_k(i)}{\sigma_i}\right)^2, \left(\frac{Z_k(j)}{\sigma_j}\right)^2\right] \\
&= \sum_{k=1}^{\nu}\left[E\left(\frac{Z_k^2(i)}{\sigma_i^2}\frac{Z_k^2(j)}{\sigma_j^2}\right) - E\left(\frac{Z_k^2(i)}{\sigma_i^2}\right)E\left(\frac{Z_k^2(j)}{\sigma_j^2}\right)\right].
\end{aligned}
$$

Since

$$\frac{Z_k(i)}{\sigma_i}\bigg|\frac{Z_k(j)}{\sigma_j} \sim N(r_{ij}\frac{Z_k(j)}{\sigma_j}, 1 - r_{ij}^2),$$

then

$$
\begin{aligned}
\text{Cov}(U_i, U_j) &= \sum_{k=1}^{\nu}\left\{E\left[\frac{Z_k^2(j)}{\sigma_j^2}E\left(\frac{Z_k^2(i)}{\sigma_i^2}\bigg|\frac{Z_k^2(j)}{\sigma_j^2}\right)\right] - 1\right\} \\
&= \sum_{k=1}^{\nu}(1 + 2r_{ij}^2 - 1) \\
&= 2\nu \cdot r_{ij}^2.
\end{aligned}
$$

Let R_{ij} denote the correlation of U_i and U_j in a χ^2 random field. Then

$$R_{ij} = \frac{\text{Cov}(U_i, U_j)}{\sqrt{\text{Var}(U_i)}\sqrt{\text{Var}(U_j)}} = \frac{2\nu \cdot r_{ij}^2}{\sqrt{2\nu}\sqrt{2\nu}} = r_{ij}^2.$$

Therefore, if the correlation between the ith and j^{th} voxel is r_{ij} in a Gaussian random field, in the corresponding χ^2 random field, which is generated with ν i.i.d Gaussian random fields, the correlation between the ith and j^{th} voxel is r_{ij}^2.

B.3.2 Preparation before smoothing a true variance image Σ

Assume

$$\frac{\nu \sigma_j^2}{\sigma_0^2} \sim \chi_\nu^2,$$

and

$$Corr\left(\frac{\nu \sigma_i^2}{\sigma_0^2}, \frac{\nu \sigma_j^2}{\sigma_0^2}\right) = R_{ij} = r_{ij}^2,$$

where σ_0^2 is a constant. Also,

$$X_{jm}|\sigma_j^2 \quad \sim \quad N(\mu_j, \sigma_j^2),$$

$$\text{Cov}(X_{km}, X_{ln}|\sigma_j^2, \sigma_l^2) \quad = \quad \rho_{kl}\sigma_k\sigma_l\delta_{m-n},$$

where $j\,k, l = 1, 2, \cdots, N_p$ index voxels of an image, $m, n = 1, 2, \cdots, \eta$ index subjects in an experiment. Similar to B.1 assumption, $\rho_{kl} = g(\mathbf{u}_{kl}, s_d^2) = \exp\{-\|\mathbf{u}_{kl}\|_2^2/(4s_d^2)\} = \exp\{-\|\mathbf{P}_k - \mathbf{P}_l\|_2^2/(4s_d^2)\}$, $s_d^2 = F_D^2/(8\ln 2)$, here, F_D is the FWHM from the data conditional on σ_0^{22}, $\mathbf{P}_k = (x_k, y_k, z_k)^t$ denotes the coordinate of the kth voxel, and $\delta_{m-n} = 1$ if $m = n$, 0 otherwise.

Under the null hypothesis, we have $\mu_j = 0$.

The sample variance for the jth voxel is

$$V_j = \frac{1}{\eta}\sum_{i=1}^{\eta} X_{ji}^2.$$

---

[2]$\|X\|_2 = \|X\|$ is l_2 norm, and $\|X\|_2^2$ denotes the distance between two voxels

Conditional mean, variance and covariance of variance images

$$
\begin{aligned}
E(V_j|\sigma_j^2) &= E\left(\frac{1}{\eta}\sum_{i=1}^{\eta}X_{ji}^2|\sigma_j^2\right) = \frac{1}{\eta}\sum_{i=1}^{\eta}E(X_{ji}^2|\sigma_j^2) \\
&= \frac{1}{\eta}\sum_{i=1}^{\eta}\sigma_j^2 = \sigma_j^2.
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(V_j|\sigma_j^2) &= E(V_j^2|\sigma_j^2) - \left[E(V_j|\sigma_j^2)\right]^2 \\
&= \frac{1}{\eta^2}E\left(\sum_{m=1}^{\eta}X_{jm}^2\sum_{n=1}^{\eta}X_{jn}^2 \mid \sigma_j^2\right) - \sigma_j^4 \\
&= \frac{1}{\eta^2}\{\sum_{m=1}^{\eta}E(X_{jm}^4|\sigma_j^2) + \sum_{m\neq n;m,n=1}^{\eta}E(X_{jm}^2X_{jn}^2|\sigma_j^2)\} - \sigma_j^4 \\
&= \frac{1}{\eta^2}(\eta\cdot 3\sigma_j^4 + \eta(\eta-1)\sigma_j^4) - \sigma_j^4 \\
&= \frac{2\sigma_j^4}{\eta}.
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}(V_k,V_l|\sigma_k^2,\sigma_l^2) &= \frac{1}{\eta^2}\sum_{m=1}^{\eta}\sum_{n=1}^{\eta}\mathrm{Cov}(X_{km}^2,X_{ln}^2|\sigma_k^2,\sigma_l^2) \\
&= \frac{1}{\eta^2}\sum_{m=1}^{\eta}\{E(X_{km}^2X_{lm}^2|\sigma_k^2,\sigma_l^2) - E(X_{km}|\sigma_k^2)E(X_{ln}|\sigma_l^2)\}.
\end{aligned}
$$

Because

$$
X_{km}|\sigma_k^2 \sim N(0,\sigma_k^2),
$$
$$
X_{lm}|\sigma_l^2 \sim N(0,\sigma_l^2),
$$

and $X_{km}|\sigma_k^2$, $X_{lm}|\sigma_l^2$ follow bivariate normal distribution, which means,

$$
X_{km}|X_{lm} \sim N(\rho_{kl}X_{lm}\sigma_k/\sigma_l,\sigma_k^2(1-\rho_{kl}^2)).
$$

