

# Modelling Voxel Dependent Hemodynamic Response Function

Darren Fletcher

A thesis submitted for the degree of  
Doctor of Philosophy

School of Mathematics and Statistics

Newcastle University

Supervisor: Professor Robin Henderson

August 10, 2020



## Abstract

An important challenge of contemporary neuroscience is the detection and understanding of significant brain activity using functional magnetic resonance imaging (fMRI). One of the many motivations of this research, related to the data set used in this thesis, is to investigate brain activation and connectivity patterns aimed at identifying associations between these patterns and regaining motor functionality following a stroke. Much statistical modelling has attempted to interpret noisy fMRI data and detect changes in response to activity. However due to the large data sets usually involved in fMRI modelling, here as many as 150,000 measurements in localised spatial volumes known as voxels at each time point, many simplifying assumptions are usually made to make computation feasible. This is known to have a negative impact on detecting voxel activation. In this work we fit a space-time model to a fMRI data set using a sequential approach to allow for scalability. However, the main contribution of this work is an alternative method to detect activation in the brain. Here we take the novel approach of using topological data analysis to investigate the model residuals to detect changes in the fMRI data. In particular we analyse the spatial distribution of topological features of the residuals to provide a test for normality, and also by providing a method to analyse how the spatial distribution of such features change over time, we are able to detect changes in the data in response to activity where conventional methods cannot. A recommendation for future work is to also investigate how topological features change for different filtration levels of the field, as this may provide new insights on brain activation.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Exploratory Data Analysis</b>	<b>2</b>
2.1	Introduction To The Data . . . . .	2
2.2	Brain Anatomy . . . . .	2
2.3	Exploring The Data . . . . .	4
2.3.1	The Voxels . . . . .	4
2.3.2	The ROIs . . . . .	9
2.3.3	The Ljung-Box Test . . . . .	12
<b>3</b>	<b>Modelling</b>	<b>15</b>
3.1	Previously Proposed Models . . . . .	15
3.2	The Model Covariates . . . . .	17
3.3	Experiment Design And Modeling Issues Identified . . . . .	20
3.3.1	Collinearity - Sequentially Adding Parameters . . . . .	20
3.3.2	Collinearity - The Reciprocal Condition Number . . . . .	21
3.4	The Mean Structure . . . . .	23
3.4.1	Different Covariates For Each Period Of Exercise . . . . .	24
3.5	The Independent Model . . . . .	27
3.5.1	Simulating Independent Data . . . . .	28
3.6	The Temporal Model . . . . .	29
3.6.1	The Autoregressive Parameters . . . . .	29
3.6.2	Procedure To Fit The Model . . . . .	33
3.6.3	Simulating Temporal Data . . . . .	35
3.6.4	Restricted Maximum Likelihood Estimation . . . . .	37
3.7	The Spatio-Temporal Model . . . . .	40
3.7.1	Procedure To Fit The Model . . . . .	41
3.7.2	Simulating Spatio-Temporal Data . . . . .	42
3.8	Application To Brain Data . . . . .	44
3.8.1	The Mean Parameters . . . . .	44
3.8.2	The Autoregressive Parameters . . . . .	47
3.8.3	The Spatial Parameter: Estimating The Spatial Co- variance Using Different Functions . . . . .	48
3.9	Residual Analysis . . . . .	57
3.9.1	Residuals Autocorrelation . . . . .	57
3.9.2	Residual Noise . . . . .	58

3.9.3	Temporal Correlation Between Residuals . . . . .	64
<b>4</b>	<b>Testing For Activation</b>	<b>67</b>
4.1	Introduction . . . . .	67
4.2	Multiple Hypothesis Testing . . . . .	68
4.2.1	Bonferroni Correction . . . . .	68
4.2.2	Benjamini-Hochberg Procedure . . . . .	69
4.2.3	Example Of How The Procedures Differ . . . . .	70
4.3	Brain Simulations . . . . .	71
4.3.1	Correlation Between Test Statistics . . . . .	71
4.3.2	Testing For Activation . . . . .	73
4.4	Application To Brain Data . . . . .	75
<b>5</b>	<b>Clustering</b>	<b>79</b>
5.1	Clustering Of The Residual Regional Means . . . . .	79
5.1.1	The Number Of Clusters . . . . .	79
5.1.2	Clustering Results . . . . .	82
5.1.3	Within Cluster And Between Cluster Correlation . . . . .	83
5.1.4	Linear Discriminant Analysis . . . . .	84
5.2	Clustering Of Raw Data . . . . .	88
5.2.1	Raw Regional Means . . . . .	88
5.2.2	Raw Centered Regional Means . . . . .	90
5.2.3	Clustering Using Number Of Voxels In Each Region . . . . .	92
5.2.4	Clustering Using Mean (Over Time) Of Regional Means . . . . .	95
5.2.5	Clustering Using Standard Deviation Of Regional Means . . . . .	97
5.3	Clustering Using Brain Anatomy . . . . .	99
5.4	Comparing The Clusterings . . . . .	101
5.4.1	Rand Index . . . . .	101
5.4.2	Adjusted Rand Index . . . . .	103
5.4.3	Association Test . . . . .	107
5.4.4	Linear Discriminant Analysis . . . . .	109
5.5	Clustering Simulations . . . . .	110
5.6	Within Region Clustering . . . . .	112
5.7	Clustering Conclusions And Discussion . . . . .	114
<b>6</b>	<b>Topological Data Analysis</b>	<b>115</b>
6.1	Important TDA Concepts . . . . .	115
6.2	Persistence Diagrams . . . . .	123

6.3	Topology Of The Model Residuals - Full Brain . . . . .	132
6.3.1	Convex Peels And Landscape Functions . . . . .	132
6.3.2	Bottleneck Distances . . . . .	135
6.4	Topology Of The Model Residuals - Full Brain Standardised .	136
6.4.1	Convex Peels And Landscape Functions . . . . .	136
6.4.2	Bottleneck Distances . . . . .	139
6.5	Summary . . . . .	141
<b>7</b>	<b>TDA: The Number Of Components And Holes</b>	<b>143</b>
7.1	Expected Number Of Components And Holes For A Gaussian Random Field . . . . .	143
7.2	Simulations . . . . .	147
7.2.1	Simulation Parameters . . . . .	147
7.2.2	Images Of The Fields On Two Dimensional Lattices . .	150
7.2.3	Simulations On Regular Two Dimensional Lattices . .	155
7.2.4	Simulations On Regular Three Dimensional Lattices . .	173
7.2.5	Investigating Random Fields Package In R . . . . .	179
7.2.6	Simulations On Two Dimensional Brain Slices . . . . .	182
7.2.7	Simulations On Three Dimensional Full Brain . . . . .	189
7.3	Application . . . . .	190
7.3.1	Number Of Components And Holes On Brain Slices . .	191
7.3.2	Number Of Components And Voids On The Full Brain	196
7.4	Conclusion . . . . .	203
<b>8</b>	<b>Spatial TDA</b>	<b>204</b>
8.1	Introduction . . . . .	204
8.2	Complete Spatial Randomness - Quadrant Count Analysis . .	205
8.3	Complete Spatial Randomness - Ripley's K-function . . . . .	206
8.3.1	Edge Effects . . . . .	208
8.3.2	Confidence Envelopes . . . . .	209
8.3.3	Example: Homogeneous Data On A Square Region . .	209
8.3.4	Example: Clustered And Regular Data On A Square Region . . . . .	210
8.3.5	Testing For Homogeneity On A Lattice . . . . .	211
8.4	Comparing Spatial Point Patterns . . . . .	213
8.4.1	Minimum Distance Between Point Patterns . . . . .	213
8.4.2	Quadrant Count Analysis . . . . .	215
8.4.3	Cross K-function . . . . .	216

8.4.4	Confidence Envelopes . . . . .	217
8.5	Simulations I - Complete Spatial Randomness . . . . .	219
8.5.1	Quadrant Count Analysis On A Square Lattice . . . . .	219
8.5.2	K-function On A Square Lattice . . . . .	222
8.5.3	Comparing The K-function Of Different Distributions . . . . .	224
8.6	Simulations II - Comparing Point Patterns . . . . .	234
8.6.1	Minimum Distance Between Point Patterns - Cumulative Distribution Functions . . . . .	234
8.6.2	Minimum Distance Between Point Patterns - Probability Functions . . . . .	244
8.6.3	Quadrant Count Analysis . . . . .	247
8.6.4	Cross K-function . . . . .	251
8.7	Application I - Complete Spatial Randomness . . . . .	257
8.7.1	Quadrant Count Analysis On Brain Slices . . . . .	257
8.7.2	K-function On Brain Slices . . . . .	261
8.7.3	Testing For Homogeneity On The Full Brain . . . . .	266
8.8	Application II - Comparing Point Patterns . . . . .	270
8.8.1	Minimum Distance Between Point Patterns . . . . .	270
8.8.2	Comparing Gaussian And Brain Slice Point Patterns . . . . .	275
8.8.3	Cross K-function . . . . .	279
8.9	Conclusion . . . . .	281
<b>9</b>	<b>Conclusion</b>	<b>284</b>
9.1	Summary Of Findings And Contribution . . . . .	284
9.2	Suggestions For Future Work . . . . .	285
<b>10</b>	<b>Appendix</b>	<b>298</b>
10.1	The Regions Of Interest . . . . .	298
10.2	Model Residual Topology Summary Statistics . . . . .	299
10.3	Standardised Model Residual Topology Summary Statistics . . . . .	303
10.4	Images Of The Fields On Two Dimensional Lattices . . . . .	307
10.5	Locations Of Components And Holes Of The Fields On Two Dimensional Lattices . . . . .	310
10.5.1	$(\nu, \eta)=(0.5, 2)$ . . . . .	310
10.5.2	$(\nu, \eta)=(0.5, 5)$ . . . . .	314
10.5.3	$(\nu, \eta)=(1, 5)$ . . . . .	318
10.5.4	$(\nu, \eta)=(0.5, 20)$ . . . . .	320







# 1 Introduction

Detecting and understanding significant patterns of brain activity is among the most important challenges of contemporary science. To this end, functional magnetic resonance imaging (fMRI) has been at the forefront of neuroscience investigations for the past twenty years. Functional magnetic resonance imaging measures brain activity by detecting changes in neural activity associated with blood flow using the contrast between oxygenated and deoxygenated hemoglobin, which are diamagnetic and paramagnetic respectively. fMRI scans produce images of high resolution, taking as many as 150,000 measurements in localised spatial volumes known as voxels.

The statistical challenge is to develop a model which is able to detect the changes in neural activity in voxels or regions of the brain. Many works in to this area have been carried out and will be discussed in detail in Section 3.1. However the disadvantages of previously proposed models is the inability to model the spatial dependence between the large number of voxels due to computational limitations. The results are models which make simplifying assumptions, such as independence between observations at different voxels or regions. It has been shown, and will be demonstrated in this work, that such assumptions significantly increase the rate of falsely identifying activity in voxels.

In this work, we propose a model (Section 3) which captures the spatial and temporal correlation between observations at different voxels. A significant part of this thesis will be dedicated to determining if any interesting results can be found using the resulting model residuals. In Section 5 various clustering techniques will be applied, and several interesting results are found, such as a spatial partitioning of regions within the brain that appear to be very closely related to the different brain lobes.

However the main contribution will be a new method of detecting activation using the model residuals. In previous works, traditionally model parameters are estimated and inferences on activation are made. Although this will be explored in Section 4, we will use topological data analysis (TDA) techniques to detect activation within regions in the brain. The novel approach here is to analyse the spatial distribution of topological features (Section 8), and how these evolve over time to provide a method to detect neural activation in voxels or regions of a brain. The results will show that this method can reveal regions which are active, where other standard methods do not.

## 2 Exploratory Data Analysis

### 2.1 Introduction To The Data

This project is motivated from the work described in [31]. As described in this paper, the data set that is being worked with arose from a study at the neuro-rehabilitation laboratory of Dr. Cramer at the University of California at Irvine. The goal of the study was to investigate associations between motor functional deficits in stroke patients, and brain activation and connectivity. The study consisted of taking functional magnetic resonance images (fMRI) scans of a single subject over the course of an experiment. The experiment consisted of 48 scans, alternating between a task and rest condition, each lasting for a period of 12 scans. The task condition was a simple hand grasping exercise. The patient was a stroke affected, right handed male in the age group 18-35 years.

The data were collected using a Philips Achieva 3.0T MRI whole body scanner, where the repetition time between consecutive scans is 2 seconds and the resolution of the data points is  $2 \text{ mm}^3$ . The functional data was then preprocessed using SPM8 software (Wellcome Department of Imaging Neuroscience, [www.lion.ucl.ac.uk/spm](http://www.lion.ucl.ac.uk/spm)), which included steps such as image alignment, coregistration to the mean image and normalisation to a standard template.

The final data set consists of measurements at 228453 points in the brain, called voxels. Each of these voxels has an associated time series of length 48, which are the fMRI measurements for experiment. Also available is information of anatomically defined regions of interest (ROI) within the human brain (see Table 63 in the appendix). The ROIs being used are the 90 ROIs in the brain as described in [72] which are anatomically defined using the Anatomical Automatic Labeling (AAL) atlas, also available at [www.lion.ucl.ac.uk/spm](http://www.lion.ucl.ac.uk/spm). Of the 228453 voxels, 152195 are assigned to one of the 90 ROIs. More information regarding the ROIs, including some plots of some regions on the brain will be given in Section 2.3.2.

### 2.2 Brain Anatomy

In this section some background detail of the anatomy of the brain will be given, which will be referred to in later sections.

The brain consists of three main parts, known as the forebrain, midbrain

and hindbrain. The forebrain consists of the cerebrum, thalamus and hypothalamus. The midbrain consists of the tectum and tegmentum and the hindbrain is made of the cerebellum, pons and medulla. The midbrain, pons and medulla are collectively sometimes referred to as the brain stem [2].

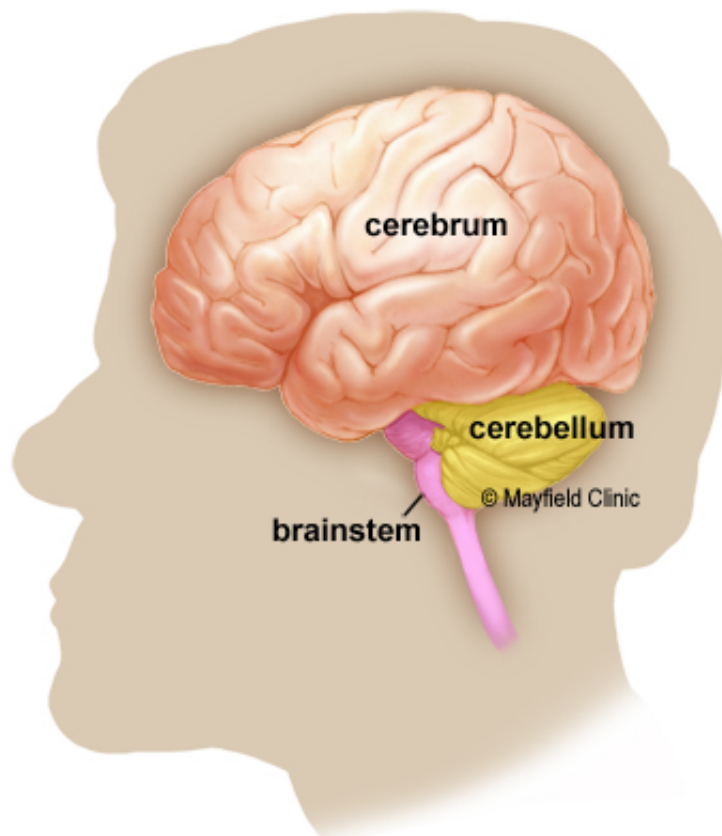


Figure 1: The three main parts of a human brain (image taken from [1])

The cerebrum is the part of the brain associated with higher brain functions such as thought, sensory perception and motor function. It is this part of the brain that we have data for. The cerebrum can be divided into smaller units called brain lobes. There are four main lobes in the forebrain, known as the frontal lobe (associated with speech, movement, emotions, problem solving), temporal lobe (audio processing), parietal lobe (movement, recognition, orientation, perception of stimuli) and occipital lobe (visual processing) [2]. These lobes can be seen in Figure 2.

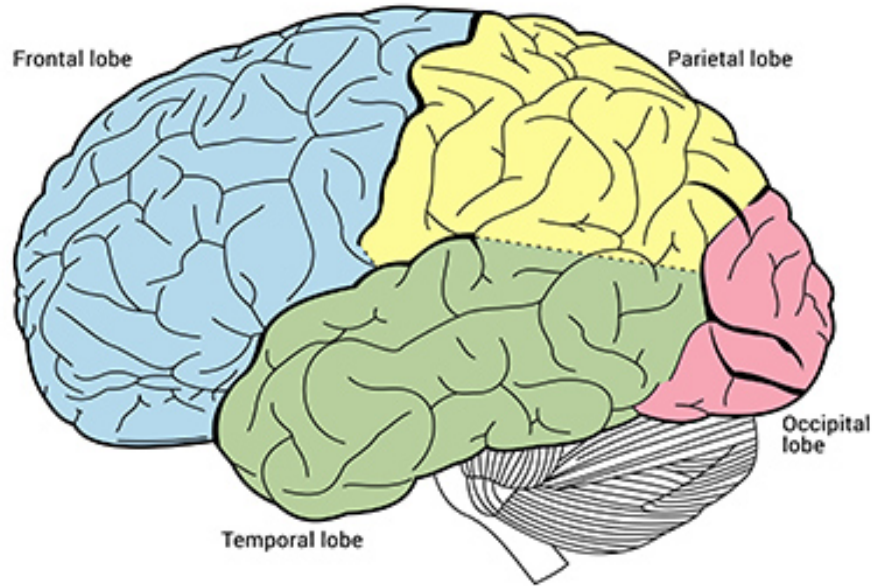


Figure 2: The four lobes of the forebrain (image taken from [7])

In later sections we will refer back to the brain anatomy presented here when investigating the model residuals, to determine if these natural partitions of the brain can be seen in the data.

## 2.3 Exploring The Data

### 2.3.1 The Voxels

At each voxel we have a time series of length 48, consisting of alternating periods of rest and task conditions lasting 12 scans each. Figure 3 shows an example of a voxel time series. The vertical dashed lines represent the partitions of the rest and activity periods. From the figure we can see that the response is very noisy and can vary a lot over the course of the experiment. It should also be noted that the magnitude and variability of the signal can vary substantially from voxel to voxel and in particular from region to region.