Hence,

$$
\begin{aligned}
\mathrm{Cov}(V_k, V_l|\sigma_k^2, \sigma_l^2) &= \frac{1}{\eta^2} \sum_{m=1}^{\eta} \{E[E(X_{km}^2 X_{lm}^2|X_{lm})|\sigma_k^2, \sigma_l^2] - \sigma_k^2 \sigma_l^2\} \\
&= \frac{1}{\eta^2} \sum_{m=1}^{\eta} \{E[\rho_{kl}^2 \sigma_k^2 X_{lm}^4/\sigma_l^2 + \sigma_k^2(1 - \rho_{kl}^2)X_{lm}^2|\sigma_k^2, \sigma_l^2] - \sigma_k^2 \sigma_l^2\} \\
&= \frac{1}{\eta^2} \sum_{m=1}^{\eta} [3\sigma_k^2 \sigma_l^2 \rho_{kl}^2 + (1 - \rho_{kl}^2)\sigma_k^2 \sigma_l^2) - \sigma_k^2 \sigma_l^2] \\
&= \frac{2\sigma_k^2 \sigma_l^2 \rho_{kl}^2}{\eta}.
\end{aligned}
$$

Unconditional mean, variance and covariance of variance images

$$
E(V_j) = E(E(V_j|\sigma_j^2)) = E(\sigma_j^2) = \sigma_0^2.
$$

$$
\begin{aligned}
\mathrm{Var}(V_j) &= E(V_j^2) - (E(V_j))^2 = E[E(V_j^2|\sigma_j^2)] - \sigma_0^4 \\
&= E\left(\frac{\eta^2 + 2\eta}{\eta^2}\sigma_j^4\right) - \sigma_0^4 \\
&= \frac{\eta^2 + 2\eta}{\eta^2}E(\sigma_j^4) - \sigma_0^4 \\
&= \frac{\eta + 2}{\eta} \cdot \frac{2\nu + \nu^2}{\nu^2}\sigma_0^4 - \sigma_0^4 \\
&= \left(\frac{\eta + 2}{\eta} \cdot \frac{\nu + 2}{\nu} - 1\right)\sigma_0^4.
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}(V_k, V_l) &= E(V_k V_l) - E(V_k)E(V_l) \\
&= E\left(\frac{1}{\eta^2} \sum_{m=1}^{\eta} X_{km}^2 \sum_{n=1}^{\eta} X_{ln}^2\right) - \sigma_0^4 \\
&= \frac{\eta + 2\rho_{kl}^2}{\eta}E(\sigma_k^2 \sigma_l^2) - \sigma_0^4 \\
&= \frac{\eta + 2\rho_{kl}^2}{\eta} \cdot \frac{\sigma_0^4}{\nu^2}\left[\mathrm{Cov}(\frac{\nu\sigma_k^2}{\sigma_0^2}, \frac{\nu\sigma_l^2}{\sigma_0^2}) + E\left(\frac{\nu\sigma_k^2}{\sigma_0^2}\right)E\left(\frac{\nu\sigma_l^2}{\sigma_0^2}\right)\right] - \sigma_0^4 \\
&= \frac{\eta + 2\rho_{kl}^2}{\eta} \cdot \frac{\nu + 2r_{kl}^2}{\nu}\sigma_0^4 - \sigma_0^4 \\
&= \left(\frac{\eta + 2\rho_{kl}^2}{\eta} \cdot \frac{\nu + 2r_{kl}^2}{\nu} - 1\right)\sigma_0^4.
\end{aligned}
$$

B.3.3 Smoothed variance

For a specific j^{th} voxel V_j, the corresponding smoothed variance at the jth voxel is,

$$\hat{V}_j = \sum_{k=1}^{N_p} w_{jk} V_k \text{ , and } \sum_{k=1}^{N_p} w_{jk} = 1.$$

Conditional mean and variance for the smoothed variance images

$$E(\hat{V}_j | \sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_p}^2) = \sum_{k=1}^{N_p} w_{jk} \sigma_k^2.$$

$$\begin{aligned}
& \text{Var}(\hat{V}_j | \sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_p}^2) \\
= & \sum_{k=1}^{N_p} \text{Var}(w_{jk} V_k | \sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_p}^2) \\
& +2 \sum_{k \neq l, k, n=1}^{N_p} \text{Cov}(w_{jk} V_k, w_{jl} V_l | \sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_p}^2) \\
= & \sum_{k=1}^{N_p} w_{jk}^2 \frac{2\sigma_k^4}{\eta} + 2 \sum_{k \neq l, k, n=1}^{N_p} w_{jk} w_{jl} \frac{2\sigma_k^2 \sigma_l^2 \rho_{kl}^2}{\eta} \\
= & \frac{2}{\eta} (\sum_{k=1}^{N_p} w_{jk}^2 \sigma_k^4 + \sum_{k \neq l, k, n=1}^{N_p} w_{jk} w_{jl} \sigma_k^2 \sigma_l^2 \rho_{kl}^2) \\
= & \frac{2}{\eta} \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} w_{jk} w_{jl} \sigma_k^2 \sigma_l^2 \rho_{kl}^2.
\end{aligned}$$

Expectation of mean and variance of the smoothed variance images

$$(\text{B.2}) \qquad E[E(\hat{V}_j | \sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_p}^2) | \sigma_0^2] = E(\sum_{k=1}^{N_p} w_{jk} \sigma_k^2 | \sigma_0^2) = \sigma_0^2$$

and

$$E[\text{Var}(\hat{V}_j|\sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_p}^2)|\sigma_0^2]$$

$$= E[\frac{2}{\eta} \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} w_{jk} w_{jl} \sigma_k^2 \sigma_l^2 \rho_{kl}^2 |\sigma_0^2]$$

$$= \frac{2}{\eta} \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} w_{jk} w_{jl} \rho_{kl}^2 E[\sigma_k^2 \sigma_l^2 |\sigma_0^2]$$

$$= \frac{2}{\eta} \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} w_{jk} w_{jl} \rho_{kl}^2 (1 + 2\frac{r_{kl}^2}{\nu}) \sigma_0^4$$

(B.3)
$$= \frac{2\sigma_0^4}{\eta} \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} w_{jk} w_{jl} \rho_{kl}^2 + \frac{4\sigma_0^4}{\eta\nu} \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} w_{jk} w_{jl} \rho_{kl}^2 r_{kl}^2.$$

Unconditional mean and variance for the smoothed variance images

$$E(\hat{V}_j|\sigma_0^2) = \sigma_0^2.$$

$$\text{Var}(\hat{V}_j|\sigma_0^2) = \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} w_{jk} w_{jl} \text{Cov}(V_k, V_l|\sigma_0^2)$$

$$= \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} w_{jk} w_{jl} \left(\frac{\eta + 2\rho_{kl}^2}{\eta} \cdot \frac{\nu + 2r_{kl}^2}{\nu} - 1 \right) \sigma_0^4$$

$$= \frac{\sigma_0^4}{\eta \cdot \nu} \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} w_{jk} w_{jl} [(\eta + 2\rho_{kl}^2)(2r_{kl}^2 + \nu) - \eta\nu]$$

$$= \frac{2\sigma_0^4}{\eta\nu} \sum_{k=1}^{N_p} \sum_{l=1}^{N_p} w_{jk} w_{jl} (\eta r_{kl}^2 + \nu\rho_{kl}^2 + 2\rho_{kl}^2 r_{kl}^2).$$

Degrees of freedom for the smoothed variance

Since $\hat{V}_j$ is a linear function of those elements from the sample variance imge V (Worsley, 1996), it is rational to assume that

$$Q_j = \frac{d\hat{V}_j}{\tau^2} \sim \chi_d^2,$$

where d is the degrees of freedom for the smoothed variance image $\hat{V}$, and τ^2 is the true variance of $\hat{V}_j$. Then, according to the Satterthwaite approximation (Worsley, 2002),

$$
\begin{aligned}
\text{EDF} \quad \approx \quad d &\approx \frac{2E[E(\hat{V}_j|\sigma_1^2,\sigma_2^2,\ldots,\sigma_{N_p}^2)|\sigma_0^2]^2}{E[\text{Var}(\hat{V}_j|\sigma_1^2,\sigma_2^2,\ldots,\sigma_{N_p}^2)|\sigma_0^2]} \\
&= \frac{2(\sigma_0^2)^2}{\frac{2\sigma_0^4}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2 + \frac{4\sigma_0^4}{\eta\nu}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2 r_{kl}^2}.
\end{aligned}
$$

Using convolution methods

We also view the smoothed variance as,

$$
\hat{V} = f \otimes V = wV
$$

where f is the smoothing kernel used for the variance images.

Also,

$$
w = \begin{pmatrix}
w_{11} & w_{12} & \cdots & w_{1N_p} \\
w_{21} & w_{22} & \cdots & w_{2N_p} \\
\cdots & \cdots & \cdots & \cdots \\
w_{N_p1} & w_{N_p2} & \cdots & w_{N_pN_p}
\end{pmatrix},
$$

$$
V = \begin{pmatrix}
V_1 \\
V_2 \\
\cdots \\
V_{N_p}
\end{pmatrix},
$$

and

$$
\sum_{k=1}^{N_p} w_{jk} = 1 \quad (j = 1,\, 2,\, \ldots,\, N_p).
$$

Then,

$$E(\hat{V}|\sigma_1^2, \ldots, \sigma_{N_p}^2) = wE(V|\sigma_1^2, \ldots, \sigma_{N_p}^2)$$

$$= \begin{pmatrix} w_{11} & w_{12} & \ldots & w_{1N_p} \\ w_{21} & w_{22} & \ldots & w_{2N_p} \\ \ldots & \ldots & \ldots & \ldots \\ w_{N_p1} & w_{N_p2} & \ldots & w_{N_pN_p} \end{pmatrix} \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \ldots \\ \sigma_{N_p}^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{k=1}^{N_p} w_{1k}\sigma_k^2 \\ \sum_{k=1}^{N_p} w_{2k}\sigma_k^2 \\ \ldots \\ \sum_{k=1}^{N_p} w_{N_pk}\sigma_k^2 \end{pmatrix},$$

$$E(\hat{V}_j|\sigma_1^2, \ldots, \sigma_{N_p}^2) = \sum_{k=1}^{N_p} w_{jk}\sigma_k^2,$$

$$\text{Var}(\hat{V}|\sigma_1^2, \ldots, \sigma_{N_p}^2) = w\text{Var}(V|\sigma_1^2, \ldots, \sigma_{N_p}^2)w^T$$

$$= \frac{2}{\eta} \cdot w \begin{pmatrix} \sigma_1^4 & \sigma_1^2\sigma_2^2\rho_{12}^2 & \ldots & \sigma_1^2\sigma_{N_p}^2\rho_{1N_p}^2 \\ \sigma_2^2\sigma_1^2\rho_{21}^2 & \sigma_2^4 & \ldots & \sigma_2^2\sigma_{N_p}^2\rho_{2N_p}^2 \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{N_p}^2\sigma_1^2\rho_{N_p1}^2 & \sigma_{N_p}^2\sigma_2^2\rho_{N_p2}^2 & \ldots & \sigma_{N_p}^4 \end{pmatrix} w^T,$$

$$\text{Var}(\hat{V}_j|\sigma_1^2, \ldots, \sigma_{N_p}^2) = \frac{2}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p} w_{jk}w_{jl}\rho_{kl}^2\sigma_k^2\sigma_l^2.$$