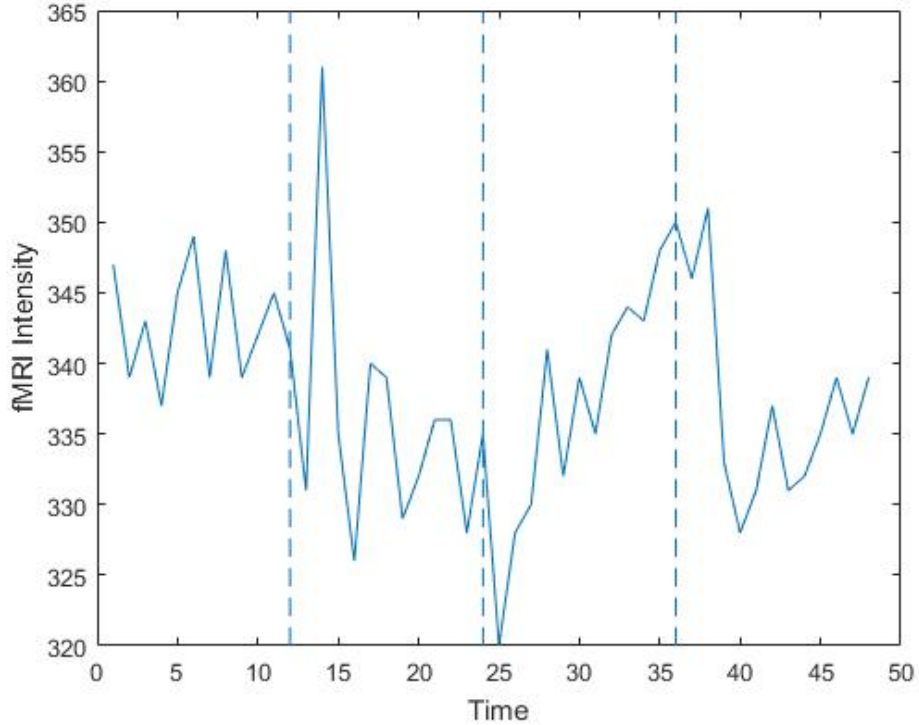


Figure 3: An example of a voxel time series.

Figures 4 and 5 shows three dimensional plots of all voxels at various time points. The coordinates range from 11 to 82 in the  $x$  direction, 11 to 100 in the  $y$  direction and 2 to 77 in the  $z$  direction. The fMRI measurements are displayed using a colour map where dark blue represents low values and lighter colours represent high values. From Figure 4 we note that very little difference can be seen at various time points, however note that from these images we can only see what is happening on the surface of the brain. If viewing online, Figure 5 can be clicked on to view how these measurements change at time changes from  $t = 1$  to  $t = 48$ , the length of the experiment. From Figure 5 we can see some areas of high activity (the lighter yellow patches) but little systematic change over time.

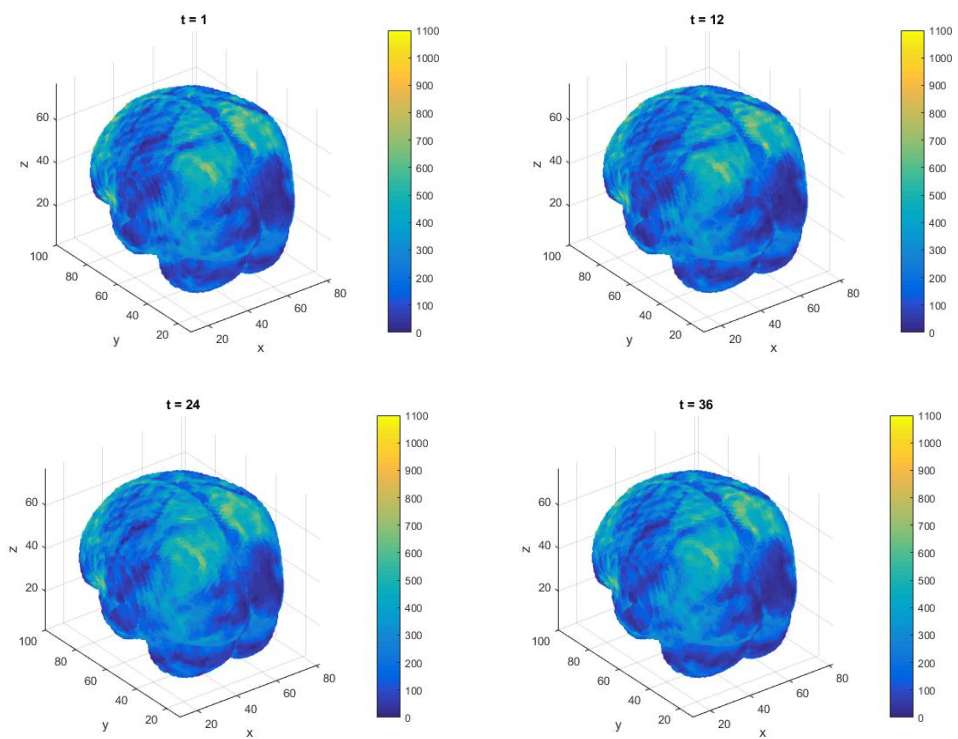


Figure 4: Four examples of the fMRI measurements at different time points on the full brain.



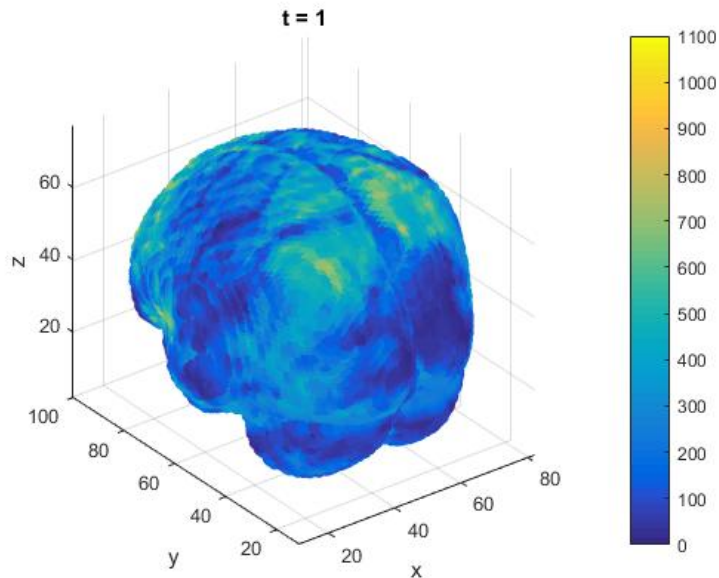


Figure 5: All voxel measurements from the fMRI dataset at time  $t = 1$ , light corresponds to high and dark to low measurements. Click the image to see how the measurements change from  $t = 1$  to  $t = 48$ .

Figure 6 shows four examples of fMRI measurements at all voxels at  $t = 1$  for four different fixed values of  $z$ , the vertical axis. From this figure we can see more information about brain activity in contrast to Figure 5, as well as how much it varies across different slices of the brain.

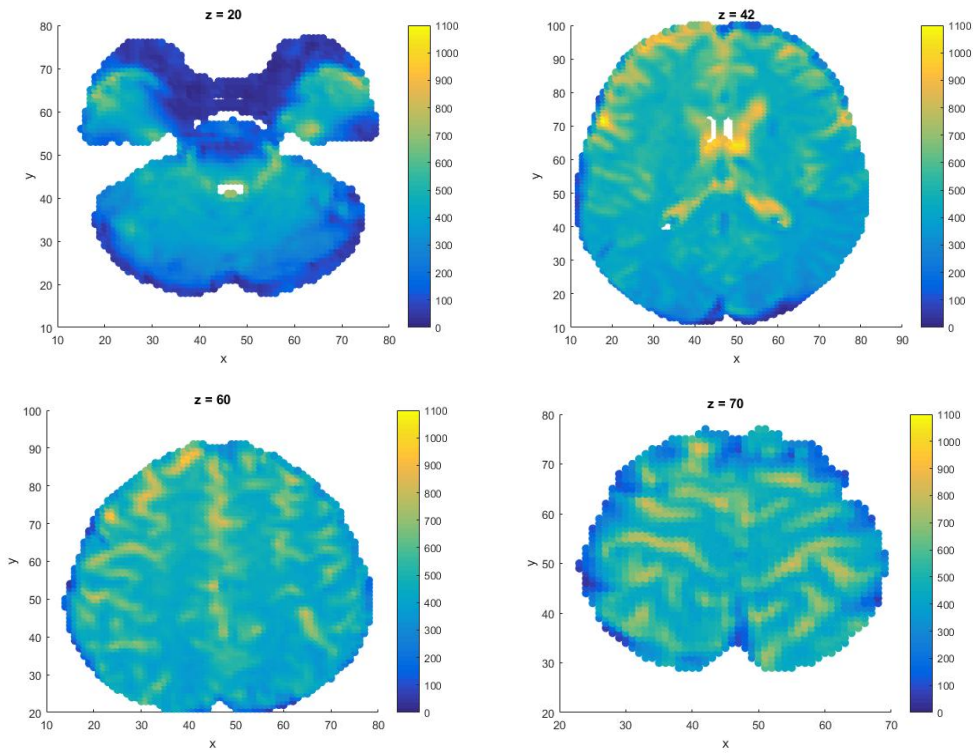


Figure 6: Four examples of the fMRI measurements at time  $t=1$  for voxels at four different fixed values of  $z$ .

Figure 7 shows how the fMRI measurements at all voxels with fixed  $z = 42$  change as time progresses from  $t = 1$  to  $t = 48$ .

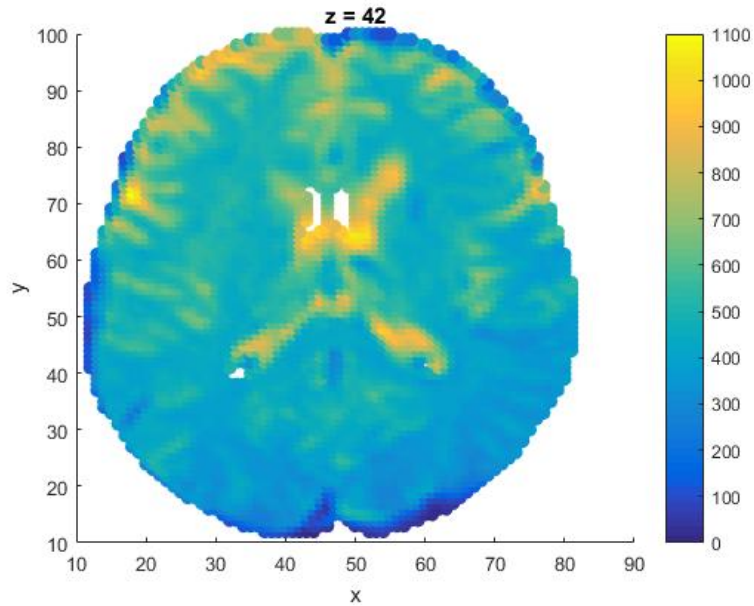


Figure 7: All voxel measurements from the fMRI dataset at time  $t=1$  for constant  $z=42$ , light corresponds to high and dark to low measurements (and white to no measurements). Click the image to see the how the measurements change from  $t=1$  to  $t=48$ .

### 2.3.2 The ROIs

As previously noted, there are 45 automatically defined ROIs in each hemisphere of the brain which are obtained using the Anatomical Automatic Labeling (AAL) atlas, see Appendix Table 63 for all ROI names. Figures 8-10 shows some examples of these regions.

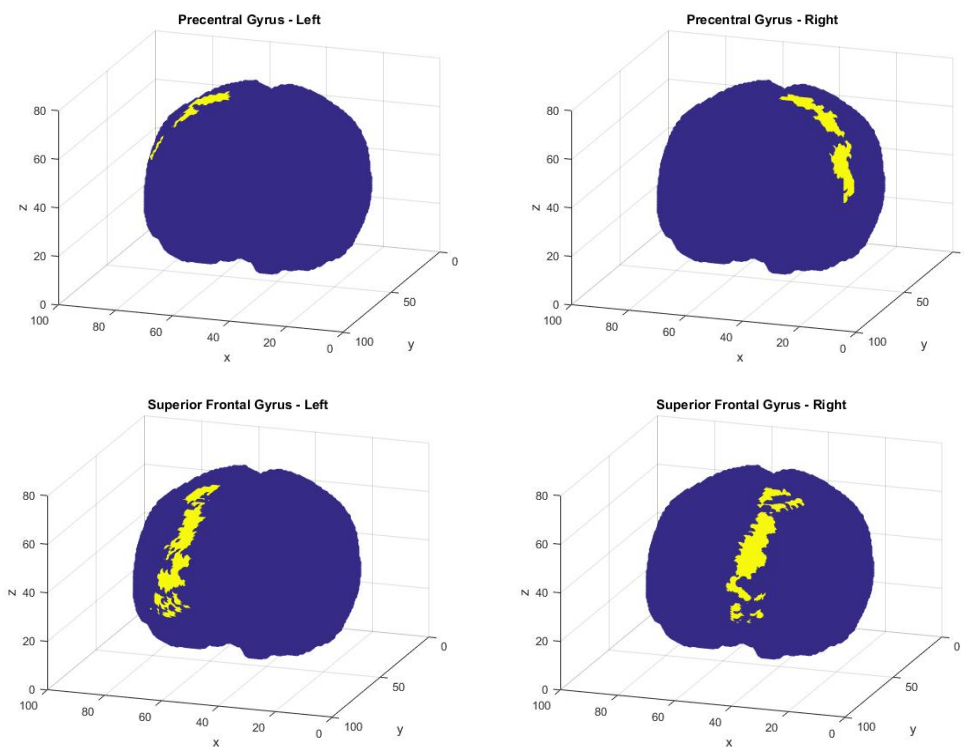


Figure 8: 4 of the 90 ROIs

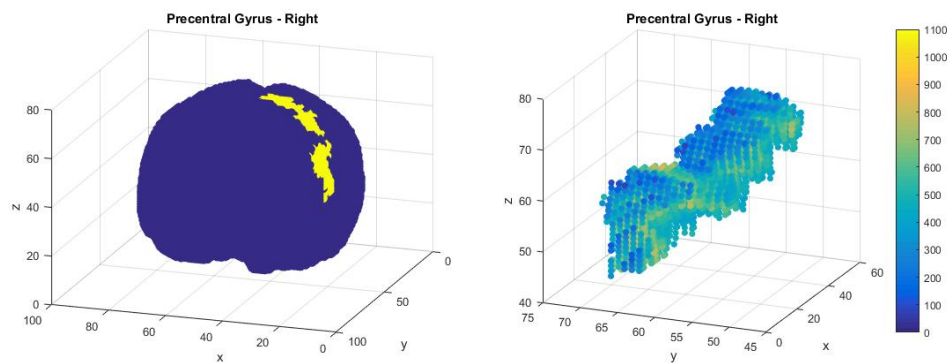


Figure 9: The Precentral Gyrus on the brain (left) and the voxel measurements in that region (right) at  $t=1$ .

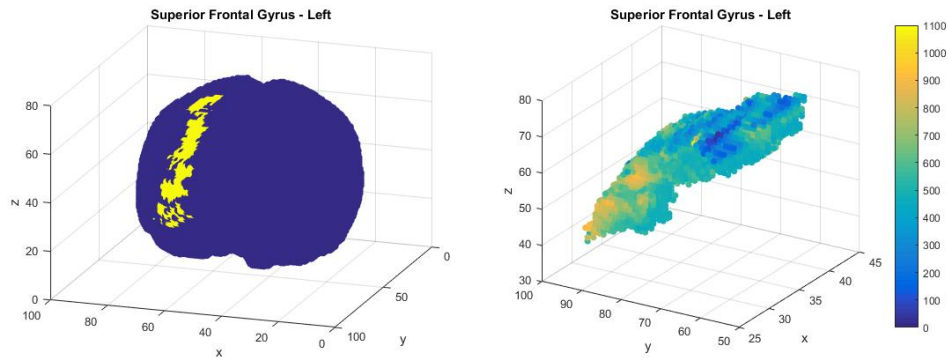


Figure 10: The Superior Frontal Gyrus on the brain (left) and the voxel measurements in that region (right) at  $t=1$ .

Of the total 228453 voxels, 152195 have an assigned region of interest label. Figure 11 (left) shows these voxels highlighted in yellow, with the voxels not assigned to a region in blue, and Figure 11 (right) shows these regions in different colours. Note only the regions on the surface of the brain and from the angle in the plot can be seen.

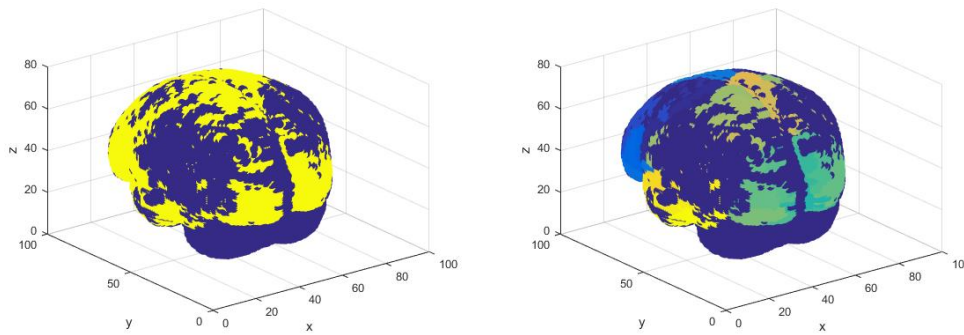


Figure 11: All regions of interest (left - yellow), and those regions separated by colour (right).

For an initial idea of the spatial correlation between the regions of interest, Figure 12 shows a simple measure of correlation between the regions of interest. To calculate this, first a mean time series was calculated for each region of interest, and the correlation between each pair was calculated. This is then plotted against the Euclidean distance between the regional centroids.