Therefore,

$$E(\hat{V}_j|\sigma_0^2) = E[E(\hat{V}_j|\sigma_1^2, \ldots, \sigma_{N_p}^2)|\sigma_0^2] = E(\sum_{k=1}^{N_p} w_{jk}\sigma_k^2|\sigma_0^2) = \sigma_0^2,$$

and

$$E[\text{Var}(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2)|\sigma_0^2] = E[\frac{2}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2\sigma_k^2\sigma_l^2|\sigma_0^2]$$

$$= \frac{2\sigma_0^4}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2 + \frac{4\sigma_0^4}{\eta\nu}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2 r_{kl}^2,$$

and

$$\text{Var}(\hat{V}_j|\sigma_0^2) = E(\text{Var}(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2)|\sigma_0^2) + \text{Var}(\text{Var}(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2)|\sigma_0^2)$$

$$= E(\frac{2}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2\sigma_k^2\sigma_l^2|\sigma_0^2) + \text{Var}(\sum_{k=1}^{N_p}w_{jk}\sigma_k^2|\sigma_0^2)$$

$$= \frac{2}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2 E(\sigma_k^2\sigma_l^2|\sigma_0^2)) + \sum_{k=1}^{N_p}\sum_{l=1}^{N_p}\text{Cov}(w_{jk}\sigma_k^2,w_{jl}\sigma_l^2|\sigma_0^2)$$

$$= \frac{2}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2(\text{Cov}(\sigma_k^2,\sigma_l^2|\sigma_0^2) + E(\sigma_k^2|\sigma_0^2)E(\sigma_l^2|\sigma_0^2))$$

$$+ \sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}r_{kl}^2 \cdot 2\nu \cdot \frac{\sigma_0^4}{\nu^2}$$

$$= \frac{2\sigma_0^4}{\eta\nu}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}(\eta r_{kl}^2 + \nu\rho_{kl}^2 + 2\rho_{kl}^2 r_{kl}^2).$$

If we assume the covariance smoothing kernel is a Gaussian, $f(\mathbf{u})$, with mean zero and variance $v^2 = F_V^2/(8\log 2)$,

$$f(\mathbf{u}) = (2\pi\,v^2)^{-D/2}\,\exp\{-\|\mathbf{u}\|^2/(2v^2)\} = \phi(\mathbf{u},v^2),$$

then

$$E[\text{Var}(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2)|\sigma_0^2]$$

$$= \frac{2\sigma_0^4}{\eta}\int\int(2\pi\,v^2)^{-D}\exp\{-\frac{t^2}{2v^2}\}\exp\{-\frac{s^2}{2v^2}\}\exp\{-\frac{(t-s)^2}{2s_d^2}\}dtds$$

$$+ \frac{4\sigma_0^4}{\eta\nu}\int\int(2\pi\,v^2)^{-D}\exp\{-\frac{t^2}{2v^2}\}\exp\{-\frac{s^2}{2v^2}\}\exp\{-\frac{(t-s)^2}{2s_v^2}\}\exp\{-\frac{(t-s)^2}{2s_d^2}\}dtds$$

$$= \frac{2\sigma_0^4}{\eta}\left(\frac{F_D^2}{F_D^2+2F_V^2}\right)^{D/2} + \frac{4\sigma_0^4}{\eta\nu}\left(\frac{F_D^2 F_{\sigma^2}^2}{F_D^2 F_{\sigma^2}^2+2F_D^2 F_V^2+2F_{\sigma^2}^2 F_V^2}\right)^{D/2},$$

and

$$
\begin{aligned}
\mathrm{Var}(\hat{V}_j|\sigma_0^2) &= \frac{2\sigma_0^4}{\eta\nu}[\eta \int\int (2\pi\,v^2)^{-D}\exp\{-\frac{t^2}{2v^2}\}\exp\{-\frac{s^2}{2v^2}\}\exp\{-\frac{(t-s)^2}{2s_v^2}\}dtds \\
&\quad +\nu\int\int (2\pi\,v^2)^{-D}\exp\{-\frac{t^2}{2v^2}\}\exp\{-\frac{s^2}{2v^2}\}\exp\{-\frac{(t-s)^2}{2s_d^2}\}dtds \\
&\quad +2\int\int ((2\pi\,v^2)^{-D}\exp\{-\frac{t^2}{2v^2}\}\exp\{-\frac{s^2}{2v^2}\}\exp\{-\frac{(t-s)^2}{2s_v^2}\} \\
&\quad \exp\{-\frac{(t-s)^2}{2s_d^2}\})dtds] \\
&= \frac{2\sigma_0^4}{\eta\nu}[\eta\left(\frac{F_{\sigma^2}^2}{F_{\sigma^2}^2+2F_V^2}\right)^{D/2} + \nu\left(\frac{F_D^2}{F_D^2+2F_V^2}\right)^{D/2} \\
&\quad +2\left(\frac{F_D^2 F_{\sigma^2}^2}{F_D^2 F_{\sigma^2}^2+2F_D^2 F_V^2+2F_{\sigma^2}^2 F_V^2}\right)^{D/2}].
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\mathrm{EDF} \;\approx\; d &\approx \frac{2E[E(\hat{V}_j|\sigma_1^2,\sigma_2^2,\ldots,\sigma_{N_p}^2)|\sigma_0^2]^2}{E[\mathrm{Var}(\hat{V}_j|\sigma_1^2,\sigma_2^2,\ldots,\sigma_{N_p}^2)|\sigma_0^2]} \\
&= \frac{2(\sigma_0^2)^2}{\frac{2\sigma_0^4}{\eta}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2 + \frac{4\sigma_0^4}{\eta\nu}\sum_{k=1}^{N_p}\sum_{l=1}^{N_p}w_{jk}w_{jl}\rho_{kl}^2 r_{kl}^2} \\
&= \frac{2\sigma_0^4}{\frac{2\sigma_0^4}{\eta}\left(\frac{F_D^2}{F_D^2+2F_V^2}\right)^{D/2}+\frac{4\sigma_0^4}{\eta\nu}\left(\frac{F_D^2 F_{\sigma^2}^2}{F_D^2 F_{\sigma^2}^2+2F_D^2 F_V^2+2F_{\sigma^2}^2 F_V^2}\right)^{D/2}} \\
&= \frac{\eta\nu}{\nu\left(\frac{F_D^2}{F_D^2+2F_V^2}\right)^{D/2}+2\left(\frac{F_D^2 F_{\sigma^2}^2}{F_D^2 F_{\sigma^2}^2+2F_D^2 F_V^2+2F_V^2 F_{\sigma^2}^2}\right)^{D/2}}.
\end{aligned}
$$

B.3.4   Bias estimation

Since $\mathrm{MSE}(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2) = (\mathrm{bias}^2(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2)) + \mathrm{Var}(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2)$,

$$
\begin{aligned}
\mathrm{bias}^2(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2) &= \mathrm{MSE}(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2) - \mathrm{Var}(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2) \\
&= [E(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2) - \sigma_j^2]^2 \\
&= \left(\sum_{k=1}^{N_p}w_{jk}\sigma_k^2 - \sigma_j^2\right)^2 \\
&= (\sum_{k=1}^{N_p}w_{jk}\sigma_k^2 - \sigma_j^2)^2,
\end{aligned}
$$

(B.4)

where $\displaystyle\sum_{k=1}^{N_p} w_{jk} = 1 \ (j = 1, 2, \ldots, N_p)$.