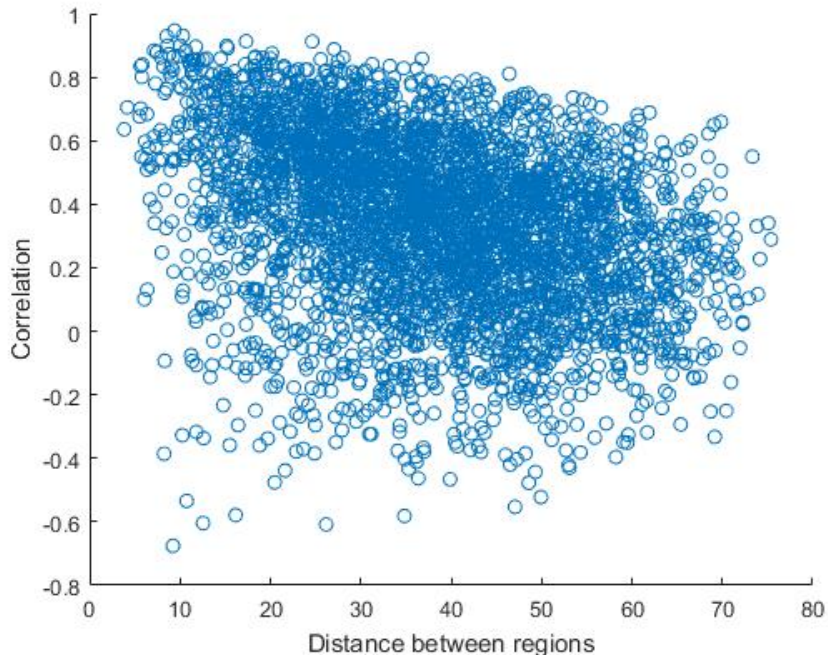


Figure 12: Correlation between ROI means.

Although there is quite a lot of noise in Figure 12 there is evidence of a trend that as distance increases, the correlation reduces to zero.

### 2.3.3 The Ljung-Box Test

In Section 2.3.1 many plots of the data were given. In particular an example of a voxel time series is given in Figure 3. From the figure it can be seen that there is some evidence of slightly different behaviour across the different stages of the experiment, however it is difficult to see if the signal shows any response to the activity or if the signal is just noise. One method to test if a time series signal is noise is the Ljung-Box Test [60]. The Ljung-Box Test statistic is given by

$$Q(K) = N(N + 2) \sum_{k=1}^K \frac{\hat{\rho}_k^2}{N - k} \quad (1)$$

where  $K$  is the number of lags to be tested,  $N$  is the number of data points and  $\rho_k$  is the autocorrelation function at lag  $k$ . The Ljung-Box statistic is

used to test the null hypothesis  $H_0 : \rho_1 = \rho_2 = \dots = \rho_h = 0$ , where the null distribution is  $\chi_K^2$ . That is, the Ljung-Box Test tests if the autocorrelations in the time series data are zero up to a lag of choice. If the null hypothesis is true this is evidence that the signal is white noise. The lag that is commonly used in the test is  $K = \ln(N)$  [71] which is the lag used here.

Here the Ljung-box test will be applied to each voxel time series, though note that the periods of activity and rest are not tested separately. Applying the Ljung-Box statistic to each voxel time series (with its mean removed) resulted in 124135 of the 228453 voxels being nothing but noise (55%) when testing at the 5% significance level, and the remaining 104318 failed the test suggesting some autocorrelations in the data. Figure 13 shows examples of a noisy signal and one for which the null hypothesis of the Ljung-Box test was rejected.

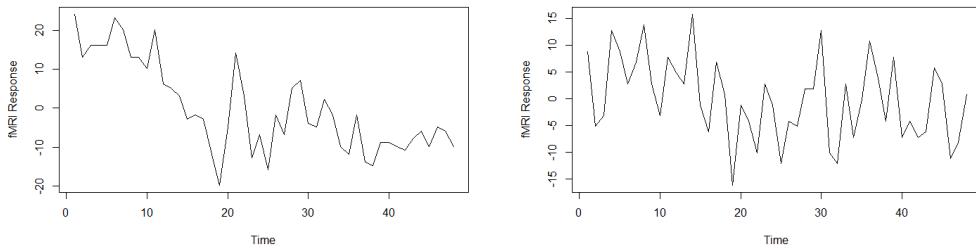


Figure 13: A voxel containing autocorrelations (left) and a voxel of noise (right) as determined by the Ljung-Box test.

As noted 104318 of 228453 voxels had their null hypothesis rejected, this is approximately 45% of all voxels. Using the results of the Ljung-Box tests, it would be interesting to investigate if the non-noisy signals were associated with particular regions of interest in the brain. Checking the total number of voxels in each of the 90 ROIs and calculating the number of which had the null hypothesis rejected means the proportion of non-noisy signals in each region of interest can be found. The proportions range from 22 – 73% across regions, meaning there are some which consist mostly of noise, and some which consist mostly of interesting signals. There are 14 regions identified from this analysis consisting mostly of noise, with less than 30% of the voxels in the regions being anything but noise. These regions are (8, 20, 33, 34, 35, 36, 38, 61, 65, 69, 71, 74, 78, 80). There are only 3 regions

of interest which contain a large number ( $> 70\%$ ) of non-noisy signals and these are (28, 41, 87). These results are based on proportions alone. Further investigation shows that, for example region number 41 has a high proportion but overall it is the smallest region of interest with only 155 voxels tested as being not just noise using the Ljung-Box test. Conversely, region number 8 has a low proportion of non-noisy voxels but is one of the largest regions of interest and contains 1441 non-noisy voxels according to the Ljung-Box test.

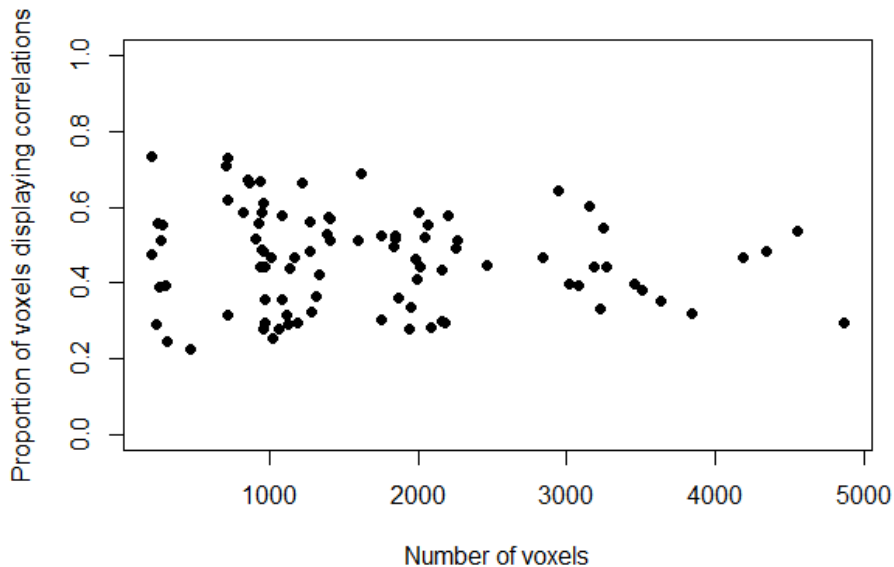


Figure 14: Number of voxels in a ROI against proportion of those which displayed correlations according to the Ljung-Box test.

The Ljung-Box test has revealed which of the voxels consist mostly of noise, and which do not. There are a large number of both noisy and non-noisy voxels. There are some regions which have a high proportion of noisy voxels, however when looking at total numbers, these are often the regions with the largest numbers of voxels in total, and they contain also some of the largest number of non-noisy voxels.



## 3 Modelling

In this chapter the aim is to describe and implement a spatio-temporal model for the brain data. Fitting a good model will help us further understand the data, allow us to test for voxel activation and use various methods to analyse the model residuals.

The structure of this chapter is as follows. Section 3.1 will begin by describing some previously proposed models, starting with an independent model and then ones which include temporal and spatial correlation between observations. The spatial model that we will initially aim to fit will be the one proposed by Castruccio et al. (2016) [31]. A number of problems were identified when fitting this model to the brain data, and so in subsection 3.3 these will be discussed and investigated. In Section 3.4 the proposed mean structure of the model will be investigated in more detail and a new mean structure will be discussed, and using an independent model, the two will be compared. Once the final mean structure has been decided, temporal and spatial models will be discussed, before moving on to application to the brain data.

### 3.1 Previously Proposed Models

There has been much analysis carried out on fMRI brain data. The simplest approach to the analysis is to fit a normal linear model to each voxel time series independently [78]. However it has been shown that if spatial correlation is ignored, the result is an increase in type one error [40] when testing for voxel activation. This means that there is an increase in detection of active voxels when the voxel is actually inactive.

Many early works on the analysis of fMRI data focused on models which allow for within voxel temporal correlation but either ignore the spatial correlation between voxels, or impose a simplistic approach due to computational issues which may not fully account for spatial dependence. For example, Worsley and Friston (1995) [76] discuss modeling smoothed fMRI data, Losascio et al. (1997) [61] included only within voxel temporal components and Bullmore et al. (2001) [28] made use of the wavelet transform of the data (which is known to whiten the data) in order to deal with temporal correlation.

Following this there have been many methods proposed to incorporate spatial dependence in fMRI modeling. For example, Worsley et al. (1992)

[75] indirectly modeled spatial correlation by assuming voxel activation test statistics were realisations from a random field, and then a threshold was used to detect active voxels. As noted in [31], this thresholding is dependent on the spatial smoothing which is assumed to be isotropic. A Bayesian approach to allow for spatial correlation has been developed by Bowman et al. in [24], [25] and [26]. The approach here is a hierarchical analysis which allowed for spatial dependence between voxels in a region. The drawbacks to this approach was that stationarity was assumed for voxels within a region (this has been shown to be not always true in [31]) and it is very computationally expensive, so much so that the analysis in these papers were limited to subsets of the data such as two-dimensional slices of the brain. Kang et al. (2012) [56] proposed a spatial model which captures both the local and regional spatial correlation by defining the model on the spectral domain, by modeling Fourier coefficients of the data. This approach was taken as 'Fourier coefficients are approximately uncorrelated under a temporal stationarity condition', thus increasing the potential for scalability. However, here activation was only assessed on the regional level and isotropy was assumed at the local voxel level. There are many more recent approaches to modeling spatial dependence, for example see [83], [81], [82], [84] and [54]. However all lack a method to adequately model the local non-stationarity at the voxel level.

Castruccio et al (2016) [31] proposed a method that deals with many of the drawbacks of the models described. This model allows for spatial dependence that is locally non-stationary, describes the connectivity between ROIs, allows for anisotropy and can be carried out on the full brain instead of subsets. Therefore the proposed model in [31] will be a starting point for our spatial model.

In this section, a basic independent model will be discussed, followed by allowing for within voxel temporal correlation, initially using the method proposed by Worsley et al. (2002) [77]. In this paper the proposed estimation method is to initially fit an independent model to the data, and use the residuals to estimate the autoregressive parameters before refitting the mean parameters using differenced data. More details can be found in Section 3.6. Finally a spatio-temporal model will be used following the proposed model by Castruccio et al. (2016) [31].

## 3.2 The Model Covariates

Following the work of [31], all models considered will initially involve three independent variables,  $x_1, x_2, x_3$ . The third covariate  $x_3$  allows for a time trend and is the vector  $x_3 = 1, 2, 3, \dots, 48$ . The remainder of this subsection will focus on the first two independent variables,  $x_1$  and  $x_2$ , which are generated using the hemodynamic response function (HRF) [9]. The HRF is a function which shows the typical changes at a voxel in terms of the increased blood flow (or delivery of oxygen) in response to a movement (see Figure 15). At  $t = 0$  a movement is carried out. It takes approximately 2 seconds for the increase in blood at a voxel to begin, and it plateaus at a maximum level at approximately 5-8 seconds after the movement begins. If the movement continues the HRF stays at this level. If not then around 8-11 seconds later the HRF will return to its normal level. Note that zero on the figure is the baseline amount of blood oxygen required at all times. An interesting feature of the HRF is the dip below zero that can be seen shortly after the activity has stopped in Figure 15. This is known as the post-stimulus undershoot. The interpretation of the undershoot is still under debate, but could be caused by a number of things. The leading explanations are delayed vascular compliance (a reduction in blood pressure due to the expansion of the vascular wall) and a sustained increase in oxygen post stimuli [73].

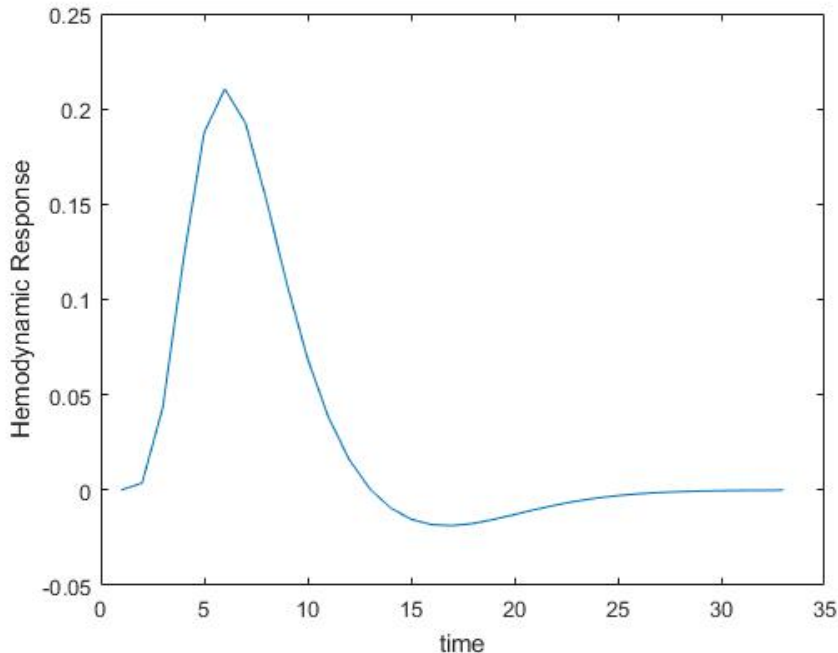


Figure 15: The HRF with temporal resolution of 1 second.

As the temporal resolution of the fMRI scans are every two seconds, the HRF that will be used has been scaled so every one unit of time represents two seconds. The way this information is used in the linear modelling is to construct covariates known as BOLD (blood oxygen level dependent) responses. In general fMRI modelling, a BOLD covariate is included for each different type of stimulus. For example in [78] a BOLD covariate was included for both 'hot' and 'warm' stimulus conditions. Following the work in [31], for the experiment this report is concerned with, a BOLD covariate is included for the 'activity' and 'rest' stimuli. Each BOLD covariate is constructed by convolving the HRF ( $hrf(t)$ ) with an indicator function  $I(t)$ , which gives the times in the experiment where the stimuli occurs. Therefore  $x_1(t)$  which is the BOLD covariate for the activity stimuli is constructed with

$$x_1(t) = hrf(t) \star I(t) \tag{2}$$

and can be seen in Figure 16.

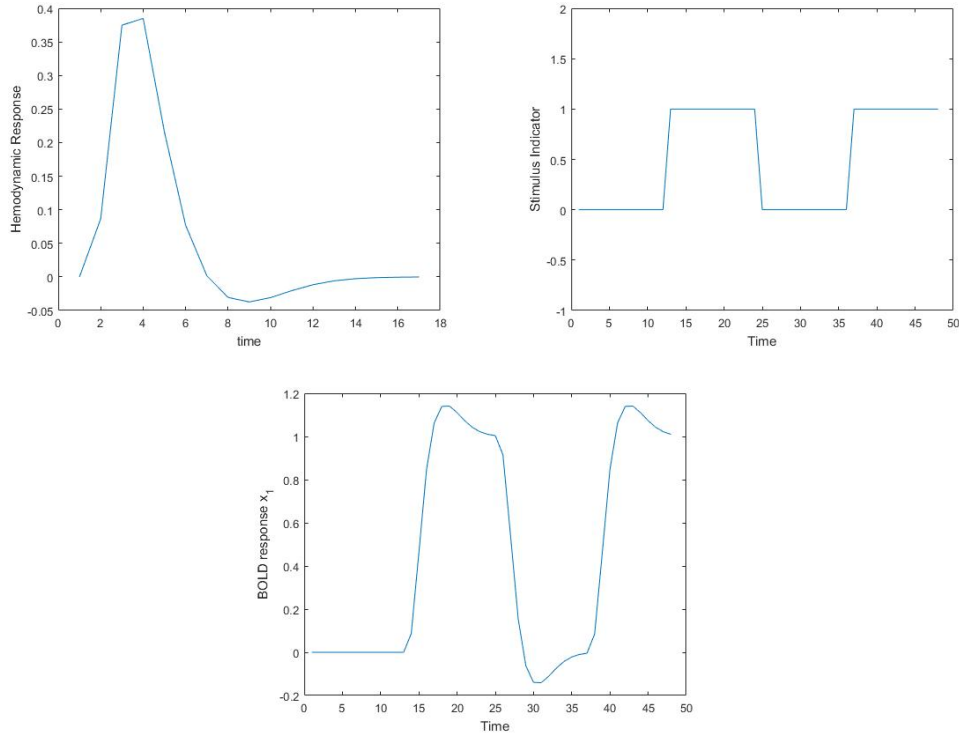


Figure 16: Top left shows the HRF with temporal resolution of two seconds. Top right shows the indicator function for the activity stimuli, and bottom centre is the BOLD response - the result of convolving the HRF with the indicator function.

The activity stimuli here is the hand grasping exercise which the participant is asked to do continually over the course each set of 12 scans. Therefore the effect of convolving the HRF with the indicator function is to maintain the maximum level of blood flow expected at a voxel whilst the movement occurs over the course of the 12 scans.

The second of the BOLD responses corresponding to the 'rest' activity as proposed by [31] is constructed by (3) where  $I_2(t) = 1 - I(t)$ .

$$x_2(t) = hrf(t) \star I_2(t) \quad (3)$$

To test for voxel activation we test if the estimated coefficients of the two BOLD covariates are statistically equal or different. As at any one time there is only one stimulus present (rest or active), if the estimated coefficients of

the BOLD covariates are statistically the same then this implies the voxel is inactive, ie a different weight need not be applied to account for a change in the signal when activity (or rest) occurs.

### 3.3 Experiment Design And Modeling Issues Identified

The experiment consisted of scanning the participant’s brain whilst they were at rest or performing a simple hand tapping exercise. There are a total of 48 scans. The experiment was designed such that the first 12 were whilst the participant was at rest, and then alternating between the stimulus task and rest conditions, each consisting of 12 scans. The result being a large set of time series of length 48. The first issue identified when modeling such data is that the short length of each time series resulted in poor estimates for certain model parameters, in particular the autoregressive parameters in the temporal model. As the time series length is short, the bias of the estimated covariance parameters obtained when using maximum likelihood was very noticeable. This can be remedied by using restricted maximum likelihood estimation, and will be further discussed in Section 3.6.4, where simulation studies will compare maximum likelihood and restricted likelihood estimation.