The unconditional bias for the jth voxel is,

$$
\begin{aligned}
(\text{bias}_j|\sigma_0^2)^2 &= (E(\hat{V}_j|\sigma_0^2) - \sigma_0^2)^2 \\
&= (\sigma_0^2 - \sigma_0^2)^2 \\
&= 0.
\end{aligned}
$$

Let

$$
\underline{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_{N_p}^2),
$$

$$
\underline{w}_j = \begin{pmatrix} w_{j1} \\ w_{j2} \\ \vdots \\ w_{jN_p} \end{pmatrix},
$$

and,

$$
\underline{e}_j = \begin{pmatrix} 0 \\ \vdots \\ 1(\leftarrow j) \\ \vdots \\ 0 \end{pmatrix}.
$$

Then, Eq. (B.4) can be expressed as,

$$
\text{bias}^2(\hat{V}_j|\sigma_1^2, \ldots, \sigma_{N_p}^2) = (\underline{\sigma}^2(\underline{w}_j - \underline{e}_j))^2 = (\underline{w}_j - \underline{e}_j)^t (\underline{\sigma}^2)^t (\underline{\sigma}^2)(\underline{w}_j - \underline{e}_j).
$$

Hence

$$
E(\text{bias}^2(\hat{V}_j|\sigma_1^2, \ldots, \sigma_{N_p}^2)|\sigma_0^2) = (\underline{w}_j - \underline{e}_j)^t E((\underline{\sigma}^2)^t (\underline{\sigma}^2)|\sigma_0^2)(\underline{w}_j - \underline{e}_j),
$$

$$
E((\underline{\sigma}^2)^t(\underline{\sigma}^2)|\sigma_0^2) = E\left[\left(\begin{array}{cccc} \sigma_1^4 & \sigma_1^2\sigma_2^2 & \ldots & \sigma_1^2\sigma_{N_p}^2 \\ \sigma_2^2\sigma_1^2 & \sigma_2^4 & \ldots & \sigma_2^2\sigma_{N_p}^2 \\ \ldots & \ldots & \ddots & \ldots \\ \sigma_{N_p}^2\sigma_1^2 & \sigma_{N_p}^2\sigma_2^2 & \ldots & \sigma_{N_p}^4 \end{array}\right)\middle|\middle|\sigma_0^2\right]
$$

$$
= \left(\begin{array}{cccc} \frac{\sigma_0^4}{n^2}(n^2+2n) & \frac{\sigma_0^4}{n^2}(n^2+2nr_{12}^2) & \ldots & \frac{\sigma_0^4}{n^2}(n^2+2nr_{1N_p}^2) \\ \frac{\sigma_0^4}{n^2}(n^2+2nr_{21}^2) & \frac{\sigma_0^4}{n^2}(n^2+2n) & \ldots & \frac{\sigma_0^4}{n^2}(n^2+2nr_{2N_p}^2) \\ \ldots & \ldots & \ddots & \ldots \\ \frac{\sigma_0^4}{n^2}(n^2+2nr_{N_p1}^2) & \frac{\sigma_0^4}{n^2}(n^2+2nr_{N_p2}^2) & \ldots & \frac{\sigma_0^4}{n^2}(n^2+2n) \end{array}\right)
$$

$$
= \sigma_0^4\left(\begin{array}{cccc} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \ldots & \ldots & \ddots & \ldots \\ 1 & 1 & \ldots & 1 \end{array}\right) + \frac{2\sigma_0^4}{n}\left(\begin{array}{cccc} 1 & r_{12}^2 & \ldots & r_{1N_p}^2 \\ r_{21}^2 & 1 & \ldots & r_{2N_p}^2 \\ \ldots & \ldots & \ddots & \ldots \\ r_{N_p1}^2 & r_{N_p2}^2 & \ldots & 1 \end{array}\right).
$$

Here, we denote

$$
R = \left(\begin{array}{cccc} 1 & r_{12}^2 & \ldots & r_{1N_p}^2 \\ r_{21}^2 & 1 & \ldots & r_{2N_p}^2 \\ \ldots & \ldots & \ddots & \ldots \\ r_{N_p1}^2 & r_{N_p2}^2 & \ldots & 1 \end{array}\right).
$$

In addition,

$$(\underline{w}_j - \underline{e}_j)^t \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \dots & \dots & \ddots & \dots \\ 1 & 1 & \dots & 1 \end{pmatrix} (\underline{w}_j - \underline{e}_j)$$

$$= \begin{pmatrix} w_{j1} & \dots & w_{jj} - 1 & \dots & w_{jN_p} \end{pmatrix} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \dots & \dots & \ddots & \dots \\ 1 & 1 & \dots & 1 \end{pmatrix} (\underline{w}_j - \underline{e}_j)$$

$$= \begin{pmatrix} \sum_{k=1}^{N_p} w_{jk} - 1 & \sum_{k=1}^{N_p} w_{jk} - 1 & \dots & \sum_{k=1}^{N_p} w_{jk} - 1 \end{pmatrix} (\underline{w}_j - \underline{e}_j) = 0.$$