The second issue identified is concerning the model covariates proposed by [31] and described in Section 3.2. By fitting some models to simulated data, it was quickly identified that there was an issue with collinearity between the BOLD responses. The result of this was that parameter estimates had large standard deviations. Here this problem will be shown in more detail.

#### 3.3.1 Collinearity - Sequentially Adding Parameters

Let  $X$  be a design matrix consisting of the intercept covariate  $x_0$ , and the covariates  $x_1$ ,  $x_2$  and  $x_3$  as described in Section 3.2, and let  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  be the corresponding covariate parameters. In this section we will use simulations from a normal linear model fitted independently to each voxel to show the issues with collinearity.

Here multiple sets of simulated independent data will be used, starting with a mean consisting of a single parameter,  $\beta_0$ , and then adding covariates one at a time to investigate at what point the problems begin. Table 1 shows the results.

Mean Parameters	Estimates Mean	Estimates SD
$\beta_0$	500.06	4.35
$\beta_0, \beta_3$	(499.85, 2.01)	(8.85, 0.34)
$\beta_0, \beta_1, \beta_3$	(499.93, 2.25, 2.00)	(8.73, 9.27, 0.34)
$\beta_0, \beta_1, \beta_2, \beta_3$	(499.75, 2.08, 2.12, 2.01)	(20.17, 23.49, 21.84, 0.37)

Table 1: Independent model parameter estimates mean and standard deviation using various ROI numbers of parameters. In all cases results are from using 4338 voxels in ROI 7 with  $\beta_0 = 500$ ,  $\beta_1 = 2$ ,  $\beta_2 = 2$ ,  $\beta_3 = 2$  and  $\sigma = 30$ .

From Table 1 it can be seen that the standard deviation of the estimates of  $\beta_0$  when fitting alone is 4.35. Adding the covariate  $x_3$  which is the time trace, increases this to 8.85. Adding in the BOLD covariates one at a time first does not change the standard deviation much when one is added, but when a second is added it increases to 20.17. Through experimentation with the order of adding covariates, we found that adding  $x_1$  to the intercept alone increased the standard deviation of  $\beta_0$  to approximately eight like above, and it seems that only when both the BOLD responses are included that the standard deviation increases a lot to approximately 20. This leads us to believe that the design matrix may be ill conditioned.

### 3.3.2 Collinearity - The Reciprocal Condition Number

The condition number of a square matrix is defined as the product of the matrix norm with the norm of its inverse [8]. If the condition number is very large then the matrix is close to singular which can cause computation issues which can often lead to inaccurate results.

Consider the example of simple linear regression where we aim to solve

$$\hat{\beta} = AX^T\mathbf{Y}$$

where  $A = (X^T X)^{-1}$ . The condition number of the matrix  $A$  tells us how close to singular the matrix is, and so how stable the solution of this equation is under small perturbations of the elements of  $A$ . The condition number is high if small perturbations of  $A$  result in very different estimates for  $\hat{\beta}$ . If the condition number is small then the solution is stable. For matrices with a high condition number, this can lead to unreliable estimates, and high variances of those estimates. Conversely, the reciprocal condition number is very close to zero for ill-conditioned matrices. Table 2 shows the reciprocal condition

number of the different versions of the matrix  $(X^T X)^{-1}$  as parameters are added to the model.

Parameters	Mean	SD	RCN
$\beta_0$	500.06	4.35	1
$\beta_0, \beta_3$	(499.85, 2.01)	(8.85, 0.34)	$3 \times 10^{-4}$
$\beta_0, \beta_1, \beta_3$	(499.93, 2.25, 2.00)	(8.73, 9.27, 0.34)	$2 \times 10^{-4}$
$\beta_0, \beta_1, \beta_2, \beta_3$	(499.75, 2.08, 2.12, 2.01)	(20.17, 23.49, 21.84, 0.37)	$2 \times 10^{-5}$

Table 2: Independent model parameter estimates mean and standard deviation using various numbers of parameters and reciprocal condition number (RCN) of  $(X^T X)^{-1}$ . In all cases results are from using 4338 voxels in ROI 7 with  $\beta_0 = 500$ ,  $\beta_1 = 2$ ,  $\beta_2 = 2$ ,  $\beta_3 = 2$  and  $\sigma = 30$ .

From Table 2 it can be seen that with a single parameter, the RCN is one, that is it is very well conditioned and the solution is stable. Adding a second parameter (the time trend) makes the RCN drop to almost zero showing an unstable solution. Adding the first of the BOLD covariates does not change the RCN much, hence the standard deviations stay roughly the same, but adding the second BOLD responses reduces the RCN by another order of magnitude which will explain the large increase in the variance here. Therefore this shows that as more parameters are added,  $(X^T X)^{-1}$  is becoming more and more unstable, which explains the increase in the variance of the parameter estimates. The reason for this may be in part explained by Figure 17 which shows that the two BOLD responses are highly correlated. In fact the correlation is  $-0.9162$  which may be causing collinearity problems. However, note that simply adding a time trend caused a massive decrease in RCN, so this is also an issue. One way to address this large decrease when the time trend is added, is to centre the time trend covariate. The result when added to the intercept term is a decrease in RCN from 1 to 0.005 which is some improvement.



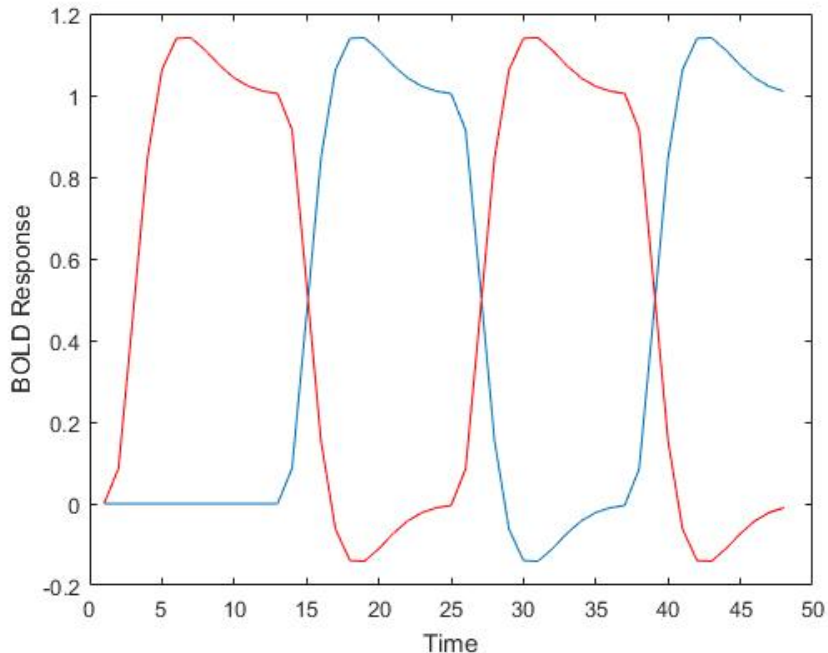


Figure 17: The BOLD covariate during activity ( $x_1$  - blue) and rest ( $x_2$  - red). The activity stimuli is the hand grasping exercise which the participant is asked to do continually over the course each set of 12 scans. The effect of convolving the HRF with the indicator function is to maintain the maximum level of blood flow expected at a voxel whilst the movement occurs over the course of the 12 scans.

### 3.4 The Mean Structure

As discussed in Section 3.2, the model proposed by Castruccio et al. (2016) [31] consists of a mean with two BOLD covariates, as well as a term for time trend. For the most simple case where the errors of the model are assumed independent, the model equation is given by

$$y_i(t) = \beta_{0,i} + \beta_{1,i}x_1(t) + \beta_{2,i}x_2(t) + \beta_{3,i}x_3(t) + \epsilon_i(t) \quad (4)$$

where  $y_i(t)$  is the fMRI response at voxel  $i$  for  $t = 1, \dots, 48$ , the covariates are as described in Section 3.2 and  $\epsilon_i \in N(0, \sigma_i^2)$ . Note that there is no  $i$  subscript on the covariates as they are the same at all voxels, ie we assume

the HRF is not dependent on location, however we do allow for a different parameter estimate at all voxels.

It has been seen in Section 3.3 that the two BOLD responses are very highly correlated, which may then cause collinearity problems. Castruccio et al. (2016) included the two BOLD responses to account for the two stimulus conditions; activity and rest. In general fMRI modeling, a BOLD covariate is included for each different type of stimulus. This method is described in [77], and an example of application was given in [78] where a BOLD covariate was included for both 'hot' and 'warm' stimulus conditions of an experiment. This method can be seen in many other works, for example in [74] or [34] to name just a few. In all these papers note that the rest phases of an experiment is not modelled with a BOLD response.

In this section, an alternative mean structure will be investigated. Firstly, we will remove the BOLD covariate corresponding to periods of rest. This will remove the collinearity issues. Then we will investigate if anything can be gained by allowing for the possibility that the BOLD response may differ for each of the activity parts of the experiment. That is, an extra covariate will be included to allow for this.

### 3.4.1 Different Covariates For Each Period Of Exercise

In this section, we will propose a new independent model with

$$y_i(t) = \beta_{0,i} + \beta_{1,i}x_1(t) + \beta_{3,i}x_3(t) + \epsilon_i(t) \quad (5)$$

where  $y_i(t)$  is the fMRI response at voxel  $i$  for  $t = 1, \dots, 48$ , with the covariates as described in Section 3.2 and  $\epsilon_i \in N(0, \sigma_i^2)$ . Note that here we have dropped the BOLD response for periods of rest, and have kept the BOLD response  $x_1(t)$  for periods of activity (see Figure 17). Note that this covariate assumes that the response during the two periods of activity is the same. Therefore a different model that we can consider is one where the BOLD response is allowed to differ for the two periods of activity.

Let us define  $x_1^*(t) = hrf(t) \star I_1^*(t)$  and  $x_2^*(t) = hrf(t) \star I_2^*(t)$  where  $I_1^*(t)$  and  $I_2^*(t)$  are new indicator functions which describe the first and second periods of activity respectively. Then an alternative model to consider is

$$y_i(t) = \beta_{0,i} + \beta_{1,i}x_1^*(t) + \beta_{2,i}x_2^*(t) + \beta_{3,i}x_3(t) + \epsilon_i(t). \quad (6)$$

To compare the models, initially model (6) will be fit to some brain data and

we will look at the estimates of  $\beta_1$  and  $\beta_2$  to see how similar or different they look. Table 3 shows the estimates of the BOLD parameters calculated for all voxels in the brain using the two independent models (one with a single BOLD response and one with a BOLD response for each session of activity).

Model Parameters	Mean( $\beta_1$ )	SD( $\beta_1$ )	Mean( $\beta_2$ )	SD( $\beta_2$ )
$\beta_0, \beta_1, \beta_3$	1.60	12.82	-	-
$\beta_0, \beta_1, \beta_2, \beta_3$	1.48	15.73	1.94	10.47

Table 3: The mean and standard deviation of the estimates of  $\beta_1$  and  $\beta_2$  for the full brain and models which include a single BOLD covariate or two BOLD covariates (one for each activity phase).

Figure 18 shows the results from applying model (6) to all voxels in each ROI.

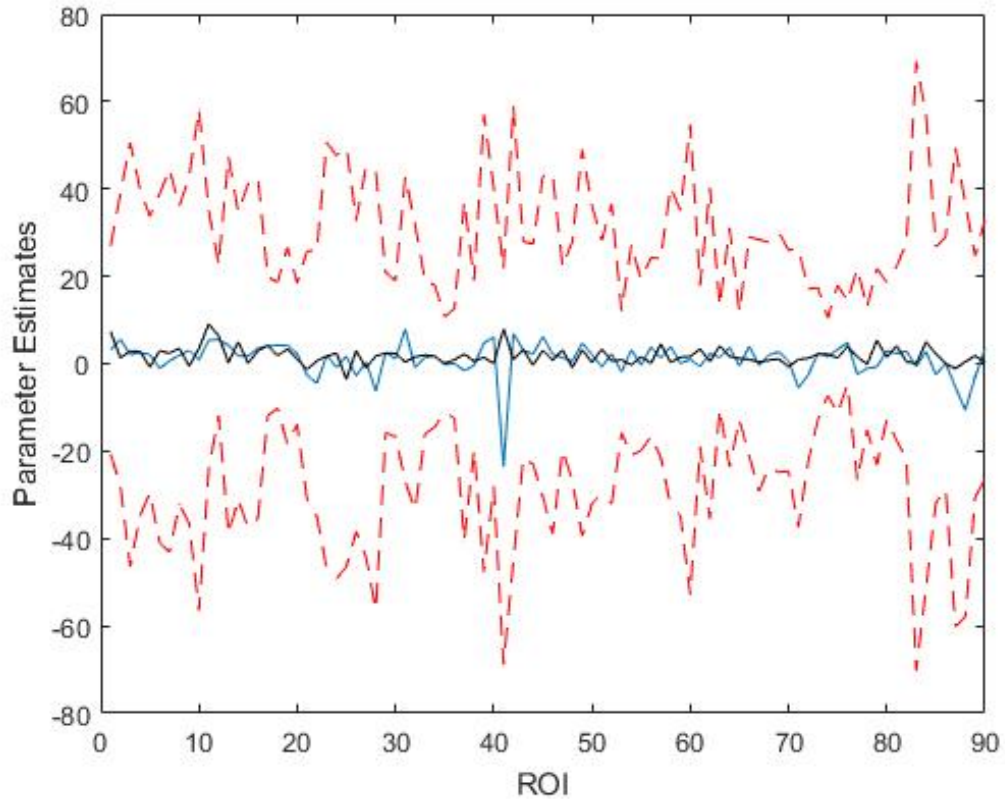


Figure 18: The mean of the estimates of  $\beta_1$  (blue) and  $\beta_2$  (black) for all ROIs, with the mean plus/minus one standard deviation (red dashed) of  $\hat{\beta}_1$

From Figure 18 it can be seen that there is a lot of variability about the estimates of  $\beta_1$  and  $\beta_2$ , and on first impression it looks like it may be unnecessary to include a separate BOLD response for each period of activity. Note that for ROI 41 there appears to be a noticeable difference between the estimates of the two parameters. This region has the fewest number of voxels of all regions (211 voxels), and we will see later that some unusual results can be found for this region as a result of this.

Next we will formally test which model is most suitable. One method is to use AIC (Akaike's Information Criterion) [13]. The AIC of a model fit is given by

$$AIC = 2k - 2\ln(\hat{L}) \quad (7)$$

where  $k$  is the number of parameters estimated and  $\ln\hat{L}$  is the value of the log-likelihood at the maximum. The idea here is to calculate the AIC for competing models, and the one with the lowest value is the preferred model. Intuitively the criterion chooses the model which is most likely subject to a penalty term for the number of model parameters.

Fitting the two competing models ((5) and (6)) to each voxel in the brain independently and calculating the AIC for each model, it was found that 89989 of the 228453 voxels had a AIC score which was lower when using the model with a BOLD response for each of the periods of activity, (6). This is approximately 40% of the voxels. Therefore although Figure 18 suggests one BOLD covariate is enough, there is an argument here to use two BOLD covariates as it will benefit the modeling of a substantial number of voxels.

From this point forward we will only consider the model (6), and will drop the star notation from the covariates  $x_1^*(t)$  and  $x_2^*(t)$ . Hence from this point forward  $x_1(t)$  and  $x_2(t)$  are to be the BOLD covariates representing the first and second period of exercise respectively. The covariates are the same at all voxels, ie we assume the HRF is not dependent on location, however we do allow for a different parameter estimate at all voxels. Activation of a voxel in response to exercise will then occur if the estimates of the model parameters  $\beta_1$  or  $\beta_2$  are statistically non-zero. We will also redefine  $x_3(t)$  so it is the centered time trend to address the issue of a large drop in RCN when the time trend is added (see Section 3.3).

### 3.5 The Independent Model

The first model to be considered in the independent model, which assumes complete independence across space and time between voxels. As noted earlier, this is not a realistic model. The inclusion here is intended to show a basic working model where data can be simulated and parameters estimated efficiently. Also later this model will be used to show the problem with the independent model, in that when applied to correlated data, type one error is inflated. The model equation is given by

$$y_i(t) = \beta_{0,i} + \beta_{1,i}x_1(t) + \beta_{2,i}x_2(t) + \beta_{3,i}x_3(t) + \epsilon_i(t) \quad (8)$$

where  $y_i(t)$  is the fMRI response at voxel  $i$  for  $t = 1, \dots, 48$ ,  $\epsilon_i \in N(0, \sigma_i^2)$  and the covariates are as described in Section 3.4, that is  $x_1(t)$  and  $x_2(t)$  are BOLD responses corresponding to the different periods of exercise, and  $x_3(t)$  is the centered time trend.

### 3.5.1 Simulating Independent Data

To ensure there are no problems with the model fitting to the brain data, here and in subsequent sections data will be simulated and parameters estimated using maximum likelihood. For the independent model, data is generated to mimic two different regions of interest which contain the largest and smallest numbers of voxels. To generate the data, the parameters are set to be  $\beta_0 = 500$ ,  $\beta_1 = \beta_2 = \beta_3 = 2$  and  $\sigma = 30$  at all voxels. These values were chosen from results of fitting the model on some of the brain data. Table 4 shows some results. In the table, number of voxels refers to the number of simulated time series, so in the first case a single set of 4338 independent time series were generated using the model described in (8). The mean and standard deviation of the set of parameter estimates is given, as well as the standard error of the mean (SEM).

Number of Voxels	Parameter	True Value	Mean	SD	SEM
4338	$\beta_0$	500	499.88	6.20	0.09
	$\beta_1$	2	1.98	10.30	0.16
	$\beta_2$	2	1.67	15.05	0.23
	$\beta_3$	2	2.00	0.43	0.01
	$\sigma$	30	29.86	3.17	0.05
	211	$\beta_0$	500	500.18	6.41
$\beta_1$		2	1.70	9.73	0.67
$\beta_2$		2	1.19	14.76	1.02
$\beta_3$		2	2.05	0.42	0.03
$\sigma$		30	29.78	3.01	0.21

Table 4: Independent model parameter estimates mean, standard deviation and standard error of the mean, from simulation of different numbers of voxels.

From Table 4 it can be seen that on average the parameter estimates are quite close to the actual values used to generate the data. However the

standard deviation of these estimates is quite large, especially for the mean parameters. However from the SEM we can see that all parameters are within two standard errors of the true value.