Therefore

$$E(\text{bias}^2(\hat{V}_j|\sigma_1^2, \dots, \sigma_{N_p}^2)|\sigma_0^2)$$

$$= \frac{2\sigma_0^4}{\nu}(\underline{w}_j - \underline{e}_j)^t R(\underline{w}_j - \underline{e}_j)$$

$$= \frac{2\sigma_0^4}{\nu}[\sum_{l=1}^{N_p}\sum_{k=1}^{N_p} w_{jl}w_{jk}r_{kl}^2 - \sum_{k=1}^{N_p} w_{jk}r_{jk}^2 - \sum_{l=1}^{N_p} w_{jk}r_{kj}^2 + 1]$$

$$= \frac{2\sigma_0^4}{\nu}[\sum_{l=1}^{N_p}\sum_{k=1}^{N_p} w_{jl}w_{jk}r_{kl}^2 - 2\sum_{k=1}^{N_p} w_{jk}r_{jk}^2 + 1]$$

$$= \frac{2\sigma_0^4}{\nu}\left[\left(\frac{F_{\sigma^2}^2}{F_{\sigma^2}^2 + 2F_V^2}\right)^{D/2} - 2\left(\frac{F_{\sigma^2}^2}{F_{\sigma^2}^2 + F_V^2}\right)^{D/2} + 1\right]$$

$$= \frac{2\sigma_0^4}{\nu}\left\{\left[1 + 2\left(\frac{F_V}{F_{\sigma^2}}\right)^2\right]^{-D/2} - 2\left[1 + \left(\frac{F_V}{F_{\sigma^2}}\right)^2\right]^{-D/2} + 1\right\}.$$

We show that bias^2 is bounded. For simplicity, we assume that D=2, and let

$$\text{RHS} = \left(\frac{F_{\sigma^2}^2}{F_{\sigma^2}^2 + 2F_V^2}\right)^{D/2} - 2\left(\frac{F_{\sigma^2}^2}{F_{\sigma^2}^2 + F_V^2}\right)^{D/2} + 1.$$

Then we prove the RHS is bounded within [0,1).

$$
\begin{aligned}
\text{RHS} &= \left(\frac{F_{\sigma^2}^2}{F_{\sigma^2}^2 + 2F_V^2}\right)^{D/2} - 2\left(\frac{F_{\sigma^2}^2}{F_{\sigma^2}^2 + F_V^2}\right)^{D/2} + 1 \\
&= \frac{F_{\sigma^2}^2}{F_{\sigma^2}^2 + 2F_V^2} - 2\frac{F_{\sigma^2}^2}{F_{\sigma^2}^2 + F_V^2} + 1 \\
&= \frac{2F_V^4}{(F_{\sigma^2}^2 + 2F_V^2)(F_{\sigma^2}^2 + F_V^2)} \\
&= \frac{2F_{\sigma^2}^4}{F_V^4 + 3F_{\sigma^2}^2 F_V^2 + 2F_V^4}.
\end{aligned}
$$

Hence when $F_V = 0$, RHS $= 0$, and when $F_V \gg F_{\sigma^2}$, RHS $\to 1$.

Then, for the entireimage

$$
\begin{aligned}
\text{bias}^2 &= \frac{1}{N_p}\sum_{j=1}^{N_p} E[\text{bias}^2(\hat{V}_j|\sigma_1^2,\ldots,,\sigma_{N_p})|\sigma_0^2] \\
&= \frac{2\sigma_0^4}{\nu}\left\{\left[1 + 2\left(\frac{F_V}{F_{\sigma^2}}\right)^2\right]^{-D/2} - 2\left[1 + \left(\frac{F_V}{F_{\sigma^2}}\right)^2\right]^{-D/2} + 1\right\},
\end{aligned}
$$

$$
\begin{aligned}
\text{MSE} &= \frac{1}{N_p}\sum_{j=1}^{N_p}[E(E(\text{Var}(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2)|\sigma_0^2)) + E(\text{bias}^2(\hat{V}_j|\sigma_1^2,\ldots,\sigma_{N_p}^2)|\sigma_0^2)] \\
&= \frac{2\sigma_0^4}{\eta\nu}\{\nu\left[1 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2} \\
&\quad + 2\left[1 + 2\left(\frac{F_V}{F_{\sigma^2}}\right)^2 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2}\} \\
&\quad + \frac{2\sigma_0^4}{\nu}\left\{\left[1 + 2\left(\frac{F_V}{F_{\sigma^2}}\right)^2\right]^{-D/2} - 2\left[1 + \left(\frac{F_V}{F_{\sigma^2}}\right)^2\right]^{-D/2} + 1\right\}.
\end{aligned}
$$

1. When $F_V = 0$, i.e. no smoothing, MSE $= \text{Var}(\hat{V}|\sigma_0^2)$, and bias$^2 = 0$.

2. When $F_V \searrow$ (decrease), VAR will increase and bias^2 will be decrease.

3. WHen $F_V \nearrow$ (increase), VAR will decrease and bias^2 will be increase.

B.3.5 Summary

In summary, when the true variance image $\boldsymbol{\Sigma}$ is spatially heterogeneous, EDF, MSE, Bias^2 and VAR are

$$
\begin{aligned}
\text{EDF} \;\approx\;& \frac{\eta\nu}{\nu\left(\frac{F_D^2}{F_D^2+2F_V^2}\right)^{D/2} + 2\left(\frac{F_D^2 F_{\sigma^2}^2}{F_D^2 F_{\sigma^2}^2 + 2F_D^2 F_V^2 + 2F_V^2 F_{\sigma^2}^2}\right)^{D/2}} \\
\;=\;& \frac{\eta\nu}{\nu\left[1 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2} + 2\left[1 + 2\left(\frac{F_V}{F_{\sigma^2}}\right)^2 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2}}.
\end{aligned}
$$

Or

$$
1/\text{EDF} \approx \frac{1}{\eta}\left[1 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2} + \frac{2}{\eta\nu}\left[1 + 2\left(\frac{F_V}{F_{\sigma^2}}\right)^2 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2},
$$

$$
\begin{aligned}
\text{VAR} \;=\;& \frac{2\sigma_0^4}{\eta\nu}\Big\{\nu\left[1 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2} \\
& + 2\left[1 + 2\left(\frac{F_V}{F_{\sigma^2}}\right)^2 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2}\Big\},
\end{aligned}
$$

$$
\text{bias}^2 \;=\; \frac{2\sigma_0^4}{\nu}\left\{\left[1 + 2\left(\frac{F_V}{F_{\sigma^2}}\right)^2\right]^{-D/2} - 2\left[1 + \left(\frac{F_V}{F_{\sigma^2}}\right)^2\right]^{-D/2} + 1\right\},
$$

$$
\begin{aligned}
\text{MSE} \;=\;& \frac{2\sigma_0^4}{\eta\nu}\Big\{\nu\left[1 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2} \\
& + 2\left[1 + 2\left(\frac{F_V}{F_{\sigma^2}}\right)^2 + 2\left(\frac{F_V}{F_D}\right)^2\right]^{-D/2}\Big\} \\
& + \frac{2\sigma_0^4}{\nu}\left\{\left[1 + 2\left(\frac{F_V}{F_{\sigma^2}}\right)^2\right]^{-D/2} - 2\left[1 + \left(\frac{F_V}{F_{\sigma^2}}\right)^2\right]^{-D/2} + 1\right\}.
\end{aligned}
$$