### 3.6 The Temporal Model

The next model to be considered is to extend (8) to allow for temporal correlation at each voxel by adding autoregressive errors. Let  $\rho_{ik}$  be the  $k$ -th autoregressive parameter at voxel  $i$ , then the model is given by

$$y_i(t) = \beta_{0,i} + \beta_{1,i}x_1(t) + \beta_{2,i}x_2(t) + \beta_{3,i}x_3(t) + \epsilon_i(t) \quad (9)$$

where

$$\epsilon_i(t) = \sum_{k=1}^K \rho_{ik}\epsilon_i(t-k) + \nu_i \quad (10)$$

and  $\nu_i \in N(0, \sigma_i^2)$ .

Before any modelling can begin, the order of the autoregressive process needs to be determined.

#### 3.6.1 The Autoregressive Parameters

The order of a AR process can be determined using the partial autocorrelation function (PACF). Typically for an AR( $p$ ) process the PACF will display significant partial correlations for the first  $p$  lags. Significant correlations are usually determined as ones which are larger in magnitude than some confidence interval. The 95% confidence interval for partial correlations is  $\pm \frac{2}{\sqrt{N}}$  where  $N$  is the length of the time series. In this section we will investigate the order of the AR process for each of the voxels in some of the regions of interest. The analysis is carried out on the residuals of the linear model which initially assumes independence. Figure 19 shows an example of a PACF for a voxel time series and Table 5 summarises the results of the PACF to all voxels in some regions of interest. These regions were chosen as they later will be shown to display some interesting features when clustering or performing topological data analysis on the data.

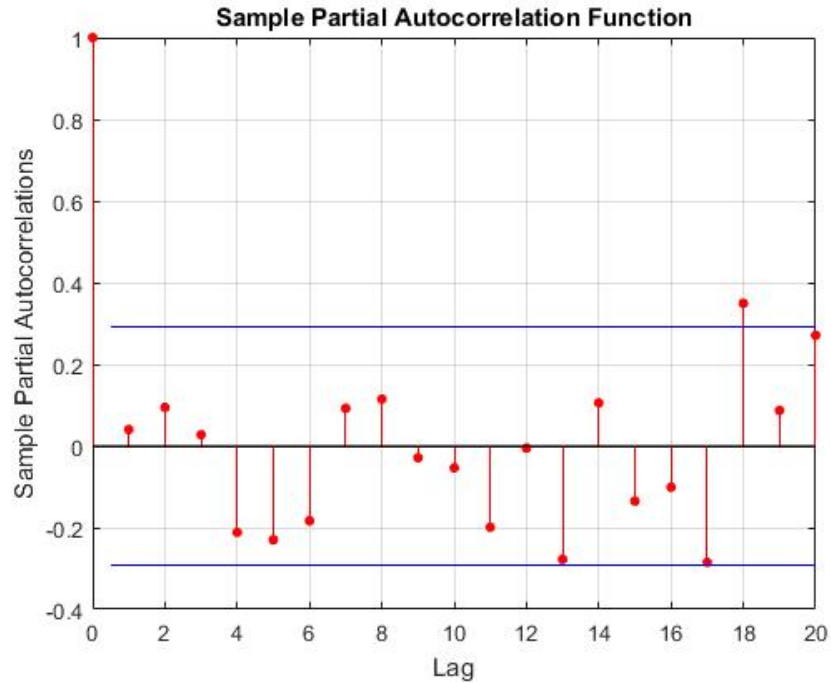


Figure 19: An example of the PACF for a voxel time series

Region	Total Number Voxels	$p = 0$	$p = 1$	$p = 2$	$p > 2$
3	3262	2961	218	69	14
7	4338	3824	397	100	17
8	4865	4540	258	62	5
21	278	240	23	15	0
41	211	118	81	12	0

Table 5: Number of voxels displaying significant partial correlations up to and including lag  $p$ .

From Table 5 it can be seen that for the majority of voxels, there is no significant partial autocorrelation at any lags. There are some which justify an AR(1) process and very few beyond that which are higher order. Fitting models using AR(1) errors and AR(2) errors and comparing using AIC, it was found that 37% of all voxels would benefit from a model with AR(2) errors. This is a significant number of voxels, and therefore from this point



on we will only consider models where the temporal correlations are described using a AR(2) process.

Figures 20 and 21 show some three dimensional plots of the values of the autoregressive parameters on the brain. It is difficult to get much information from these plots, especially those regions with large numbers of voxels. However there is some evidence of spatial correlation, ie the lighter colours are near each other and do not appear to be randomly scattered. The model is still spatially independent, so the voxels with higher values of  $\hat{\rho}_1$  correspond to signals with higher lag one correlations, and those which are dark blue are less correlated, ie more noisy.

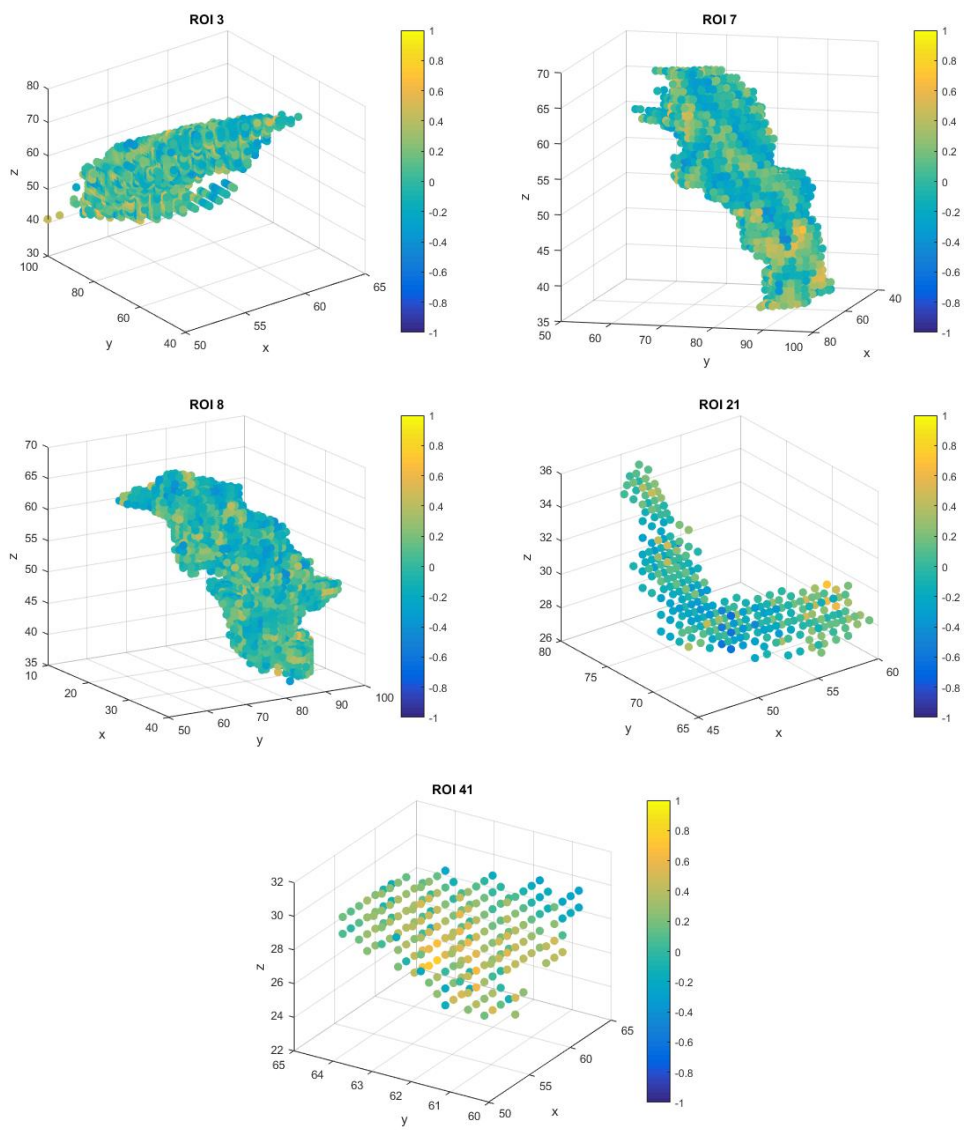


Figure 20: 3D plots of  $\hat{\rho}_1$  for different regions of interest.

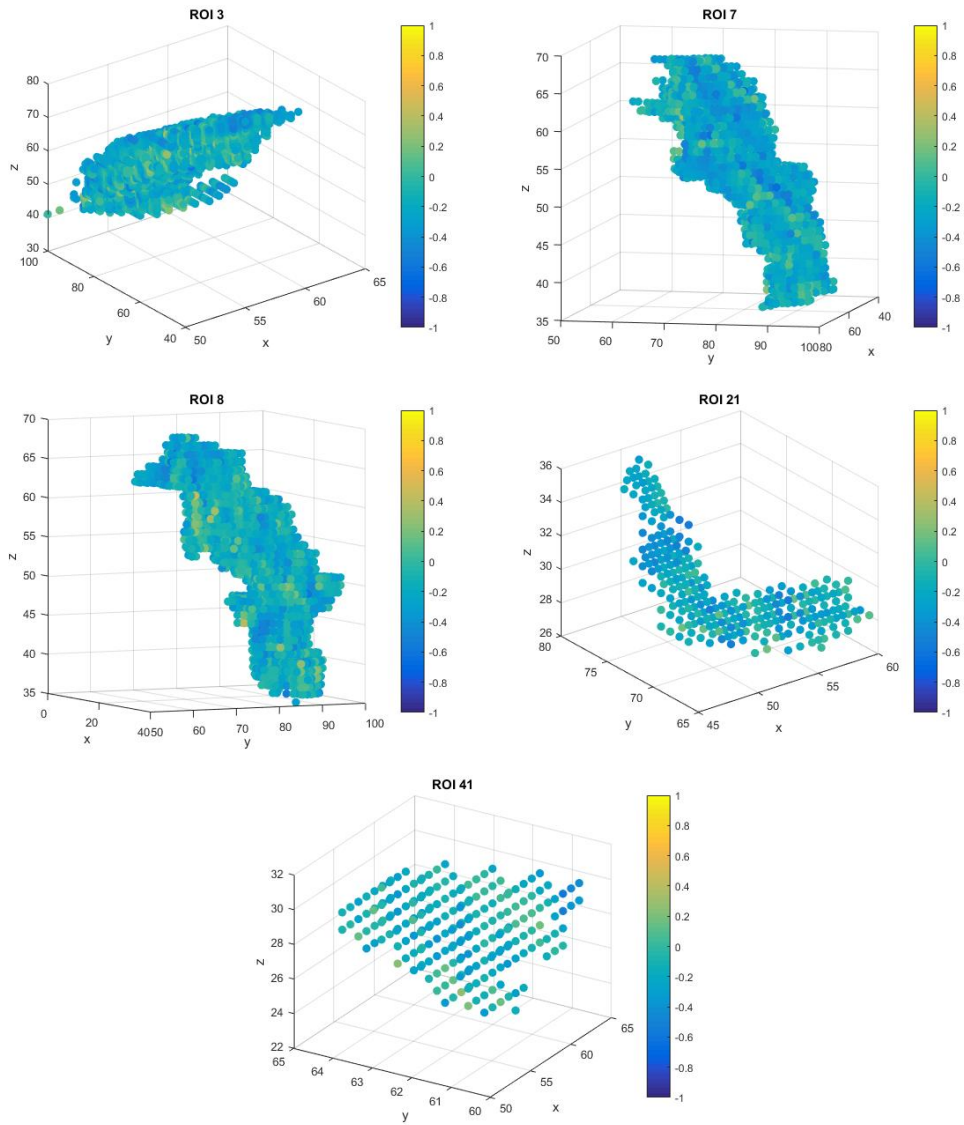


Figure 21: 3D plots of  $\hat{\rho}_2$  for different regions of interest.

### 3.6.2 Procedure To Fit The Model

Given we have seen that an AR(2) process is likely to be suitable for this data, the temporal model that we will be considering is of the form

$$y_i(t) = \beta_{0,i} + \beta_{1,i}x_1(t) + \beta_{2,i}x_2(t) + \beta_{3,i}x_3(t) + \epsilon_i(t) \quad (11)$$

where

$$\epsilon_i(t) = \rho_{i1}\epsilon_i(t-1) + \rho_{i2}\epsilon_i(t-2) + \nu_i(t) \quad (12)$$

and  $\nu_i \in N(0, \sigma_i^2)$ . Also note that we require

$$\begin{aligned} \rho_{i1} + \rho_{i2} &< 1 \\ \rho_{i2} - \rho_{i1} &< 1 \\ |\rho_{i2}| &< 1 \end{aligned} \quad (13)$$

so that the process is stationary [11].

The fitting of such a model can be achieved in a number of ways. Firstly, standard maximum likelihood estimation can be used to estimate all parameters simultaneously, however this may have limitations when scaling up if we wish to model the full brain (see 3.6.3). To overcome this Worsley et al. (2002) [77] proposed an alternative method which is computationally much more efficient. The method proposed by Worsley et al. (2002) [77] will be explored in this section, and results will be compared with those obtained using standard maximum likelihood methods. The results obtained will show that both maximum likelihood and the method proposed by Worsley et al. (2002) [77] perform poorly on shorter time series, and thus this section will conclude with modeling data using restricted maximum likelihood estimation (REML).

For a given time series, Worsley et al. (2002) suggest that the first step of fitting a temporal model to the data, is to first estimate the autoregressive parameters,  $\rho_1$  and  $\rho_2$ . It is suggested by [77] that these estimates can be obtained by the residuals of ordinary least squares regression on the original data. Let  $r_t$  denote these residuals at time  $t$ , then the estimate for  $\rho_1$  is given by

$$\hat{\rho}_1 = \frac{\sum_{t=2}^n r_t r_{t-1}}{\sum_{t=1}^n r_t^2} \quad (14)$$

with a similar expression for  $\hat{\rho}_2$ . Note that if these estimates of the AR(2) parameters do not meet the stationarity conditions above, then before the modelling one could take steps to make the data stationary, for example by

differencing [53]. Once these parameters are estimated, the next step is to take differences of the response to recover independent errors, ie

$$y_i(t) - \hat{\rho}_1 y_i(t-1) - \hat{\rho}_2 y_i(t-2) = \beta_0(1 - \hat{\rho}_1 - \hat{\rho}_2) + \sum_{j=1}^3 \beta_j(x_j(t) - \hat{\rho}_1 x_j(t-1) - \hat{\rho}_2 x_j(t-2)) + \nu_i(t). \quad (15)$$

The result of this differencing means we have a transformed equation of the form

$$\tilde{y}_i(t) = \tilde{\beta}_0 + \beta_1 \tilde{x}_1(t) + \beta_2 \tilde{x}_2(t) + \beta_3 \tilde{x}_3(t) + \nu_i(t) \quad (16)$$

where  $\tilde{y}_i(t) = y_i(t) - \hat{\rho}_1 y_i(t-1) - \hat{\rho}_2 y_i(t-2)$ ,  $\tilde{\beta}_0 = \beta_0(1 - \hat{\rho}_1 - \hat{\rho}_2)$ ,  $\tilde{x}_j(t) = x_j(t) - \hat{\rho}_1 x_j(t-1) - \hat{\rho}_2 x_j(t-2)$  (for  $j = 1 - 3$ ) and  $\nu_i \in N(0, \sigma_i^2)$ . Note that in the transformation we let  $\tilde{y}_i(1) = y_i(1)$ ,  $\tilde{x}_1(1) = x_1(1)$  and  $\tilde{x}_2(1) = x_2(1)$ .

In [77] it is noted that this procedure could be iterated, ie use  $\hat{\rho}$  to find  $\hat{\beta}$  and then use the residual to gain a new estimate for  $\rho$ . However this does not give much improvement to the fit.

In summary, the procedure for fitting a model with correlated autoregressive errors is to first fit an ordinary least squares model to the original data, estimate  $\rho$  using the residuals, then transform the response and design matrix then refit.

### 3.6.3 Simulating Temporal Data

In this section temporal data will be simulated and parameters estimated using the proposed method by Worsley et al. (2002). For this model, data is generated to mimic two different regions of interest which contain the largest and smallest numbers of voxels in order to determine how the model performs on large and small regions of interest.. To generate the data, the parameters are set to be  $\beta_0 = 500$ ,  $\beta_1 = \beta_2 = \beta_3 = 2$ ,  $\sigma = 30$ ,  $\rho_1 = 0.15$  and  $\rho_2 = -0.30$  at all voxels. These values were chosen from results of fitting the model on some of the brain data. Table 6 shows some results. In the table, number of voxels refers to the number of simulated time series, so in the first case a single set of 4338 time series with AR(2) errors were generated using the model described in (17) with parameters chosen as above. The mean and standard deviation of the set of parameter estimates is given, as well as the standard error of the mean (SEM).

Number of Voxels	Parameter	True Value	Mean	SD	SEM
4338	$\beta_0$	500	500.00	5.44	0.08
	$\beta_1$	2	1.95	9.15	0.14
	$\beta_2$	2	2.27	13.71	0.21
	$\beta_3$	2	2.00	0.38	0.01
	$\sigma$	30	29.42	3.28	0.05
	$\rho_1$	0.15	0.04	0.11	0.002
	$\rho_2$	-0.30	-0.31	0.12	0.002
211	$\beta_0$	500	499.58	5.24	0.36
	$\beta_1$	2	2.75	9.27	0.64
	$\beta_2$	2	2.25	13.31	0.92
	$\beta_3$	2	1.97	0.38	0.03
	$\sigma$	30	29.17	3.32	0.23
	$\rho_1$	0.15	0.03	0.11	0.01
	$\rho_2$	-0.30	-0.33	0.12	0.01

Table 6: Temporal model parameter estimates mean, standard deviation and SEM from simulation of different numbers of voxels and estimation using the Worsley method.

From Table 6 it can be seen that the mean parameters have large standard deviations as seen before with the independent model. However what we can also see here is that the first order autoregressive parameter is poorly estimated, and has a large standard deviation with respect to the mean.

The results in Table 6 show simulation results using the model and covariates that will be used on the brain data. The poor estimates for the autoregressive parameter may be due to the time series at each voxel being only of length 48, or perhaps the number of voxels is not large enough. To investigate the impact on the first order autoregressive parameter, Table 7 shows some results from simulating zero mean temporal data, where the voxels have time series of differing lengths, and different numbers of voxels are used. In all cases the true value of the first order autoregressive parameter is  $\rho_1 = 0.25$ .