Two specific cases, complete smoothing and no smoothing for random field variance images

Case 1: $w_{jk} = \frac{1}{N_p}$, homogeneous smoothing (complete smoothing),

$$
\text{EDF} \;\approx\; d \approx \frac{2E(\hat{V}_j)^2}{\text{Var}(\hat{V}_j)} = \frac{\eta\nu N_p^2}{\nu\sum_{k,l}^{N_p}\rho_{kl}^2 + 2\sum_{k,l}^{N_p}\rho_{kl}^2 r_{kl}^2}.
$$

If we assume

$$r_{kl} = \exp\{-\|\mathbf{P}_k - \mathbf{P}_l\|_2^2/(4s_v^2)\}, \ \ s_v^2 = F_{\sigma^2}^2/(8\log 2),$$

then

$$\text{EDF} \approx \frac{\eta\nu N_p^2}{\nu \left(2\pi\frac{2F_D^2}{16\log 2}\right)^{D/2} + 2\left(2\pi\frac{2F_D^2 F_{\sigma^2}^2}{16\log 2(F_D^2+F_{\sigma^2}^2)}\right)^{D/2}},$$

where $N_p \gg \nu \gg \eta$, $\text{EDF} \longrightarrow +\infty$,

Case 2: When $w_{jj} = 1$, and $w_{jk} = 0$ if $j \neq k$ (no variance smoothing)

$$\begin{aligned}
\text{EDF} &\approx& d = \frac{2E(\hat{V}_j)^2}{\text{Var}(\hat{V}_j)} = \frac{\eta\nu}{\nu+2}, \\
\text{VAR} &=& \frac{2\sigma_0^4}{\eta\nu}(\nu+2), \\
\text{MSE} &=& \text{VAR} = \frac{2\sigma_0^4}{\eta\nu}(\nu+2), \\
\text{Bias}^2 &=& 0.
\end{aligned}$$

When $n \gg \eta$, $\text{SmVarDF} \longrightarrow \eta$, the degrees of freedom of the sample variance before smoothing.

B.4 Hot spot with constant variance images

Assume the true variance image is constant (σ_0^2), but there is a hot spot at a certain voxel (j) with variance σ_j^2 $(\sigma_j^2 \gg \sigma_0^2)$. We want to know the bias in $\hat{V}$ at the jth voxel.

$$V_j = \frac{1}{\eta}\sum_{i=1}^{\eta} X_{ji}^2,$$

where i indexes subjects, j indexes voxel.

Then

$$E(V_j|\sigma_j^2) = \frac{1}{\eta}\sum_{i=1}^{\eta} E(X_{ji}^2|\sigma_j^2) = \sigma_j^2.$$

After the variance is smoothed, the smoothed variance at the jth voxel is

$$\hat{V}_j = \sum_{k=1}^{N_p} w_{jk} V_k \text{ , and } \sum_{k=1}^{N_p} w_{jk} = 1.$$

Then,

$$E(\hat{V}_j | \sigma_0^2, \sigma_j^2) = \sum_{k=1}^{N_p} w_{jk} E(V_k | \sigma_0^2, \sigma_j^2) = \sum_{k=1}^{N_p} w_{jk} \sigma_0^2 + w_{jj}(\sigma_j^2 - \sigma_0^2) = \sigma_0^2 + w_{jj}(\sigma_j^2 - \sigma_0^2)$$

Therefore,

$$
\begin{aligned}
(\text{bias}_j | \sigma_0^2, \sigma_j^2)^2 &= \text{MSE}(\hat{V}_j | \sigma_0^2, \sigma_j^2) - \text{Var}(\hat{V}_j | \sigma_0^2, \sigma_j^2) \\
&= (E(\hat{V}_j | \sigma_j^2, \sigma_0^2) - \sigma_j^2)^2 \\
&= (\sigma_0^2 + w_{jj}(\sigma_j^2 - \sigma_0^2) - \sigma_j^2)^2 \\
&= (1 - w_{jj})^2 \cdot (\sigma_j^2 - \sigma_0^2)^2 .
\end{aligned}
$$

References

References

Adler, R.J. 1981. *The Geometry of Random Fields.* New York: Wiley.

Bullmore, E.T., Brammer, M., Williams, S., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., & Sham, P. 1996. Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance Medicine*, **35**, 261–277.

Bullmore, E.T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M.J. 1999. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans. Med. Imaging*, **18**, 32–42.

Cao, J. 1999. The size of the connected components of excursion sets of χ^2, t, and F fields. *Advances in Applied Probability*, **31**, 579–595.

Cao, J., & Worsley, K.J. 2001. *Applications of random fields in human brain mapping, in: Moore, M.(Ed), Spatial Statistics: Methodological Aspects and Applications, Springer lecture Notes in Statistics.* Vol. 159. New York: Springer. Pages 169–182.

Chee, M., Venkatraman, V., Westphal, C., & Siong, S.C. 2003. Comparison of Block and Event-Related fMRI designs in Evaluating the Word-Frequency Effect. *Human Brain Mapping*, 186–193.

Forman, S.D., Cohen, J.D., Fitzgerald, J.D., Eddy, W.F., Mintun, M.A., & Noll, D.C. 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance Medicine*, **33**, 636–647.

Friston, K.J. 1997. Testing for anatomically specified regional effects. *Human Brain Mapping*, **5**, 133–136.

Friston, K.J., Frith, C.D., Liddle, P.F., & Frackowiak, R.S.J. 1991. Comparing functional

(PET) images: The assessment of significant change. *J. Cereb. Blood Flow Metab.*, **11**, 690–699.

Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., & Evans, A.C. 1994. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, **1**, 210–220.

Friston, K.J., Holmes, A., Poline, J.-B, Price, C.J., & Frith, C.D. 1996. Detecting activations in PET and fMRI: levels of inference and power. *NeuroImage*, **4**, 223–235.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A.P., Rugg, M.D., & Turner, R. 1998. Event-related fMRI: Characterising differential responses. *Neuroimage*, **7**, 30–40.

Friston, K.J., Zarahn, E., Josephs, O., Henson, R.N., & Dale, A.M. 1999. Stochastic designs in event-related fMRI. *Neuroimage*, **10**, 607–619.

Gordon, R.D. 1941. Values of mill's ratio of area to bounding ordinate of the normal probability integral for large values of the argument. *Annals of Mathematical Statistics*, **12**, 364–366.

Hayasaka, S., & Nichols, T.E. 2003. Validating cluster size inference: Random field and permutation methods. *Neuroimage*, **20(4)**, 2343–2356.

Hayasaka, S., & Nichols, T.E. 2004. Combining voxel intensity and cluster extent with permutaion test framework. *Neuroimage*, **23**, 54–63.

Hayasaka, S., Phan, K.L., Liberzon, I., Worsley, K.J., & T.E., Nichols. 2004. Nonstationary cluster size inference with random field and permutation methods. *Neuroimage*, **22(2)**, 676–687.

Holmes, Andrew. 1994. *Statistica Issues in functional Brain Mapping*. Ph.D. thesis, University of Glasgow.

Jenkinson, M. 2000. *Estimation of Smoothness from the Residual Field*. Tech. rept. Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB).

Kiebel, S., J.B., Poline, Friston, K.J., Holmes, A., & Worsley, K.J. 1999. Robust smooth-

ness estimation in statistical parametric maps using standardized residuals from the general linear model. *Neuroimage*, **10**, 756–766.

Kullback, S., & Leibler, R.A. 1951. On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

Kwong, K.K., Belliveau, J.W., Chesler, D.A., I.E., Goldberg, Weisskoff, R.M., Poncelet, B.P., D.N., Kennedy, Hoppel, B.E., Cohen, M.S., Turner, R., Cheng, H., Brady, T.J., & Rosen, B.R. 1992. Dynamic Magnetic Resonance Imaging of Human Brain Activity During Primary Sensory Stimulation. *PNAS*, **89**, 5675–79.

Lange, N., & Zeger, S.L. 1997. Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonanace imaging (with Discussion). *Applied Statistics*, **46**, 1–29.

Lazar, N.A., Luna, B., Sweeney, J.A., & Eddy, W.F. 2002. Combining brains: a survey of methods for statitical pooling of information. *Neuroimage*, **16**, 538–550.

Lukic, A.S., Wernich, M.N., & Strother, S.C. 2002. An evaluation of methods for detecting brain activations from functional neuroimages. *Artiicial Intelligence in Medicine*, **1**, 69–88.

Luo, W.-L., & Nichols, T.E. 2003. Diagnosis and Exploration of Massively Univariate fMRI Models. *NeuroImage*, **19(3)**, 1014–1–32.

Madic, & Group. 2005. *Functional Imaging Analysis Contest (FIAC).* `http://www.madic.org/fiac/how_to_participate.html`.

Marchini, J., & Ripley, B. 2000. A new statistical approach to detecting significant activation in functional MRI. *NeuroImage*, **12**, 366–380.

Marshuetz, C., Smith, E.E., Jonides, J., DeGutis, J., & Chenevert, T.L. 2000. Order information in working memory:fMRI evidence for partietal and prefrontal mechanism. *J. Cogn. Neurosci.*, **12(S2)**, 130–144.

Nichols, T.E., & Hayasaka, S. 2003. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, **12**,

419–446.

Nichols, T.E., & Holmes, A.P. 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, **15**, 1–25.

Ogawa, S., & Lee, T.M. 1992. Functional brain imagining with physiologically sensitivie image signals. *Journal of Magnetic Resonance Imaging*, **2(P)-WIP (Suppl)**, S22.

Ogawa, S., Lee, T.M., Nayak, A.S., & Glynn, P. 1990. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Mangnetic Resonance Medicine*, **14**, 68–78.

Pesarin, F. 2001. *Multivariate Permutation Tests.* New York: Wiley.

Poline, J.B., Worsley, K.J., Evans, A.C., & Friston, K.J. 1997. Combining Spatial Extent and Peak Intensity to Test for Activations in Functional Imaging. *Neuroimage*, **5**, 83–96.

Roy, C.S., & Sherrington, C.S. 1890. On the regulation of the blood-supply of the brain. *Journal of Physiology*, **11(1-2)**, 85–158.

Siegmund, D.O., & Worsley, K.J. 1995. Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Annals of Statistics*, **23**, 608–639.

Soltysik, D.A., & J.S., Hyde. 2006. Strategies for block-design fMRI experiments during task-related motion of structures of the oral cavity. *Neuroimage*, **29**, 1260–1271.

Springer, C.S., Patlak, C.S., Palyka, I., & Huang, W. 1999. Principles of susceptibility contrast-based functional MRI: The sign of the functional MRI response. *Functional MRI*, 91–102.

Woolrich, M.W., Ripley, B.D., Brady, M., & Smith, S.M. 2001. Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *Neuroimage*, **14**, 1370–1386.

Worley, K.J. 1994. Local maxima and the expected Euler characteristic of excursion sets of χ^2, f and t fields. *Advances in Applied Probability*, **26**, 13–42.

Worsley, K.J., Evans, A.C., Marrett, S., & Neelin, P. 1992. Three-dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab.*,

12, 900–918.

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., & Evans, A.C. 1996. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, **4**, 58–73.

Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., & Evans, A.C. 2002. A General Statistical Analysis for fMRI Data. *NeuroImage*, **5**, 179–197.

Yendiki, A., & Fessler, J.A. 2007. Analysis of observer performance in unknown-location tasks for tomographic image reconstruction. *Journal of the Optical Society of America A*, **24(12)**, B99–109.