		Number of voxels			
		100	500	1000	5000
	48	0.20 (0.15)	0.21 (0.14)	0.21 (0.14)	0.21 (0.14)
Length	144	0.23 (0.09)	0.24 (0.08)	0.24 (0.08)	0.24 (0.08)
	500	0.24 (0.04)	0.25 (0.04)	0.25 (0.05)	0.25 (0.04)
	1000	0.25 (0.02)	0.25 (0.03)	0.25 (0.03)	0.25 (0.03)

Table 7: Mean and standard deviation of estimates of the autoregressive parameter, using different numbers of voxels with corresponding time series of different lengths. In all cases the true values of the autoregressive parameter is 0.25.

Table 7 shows that the voxel time series length is what is important to get accurate estimates of  $\rho$ . Increasing the number of voxels has very little impact on the estimates. Hence this may provide some explanation why  $\rho_1$  was poorly estimated in Table 6.

### 3.6.4 Restricted Maximum Likelihood Estimation

From Table 6 it can be seen that even for large numbers of voxels that the estimates of the first autoregressive parameter is very poor. Table 7 shows that this is likely a problem due to the length of each voxel time series. Recall the temporal model

$$y_i(t) = \beta_{0,i} + \beta_{1,i}x_1(t) + \beta_{2,i}x_2(t) + \beta_{3,i}x_3(t) + \epsilon_i(t) \quad (17)$$

where

$$\epsilon_i(t) = \rho_{i1}\epsilon_i(t-1) + \rho_{i2}\epsilon_i(t-2) + \nu_i(t) \quad (18)$$

and  $\nu_i \in N(0, \sigma_i^2)$ . In vector form the model can be written as

$$\mathbf{Y}_i \in N(X\boldsymbol{\beta}_i, \Sigma_i)$$

where  $\Sigma_i = \sigma_i^2 V$ , and  $V$  is the correlation matrix. For a general linear model such as this one, it can be shown that the standard maximum likelihood estimator for the variance component of the model is biased. If  $\mathbf{Y} \in N(X\boldsymbol{\beta}, \sigma^2 V)$  for some correlation matrix  $V$  then it can be shown that [37]

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - X\boldsymbol{\beta})V^{-1}(\mathbf{Y} - X\boldsymbol{\beta})}{n} \quad (19)$$

which is a biased estimator. The unbiased result is

$$\hat{\sigma}_{unbiased}^2 = \frac{(\mathbf{Y} - X\boldsymbol{\beta})V^{-1}(\mathbf{Y} - X\boldsymbol{\beta})}{n - p}. \quad (20)$$

If the length  $n$ , of the time series is large with respect to the number of parameters  $p$ , then the biased version is a good estimate for  $\sigma^2$ . However if this is not the case, then the variance parameters of the model could be significantly biased. A solution to correct for this bias when using small  $n$  is to use restricted maximum likelihood estimation (REML). The reason that the estimator for  $\sigma^2$  is biased when using standard maximum likelihood estimation is that the fixed effects are ignored, and the degrees of freedom used in deriving the estimators do not adjust for this [41]. The idea behind REML is to perform maximum likelihood estimation on a set of linearly transformed data  $\mathbf{Y}^* = H\mathbf{Y}$  where  $H$  is any matrix such that  $E[H\mathbf{Y}] = 0$ . The result is estimation based on data which is independent of the mean parameters  $\boldsymbol{\beta}$  and only contains information about the variance components. The REML log-likelihood is given by

$$l^R = -\frac{1}{2}\log|\Sigma| - \frac{n-p}{2}\log((\mathbf{Y} - X\boldsymbol{\beta})\Sigma^{-1}(\mathbf{Y} - X\boldsymbol{\beta})) - \frac{1}{2}\log|X^T\Sigma^{-1}X| \quad (21)$$

which, when maximised results in an unbiased estimator for  $\sigma^2$ . Note that (21) does not depend on  $H$ , and this result can be derived using any choice of  $H$  such that  $E[H\mathbf{Y}] = 0$ . Full algebraic details can be found in [37], where it is shown that the choice of  $H$  is independent of the resulting log-likelihood.

The following table shows the estimates for the mean and variance parameters of the temporal model using three methods. The first will be the method described in Section 3.6 based on the work of Worsley [77]. The second method is standard maximum likelihood estimation, and finally REML is used. In all cases data has been simulated using all model covariates (including two BOLD responses and a time trend) with parameters set as described in Table 8.



Estimation Method	Parameter	True Value	Mean	SD	SEM
Worsley	$\beta_0$	500	499.95	5.53	0.09
	$\beta_1$	2	1.97	9.13	0.14
	$\beta_2$	2	2.03	13.79	0.22
	$\beta_3$	2	1.99	0.39	0.01
	$\sigma$	30	29.38	3.30	0.05
	$\rho_1$	0.15	0.03	0.11	0.002
	$\rho_2$	-0.30	-0.32	0.12	0.002
Maximum Likelihood	$\beta_0$	500	499.96	5.58	0.09
	$\beta_1$	2	2.01	9.06	0.14
	$\beta_2$	2	1.90	13.88	0.22
	$\beta_3$	2	2.01	0.39	0.01
	$\sigma$	30	30.49	3.51	0.06
	$\rho_1$	0.15	0.04	0.16	0.002
	$\rho_2$	-0.30	-0.36	0.12	0.002
REML	$\beta_0$	500	500.05	5.42	0.09
	$\beta_1$	2	2.03	9.17	0.14
	$\beta_2$	2	1.76	13.88	0.22
	$\beta_3$	2	2.01	0.39	0.01
	$\sigma$	30	30.08	3.35	0.05
	$\rho_1$	0.15	0.12	0.16	0.002
	$\rho_2$	-0.30	-0.29	0.14	0.002

Table 8: Temporal model parameter estimates mean, standard deviation (SD) and standard error of the mean (SEM), using 4000 simulated voxels of length 48 and three different estimation methods.

From Table 8 we can see that REML has given a much better estimate of the autoregressive parameters than the other methods. We can also see that maximum likelihood estimation gives very similar results to those found using Worsley’s method. One reason to use the Worsley et al. (2002) method is that it is computationally much quicker than using maximum likelihood. Table 9 compares the time taken to perform the simulations described using the different methods.

Time Series Length	Worsley	Maximum Likelihood	REML
48	2	25	29
96	4	70	75
144	8	142	150

Table 9: Time (seconds) taken to fit the temporal model to 4000 time series of different lengths using three different estimation methods.

From Table 9 it can be seen that increasing each voxel length results in a much longer computation time for maximum likelihood and REML than Worsley’s method.

In summary, Worsley’s method may be preferable for situations where the time series is very large, or if there are a large number of time series to model, due to the significantly quicker estimation method. However REML gives better estimates of the autoregressive parameter, especially if the time series are of shorter length, though the computation time is higher.

### 3.7 The Spatio-Temporal Model

The final model to be considered is to extend the temporal model and include spatial correlation between voxels at different sites. Due to the size of the data, the approach of constructing a covariance matrix which can model the spatial dependence between observations and estimating parameters using likelihood methods is extremely computationally intensive. The main issues are the very large number of parameters that would need to be estimated jointly (four mean parameters for each voxel plus temporal parameters), and also the size of the covariance matrix that would be required. For example, modelling 4000 time series, each of length 48, would result in a covariance matrix consisting of approximately  $200,000^2$  elements, which cannot be stored in most computers. Therefore following the work of [31], a sequential approach to spatial modeling will be carried out here. From Section 3.6, the error term of the temporal model for voxel  $i$  is now expressed by

$$\epsilon_i(t) = \rho_{i1}\epsilon_i(t - 1) + \rho_{i2}\epsilon_i(t - 2) + \nu_i(t) \quad (22)$$

where  $\epsilon_i(t)$  is the time series of errors for voxel  $i$ ,  $\rho_{ik}$  is the  $k$ -th autoregressive parameter for voxel  $i$ , and  $\nu_i(t) \in N(0, \sigma_i^2)$ . Now suppose we wish to model a set of voxels  $i = 1, \dots, V$  then we have

$$\begin{aligned}
\epsilon_1(t) &= \rho_{11}\epsilon_1(t-1) + \rho_{12}\epsilon_1(t-2) + \nu_1(t) \\
\epsilon_2(t) &= \rho_{21}\epsilon_2(t-1) + \rho_{22}\epsilon_2(t-2) + \nu_2(t) \\
&\vdots \\
\epsilon_V(t) &= \rho_{V1}\epsilon_V(t-1) + \rho_{V2}\epsilon_V(t-2) + \nu_V(t).
\end{aligned} \tag{23}$$

The spatial dependence between the voxels will now be included by allowing a spatial covariance between  $\nu_i$ ,  $i = 1, \dots, V$ . We will use a scaled version

$$\begin{aligned}
\epsilon_1(t) &= \rho_{11}\epsilon_1(t-1) + \rho_{12}\epsilon_1(t-2) + \sigma_1 h_1(t) \\
\epsilon_2(t) &= \rho_{21}\epsilon_2(t-1) + \rho_{22}\epsilon_2(t-2) + \sigma_2 h_2(t) \\
&\vdots \\
\epsilon_V(t) &= \rho_{V1}\epsilon_V(t-1) + \rho_{V2}\epsilon_V(t-2) + \sigma_V h_V(t)
\end{aligned} \tag{24}$$

such that  $h_i(t)$  have unit variance. In vector form we have

$$\boldsymbol{\epsilon}(t) = \Phi_1 \boldsymbol{\epsilon}(t-1) + \Phi_2 \boldsymbol{\epsilon}(t-2) + S \mathbf{h}(t). \tag{25}$$

Hence

$$\mathbf{h}(t) = S^{-1}(\boldsymbol{\epsilon}(t) - \Phi_1 \boldsymbol{\epsilon}(t-1) - \Phi_2 \boldsymbol{\epsilon}(t-2)) \tag{26}$$

where  $\Phi_k$  is a diagonal matrix containing the  $k$ -th autoregressive parameters at each voxel, and  $S$  is a diagonal matrix containing the voxel wise standard deviations. These matrices are of dimension  $V \times V$ . Now we assume that  $\mathbf{h}(t) \in N(0, \Sigma(\theta))$  where the covariance matrix is defined using a spatial covariance function with parameters  $\theta$ . Note that this method allows for spatial dependence between voxels at each time point. There is no inclusion of spatial dependence between voxels at different time points, ie we assume that  $\mathbf{h}(t)$  is independent from  $\mathbf{h}(t')$  where  $t \neq t'$ .

### 3.7.1 Procedure To Fit The Model

The algorithm to fit the spatio-temporal model as described in 3.7 is as follows.

1. Use restricted maximum likelihood to estimate the  $\beta_i$ ,  $\sigma_i$ ,  $\rho_{i1}$  and  $\rho_{i2}$  at each voxel  $i$ .

2. At each voxel  $i$ , obtain the residuals  $Y_i(t) - X\boldsymbol{\beta}_i = \epsilon_i(t)$ , where  $\epsilon_i(t) = \rho_{i1}\epsilon_i(t-1) + \rho_{i2}\epsilon_i(t-2) + \nu_i(t)$ .
3. Calculate  $\nu_i(t) = \epsilon_i(t) - \hat{\rho}_{i1}\epsilon_i(t-1) - \hat{\rho}_{i2}\epsilon_i(t-2)$  where  $\nu_i(1) = \epsilon_i(1)$  and  $\nu_i(2) = \epsilon_i(2)$ .
4. Using  $\hat{\sigma}_i^2$  construct the matrix  $S$ , and multiply  $S^{-1}$  by  $\boldsymbol{\nu}(t)$  to give estimates of  $\mathbf{h}(t)$ .
5. As  $\mathbf{h}(t) \in N(0, \Sigma(\theta))$ , estimate  $\theta$  by maximising the log-likelihood

$$l \propto -\frac{T}{2} \log |\Sigma(\theta)| - \frac{1}{2} \sum_{t=1}^T \hat{\mathbf{h}}^T(t) \Sigma^{-1}(\theta) \hat{\mathbf{h}}(t). \quad (27)$$

As  $\Sigma(\theta)$  is a  $V \times V$  matrix, it will be computationally intensive to maximise Equation (27) for large  $V$ . For the forms of  $\Sigma(\theta)$  that we consider in subsequent sections, there are no closed form solutions. There are several methods one could use to speed up the calculations at the expense of some approximation, for example see [57], [29] or [45]. However, in this case, although the computation is time-consuming it is still feasible without using approximations.

### 3.7.2 Simulating Spatio-Temporal Data

The algorithm to simulate Spatio-Temporal data is as follows.

1. Choose some parameters  $\boldsymbol{\beta}$  ( $\dim 4 \times V$ ),  $\boldsymbol{\rho}$  ( $\dim V \times 2$ ),  $\boldsymbol{\sigma}$  ( $\dim V \times 1$ ) and  $\theta$  ( $\dim 1 \times 1$ ) for a set of  $V$  voxels.
2. Simulate  $\mathbf{h} \in N(0, \Sigma(\theta))$  using the following steps.
3. Construct  $\Sigma$  using the distance matrix between the voxels, and a spatial covariance function with parameters  $\theta$ .
4. Calculate the Choleski decomposition  $\Sigma = LL^T$  where  $L$  is a lower triangle matrix.
5. Obtain a realisation  $\mathbf{e} \in N(0, I_V)$ , where  $I_V$  is the  $V$  by  $V$  identity matrix, then  $\mathbf{h} = L\mathbf{e} \in N(0, \Sigma)$

6. Repeat Step 5  $T$  times to obtain  $\mathbf{h}(t)$  for  $t = 1 \dots T$ .
7. Construct the matrix  $S = \text{diag}(\boldsymbol{\sigma})$ , and then  $\boldsymbol{\nu}(t) = S\mathbf{h}(t)$  for  $t = 1 \dots T$ .
8. Add in the autoregressive part of the model by letting  $\epsilon_i(1) = \nu_i(1)$ ,  $\epsilon_i(2) = \nu_i(2)$  and  $\epsilon_i(t) = \rho_{i1}\epsilon_i(t-1) + \rho_{i2}\epsilon_i(t-2) + \nu_i(t)$  for  $t = 3, \dots, T$ .
9. Generate a response vector  $\mathbf{Y}_i = X\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$  at each voxel  $i$ .

Table 10 shows some results of simulating spatial data and estimating the parameters as described in Section 3.7.1. The simulated data was generated to be similar to the brain data in the smallest ROI initially. In the table the true values refer to the values of the parameters used to generate the data. For example  $\beta_0 = 500$  for all voxels in the region and for all simulations. For the purpose of testing the estimation algorithm, the spatial covariance function is set to be

$$\Sigma(\theta) = \theta^D \tag{28}$$

where  $D$  is the  $V \times V$  Euclidean distance matrix between the voxels within a region and  $|\theta| < 1$ . Note that in this case a higher value for  $\theta$  equates to a higher correlation between voxels.

The results of the simulation was a set of parameters for each voxel and for each simulation. Therefore here the results gave an array of size 4 by  $V$  by  $nsim$  (for the mean parameters) where  $V$  is the number of voxels in the ROI, and  $nsim$  is the number of simulations. To obtain the mean of the estimates for the mean parameters, the average of the average for each simulation was taken, the standard deviation (SD) is calculated as the square root of the average of the variances for each simulation, and the SEM is calculated in a similar way. The same logic was used to calculate the statistics for the other parameters.

Parameter	True Value	Mean	SD	SEM
$\beta_0$	500	500.03	3.82	0.26
$\beta_1$	2	2.04	6.33	0.44
$\beta_2$	2	1.84	9.60	0.66
$\beta_3$	2	2.00	0.27	0.02
$\rho_1$	0.15	0.12	0.13	0.01
$\rho_2$	-0.30	-0.29	0.12	0.01
$\sigma$	30	30.26	5.71	0.39
$\theta$	0.85	0.85	0.02	0.001

Table 10: Spatial model parameter estimates mean and standard deviation from 2000 simulations of voxels in region 41 (211 voxels)

From Table 10 it can be seen that the estimation procedure proposed gives adequate estimates for the parameters of the model.

## 3.8 Application To Brain Data

### 3.8.1 The Mean Parameters

As the mean parameters in the final model 3.7 are estimated using the temporal model 3.6, here it is computationally possible to estimate the parameters for all voxels in the brain. Figure 22 shows some results. The blue lines in all plots show the mean of the estimates of the parameter in each region of interest. The red lines show the mean plus and minus two standard deviations of the estimates of the parameters in each region of interest. From the figure it can be seen that all parameters are highly variable. In particular it can be seen that the mean of all parameters except the intercept term are within two standard deviations of zero for all regions of interest. This is interesting as if the estimates of the BOLD response parameters  $\beta_1$  and  $\beta_2$  are close to zero, this would imply inactivity. This will be returned to in Section 4 where the voxels will be formally tested for activation.

In Section 2 we saw that when adding the time trend to the intercept, the reciprocal condition number dropped to near zero ( $2.88 \times 10^{-4}$ ). We then centered the time trend, and the RCN reduced to 0.005 when this covariate was added to the intercept alone. This is still suggesting a poorly conditioned matrix, and therefore removing this covariate may give better parameter estimates. Here we will investigate if the time trend is required

in the model. Applying the temporal model to all voxels, we can test for each voxel if  $\beta_3 = 0$ . If this is true for the majority of voxels, then it may make sense to remove the covariate. Applying the model to all 228453 voxels resulted in 94407 which had significant p-values when testing at the 5% level, implying the time trend is required in these cases. This is 41% of all voxels and so it seems that the time trend is required for a large proportion of the voxels.

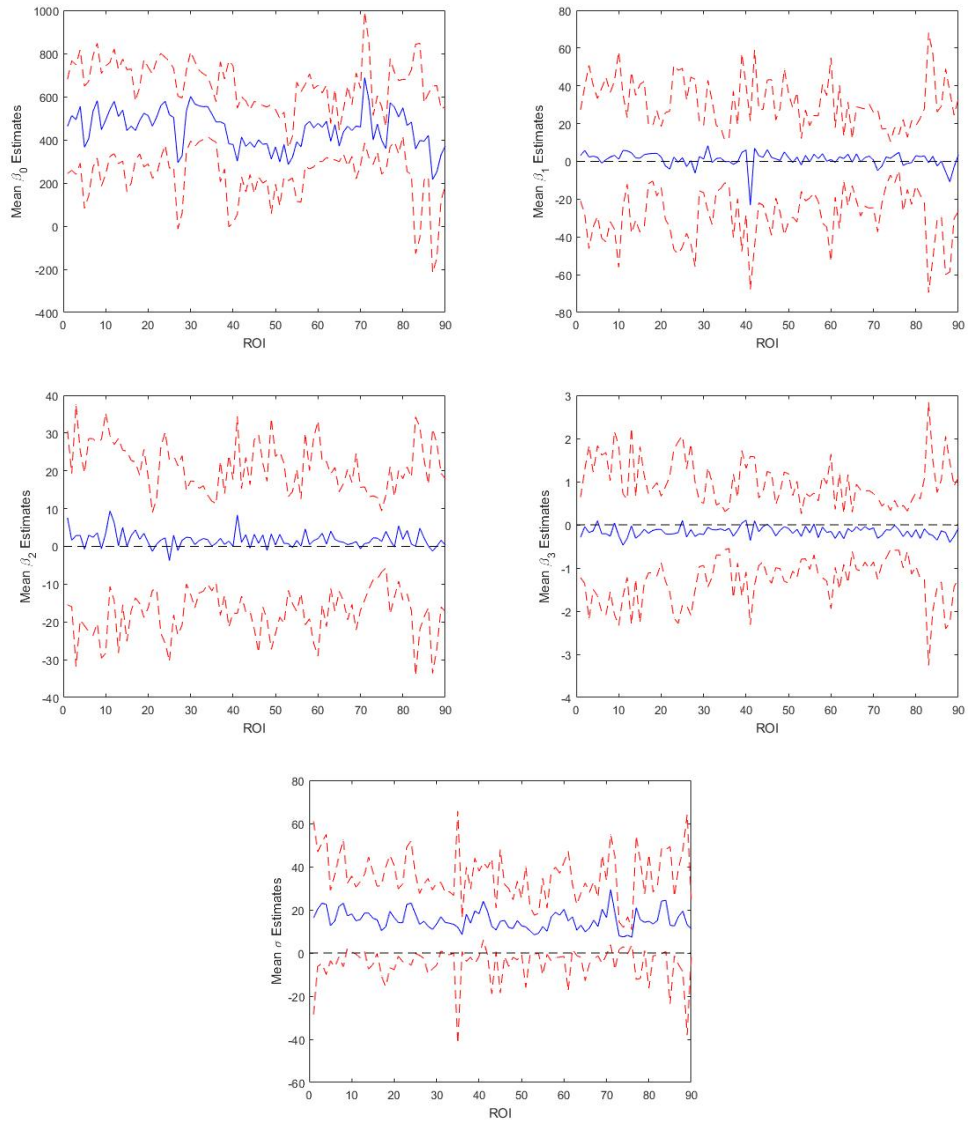


Figure 22: Mean (blue) and mean plus/minus two standard deviation (red) of the parameter estimates in all regions of interest. Black dashed line shows reference line at zero.



### 3.8.2 The Autoregressive Parameters

Figure 23 shows the mean of all autoregressive parameters in each region of interest. From the plot we can see that in each ROI there is a lot of variability of the values of the parameter in the region.

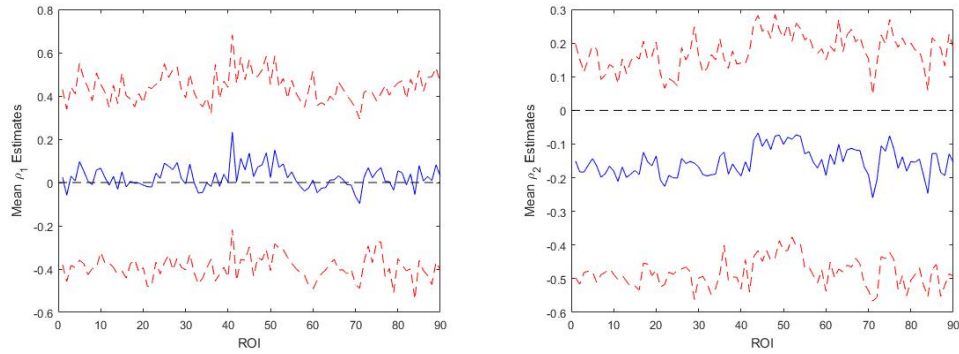


Figure 23: The mean of the autoregressive parameters in each ROI (blue) and plus/minus two standard deviations (red dashed). Black dashed line shows reference line at zero.

Figure 24 shows some plots of the values of the autoregressive parameter on the full brain. See Section 3.6.1 for some examples of the parameters on some ROIs. From these plots it is difficult to see any real pattern, though there are some areas where the parameter looks to be somewhat clustered, suggesting some spatial correlation between the estimates.

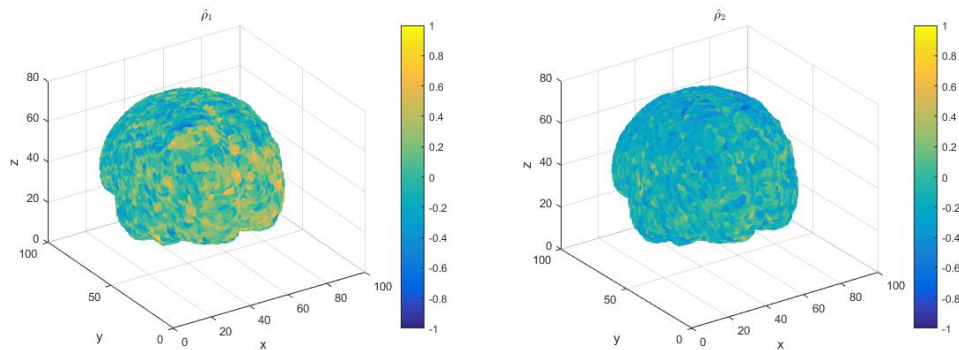


Figure 24: The estimated autoregressive parameters on the surface of the brain.

### 3.8.3 The Spatial Parameter: Estimating The Spatial Covariance Using Different Functions

In this section the spatial parameter will be estimated for the regions of interest using some different spatial covariance functions. The first covariance function that will be considered will make the assumption that

$$\Sigma(\theta) = \theta^D \tag{29}$$

where  $D$  is the  $V \times V$  Euclidean distance matrix between the voxels within a region and  $|\theta| < 1$ . Note that in this case a higher value for  $\theta$  equates to a higher correlation between voxels. Figure 25 shows the value of the spatial parameter estimate for each ROI and Table 11 shows some of the interesting results. Note that the figure does not show the standard deviations as they are so small it makes it difficult to see the estimates clearly. In all cases the parameter estimate is fairly high indicating spatial correlation between voxels within regions of interest. Some interesting points are the two regions with highest spatial dependence which are region 41 and 71 with spatial parameter estimates of 0.85 and 0.88 respectively. Also of note is region 87 which has by far the lowest estimate of 0.63, but still indicates a high correlation between voxels in that region.

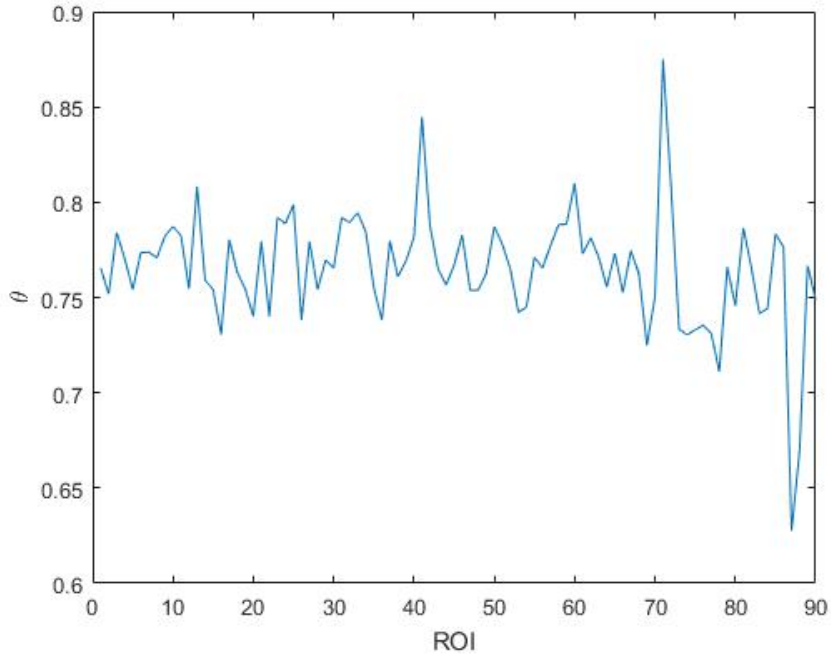


Figure 25: The estimate of the spatial parameter in each ROI.

Region	Number of Voxels	$\hat{\theta}$	SD	Time
3	3262	0.78	0.02	14 mins
7	4338	0.77	0.01	19 mins
8	4865	0.77	0.01	54 mins
21	278	0.78	0.05	6 secs
41	211	0.85	0.02	5 secs
71	965	0.88	0.04	58 secs

Table 11: Estimates of  $\theta$  for some different regions of interest assuming covariance function (29).

From Table 11 it can be seen that for the set of different regions of interest, the value of the parameter  $\theta$  is fairly high in most cases. The standard deviation was obtained here using the output Hessian matrix from the Matlab function *fmincon*.

An alternative covariance function that will be considered is the Matérn covariance function given by

$$\Sigma(\alpha, \nu) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{d\sqrt{2\nu}}{\alpha} \right)^\nu K_\nu \left( \frac{d\sqrt{2\nu}}{\alpha} \right). \quad (30)$$

where  $d$  is the distance between two voxels, and  $K_\nu$  is the modified Bessel function of the second kind of order  $\nu$ . Applying this to some regions of interest gives the following parameter estimates.

Region	Number of Voxels	$\hat{\alpha}$	SD( $\hat{\alpha}$ )	$\hat{\nu}$	SD( $\hat{\nu}$ )	Time
3	3262	1.53	0.01	2.44	0.03	56 mins
7	4338	1.51	0.03	2.39	0.05	130 mins
8	4865	1.48	0.01	2.49	0.02	190 mins
21	278	1.62	0.02	1.93	0.04	13 secs
41	211	1.80	0.03	2.20	0.07	20 secs
71	965	1.99	0.01	2.16	0.01	3 mins

Table 12: Estimates of  $\alpha$  and  $\nu$  for different regions of interest assuming Matérn covariance function (30).

Table 13 shows a comparison of the covariance between voxels separated by different distances in region of interest seven (chosen at random) using the different covariance functions discussed. Here the aim is to see the similarity and differences between the covariances at different distances using the two covariance functions. The Matérn parameters used are the estimates for those in region seven given in Table 12, and the exponential parameter used is the estimate for the parameter in region seven, as given in Table 11.

Distance, $d$	$\theta^d$	Matérn
1	0.770	0.727
2	0.593	0.354
3	0.457	0.141
4	0.352	0.050
5	0.271	0.017

Table 13: Estimates of the covariance between voxels separated by distance  $d$  for regions of interest seven using different covariance functions.

From Table 13 we can see that the covariance function (29) and the Matérn gives very different results. Although the lag one correlation is quite close, the Matérn correlation decays much faster than the exponential covariance function.

Figure 26 shows a comparison of the exponential and Matérn fits to the brain data. In this plot, the correlation between the residual time series  $\mathbf{h}(t)$ , (26), is plotted against distance, and the fitted exponential and Matérn correlations are shown on top.

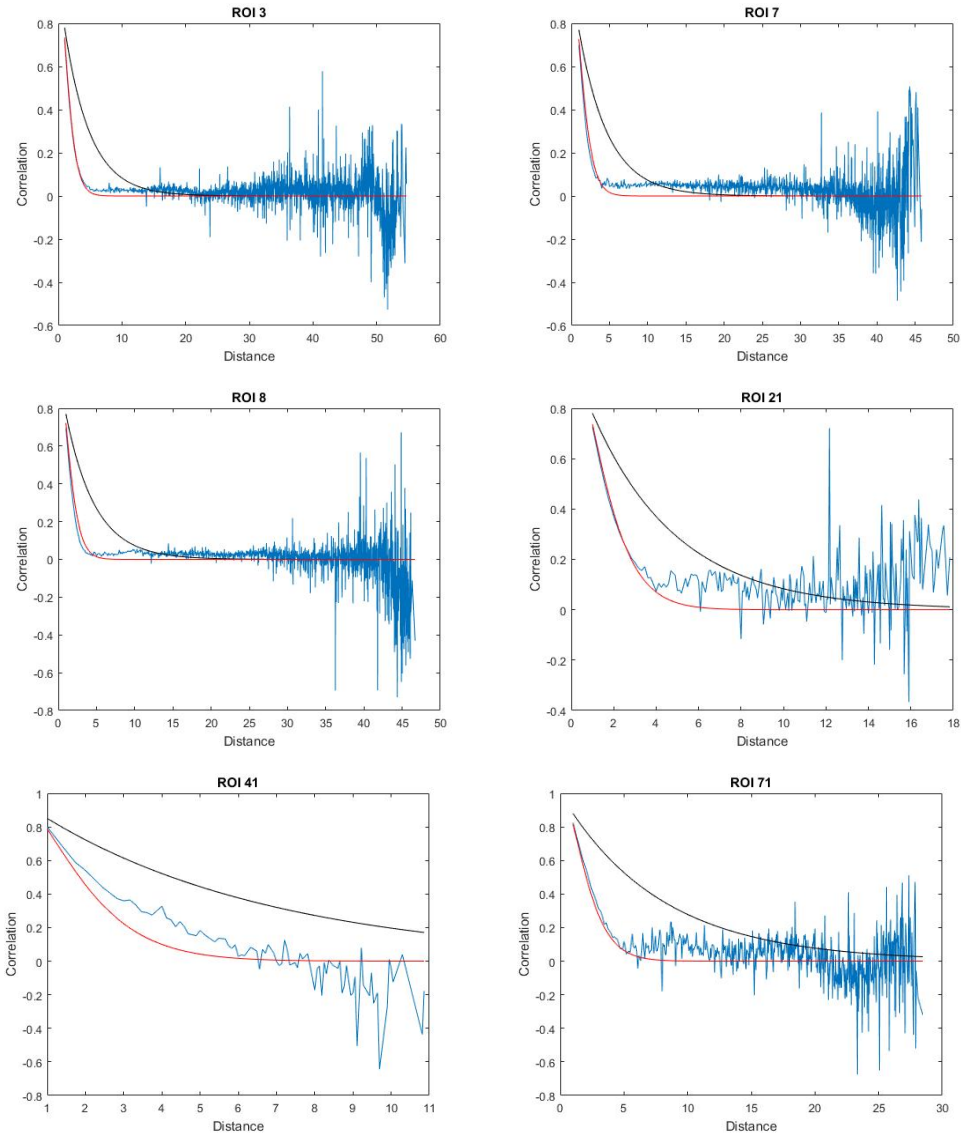


Figure 26: The correlation between estimated residuals  $\mathbf{h}(t)$  against distance (blue) with estimated spatial correlation (exponential in black, Matérn in red)

From Figure 26 it can be seen that for this selection of regions of interest, the Matérn correlation function is a better fit to the residuals than the exponential. We can also see that at larger distances the observed residual

correlation is more erratic than at smaller distances. This is probably due to the low number of pairs of voxel residuals which are separated by these larger distances, ie most of the voxels have a small separation distance. Figure 28 shows histograms of the numbers of pairs of voxels against separation distance corresponding to the regions of interest we have been working with. Binning the distances gives a smoother observed curve (see Figure 27).

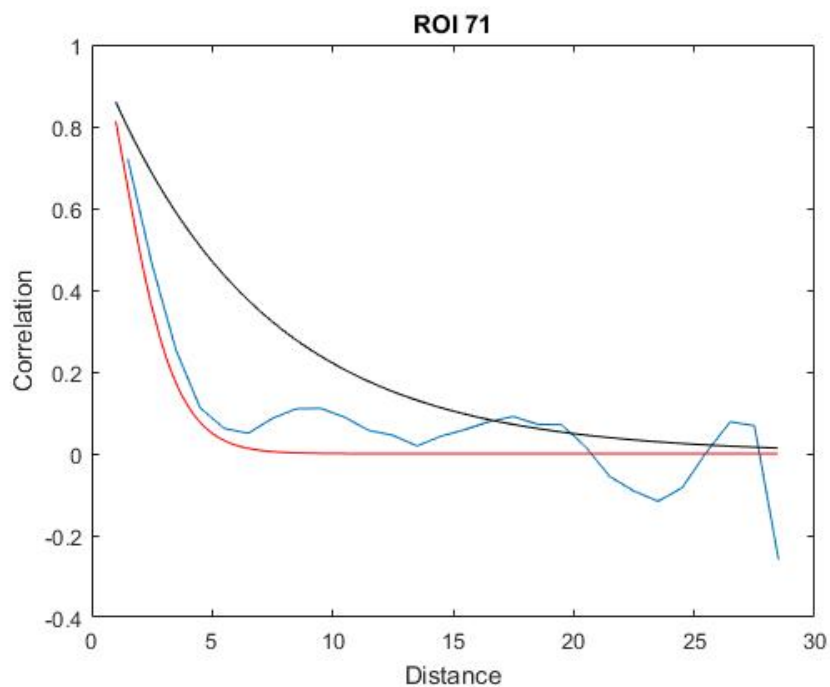


Figure 27: The correlation between estimated binned residuals  $\mathbf{h}(t)$  against distance (blue) with estimated spatial correlation (exponential in black, Matérn in red)

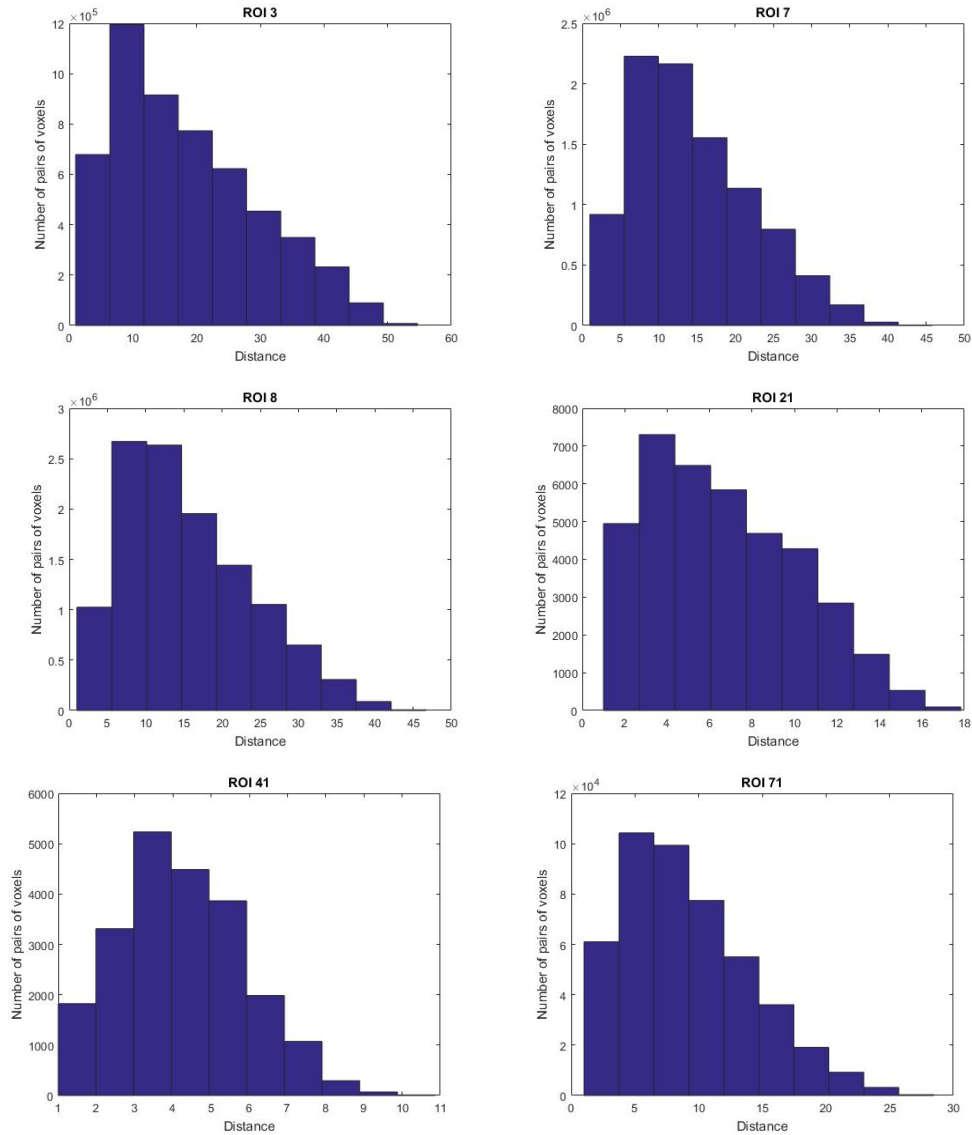


Figure 28: Histograms of the number of unique pairs of voxels separated by different distances in some regions of interest.

One assumption made up until this point is that the spatial covariance function has constant parameter values over the course of the experiment. This may not be true, and one could assume for example, that the spatial parameter varies at each stage of the experiment. The next covariance function



that will be considered is of the form

$$\Sigma(\boldsymbol{\theta}) = \begin{cases} \theta_1^D, & \text{if } 1 \leq t \leq 12 \\ \theta_2^D, & \text{if } 12 < t \leq 24 \\ \theta_3^D, & \text{if } 25 < t \leq 36 \\ \theta_4^D, & \text{if } 37 < t \leq 48 \end{cases}$$

Table 14 shows the estimates of the spatial parameters for a few different regions of interest.

ROI	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$
3	0.88 (0.001)	0.61 (0.001)	0.86 (0.007)	0.87 (0.004)
7	0.89 (0.003)	0.65 (0.55)	0.83 (0.004)	0.89 (0.002)
8	0.87 (0.002)	0.59 (0.005)	0.84 (0.006)	0.88 (0.001)
21	0.77 (1.03)	0.54 (3.75)	0.85 (2.43)	0.87 (2.63)
41	0.88 (0.002)	0.75 (0.006)	0.87 (0.002)	0.92 (0.003)
71	0.92 (0.003)	0.77 (0.002)	0.92 (0.002)	0.93 (0.001)

Table 14: Spatial parameter estimates (and standard deviation) when allowing for a different exponential parameter in each of the four periods of the experiment.

From Table 14 we can see that the spatial correlation drops in all cases during the second period, ie the first period of activity.

An alternative method to determine if the spatial correlation is time dependent is to introduce a time dependent spatial parameter. The next two functions to be considered are the time dependent exponential given by

$$\Sigma(\theta, t) = e^{-\frac{D}{\theta_0 + \theta_1 t + \theta_2 t^2}} \quad (31)$$

and the time dependent Matérn given by (30) with  $\alpha = \theta_0 + \theta_1 t + \theta_2 t^2$ . These two spatial functions allow the spatial parameter to vary with time. Tables 15 and 16 shows the results.

ROI	$\theta_0$	$\theta_1$	$\theta_2$
3	2.80 (4.25)	0.001 (1.12)	0.002 (0.03)
7	2.64 (4.45)	0.001 (0.83)	0.002 (0.02)
8	2.78 (58.85)	0.001 (1.31)	0.002 (0.03)
21	3.32 (9.38)	0.001 (1.25)	0.001 (0.03)
41	4.66 (13.09)	0.001 (1.07)	0.002 (0.02)
71	5.27 (0.10)	0.001 (0.001)	0.004 (0.001)

Table 15: Spatial parameter estimates (and standard deviation) using the time dependent exponential

ROI	$\theta_0$	$\theta_1$	$\theta_2$	$\nu$
3	1.55 (8.39)	0.001 (0.91)	0.001 (0.17)	1.20 (0.95)
7	1.59 (7.53)	0.001 (1.43)	0.001 (0.05)	1.34 (1.08)
8	1.43 (8.01)	0.001 (0.77)	0.001 (0.09)	1.27 (0.86)
21	1.80 (15.56)	0.001 (1.89)	0.001 (0.03)	0.92 (2.45)
41	1.99 (38.46)	0.001 (4.72)	0.001 (0.10)	1.17 (1.01)
71	2.10 (158.85)	0.001 (19.97)	0.001 (0.40)	1.34 (10.84)

Table 16: Spatial parameter estimates (and standard deviation) using the time dependent Matérn

From Tables 15 and 16 it appears that the additional parameters  $\theta_1$  and  $\theta_2$  are estimated near zero in all cases, indicating that they are not needed. However note that the parameter estimate standard deviations are very large in these cases.

One method of determining which of the covariance functions gives a better model fit is to use the AIC (7). Here the log-likelihood given in Equation (27) will be evaluated at the maximum for the covariance models discussed in order to compute the AIC. As usual the model which gives the lowest AIC is the preferable model. Table 17 shows the AIC for a selection of the regions of interest.

ROI	$\theta^d$	Matérn	$e^{-\frac{d}{\theta_0 + \theta_1 t + \theta_2 t^2}}$	Matérn (time trend)	Varying $\theta$
3	-0.75	-1.38	-0.80	-1.16	-1.05
7	-0.96	-1.77	-1.13	-1.58	-1.40
8	-0.93	-1.84	-1.09	-1.62	-1.39
21	-0.04	-0.09	-0.05	-0.06	-0.05
41	-0.08	-0.11	-0.08	-0.09	-0.08
71	-0.47	-0.66	-0.49	-0.63	-0.55

Table 17: The AIC for some regions of interest using different covariance functions (multiply all figures given by  $10^5$  for true values )

From Table 17 it can be seen that in all cases, the Matérn covariance function with no time trend gives the lowest AIC, which indicates that this is the best model.

### 3.9 Residual Analysis

In this section, the residuals of the model described in Section 3 will be analysed for a number of reasons. Firstly, we can assess model fit. If our statistical model is a good one, then the residuals should be approximately noise. The second reason is that the residuals can be used to assess some of our model assumptions. One assumption was that the residuals are spatially correlated, and that this correlation is independent of time. Finally, as we modeled the residuals and found evidence of spatial correlation, the residuals can be used to analyse this pattern of spatial correlation further, by using methods such as clustering.

In all of the following, the residuals of the model,  $\mathbf{h}(t)$  are given in Equation (26).

#### 3.9.1 Residuals Autocorrelation

If a good statistical model is fit to data, then the resulting residuals of such a model will be approximate white noise. In this section we will formally test if the residuals at each individual voxel are noise, or if there is any evidence of correlation still remaining in the residual time series. One method to test this is to use the Ljung-Box Test as described in Section 2.3.3.

Applying the Ljung-Box test to all voxel residuals in the brain resulted in Figure 29 which shows the proportion of voxels which still have correlation

when testing at the 5% significance level in all regions of interest.

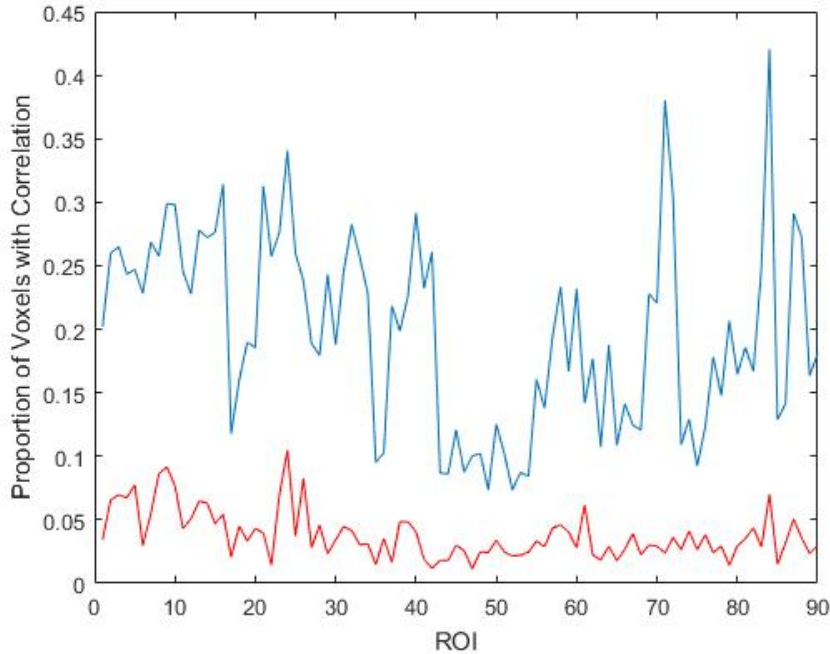


Figure 29: The proportion of residuals in each ROI which have evidence of correlation according to the Ljung-Box test. Shown are results using residuals of a model with  $AR(1)$  errors (blue), and the model with  $AR(2)$  errors (red).

From Figure 29 we can see that the average across all regions of residuals which contain some correlation is approximately 20% when applying a model with first order autoregressive errors. When allowing for second order autoregressive errors, there is a substantial decrease in the number of residuals which have evidence of autocorrelation. This is further evidence that the second order autoregressive model is more suitable to use for this data set.

### 3.9.2 Residual Noise

The statistical model assumes that the model residuals  $\mathbf{h}$  are standard normal at each voxel (see Section 3.7). Then we later assume spatial correlation between the residuals across different voxels. In this section we will investigate if the assumption that  $\mathbf{h}$  are standard normal at each voxel. If the

assumption is valid, then we should find that the estimated residuals are independent and approximately 95% should have value between  $-1.96$  and  $1.96$ . Also as the residuals are assumed to be uncorrelated noise, plotting them against the fitted values should show a random scatter of points. Given there are over 150,000 residual time series that we could plot, a random set of six are shown in Figure 30. Furthermore, Figure 32 shows histograms of the residuals in six different regions of interest.

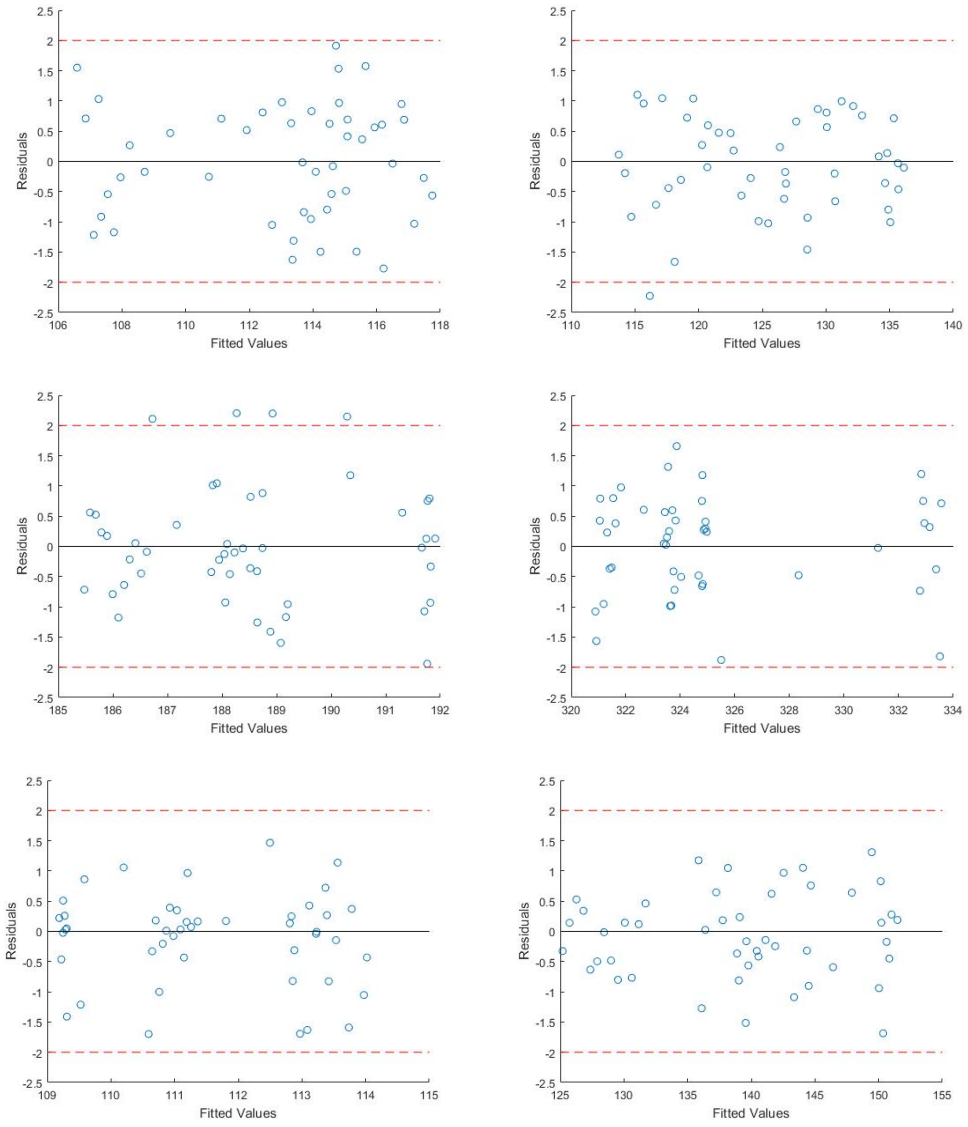


Figure 30: The model residuals plotted against the fitted values for six random voxels

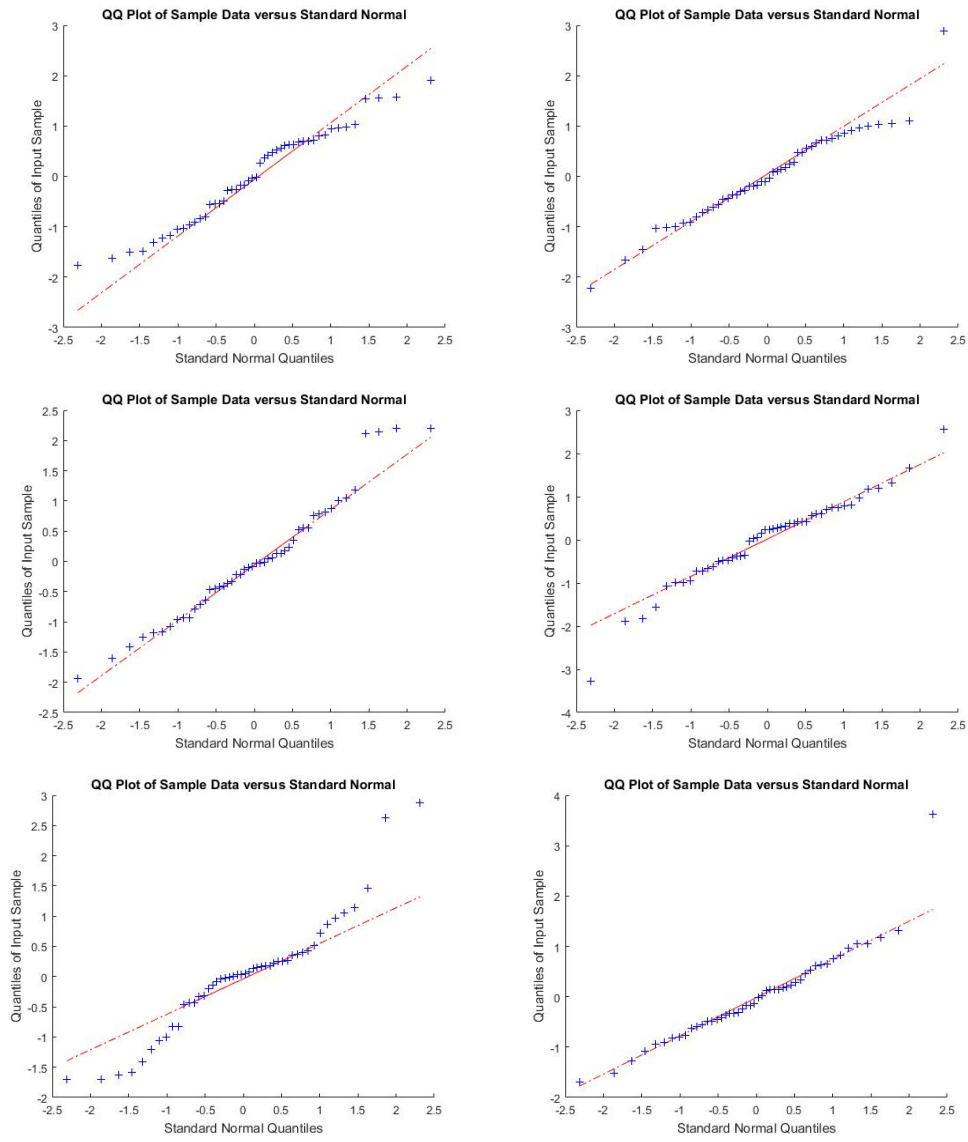


Figure 31: Q-Q plots of model residuals for six random voxels

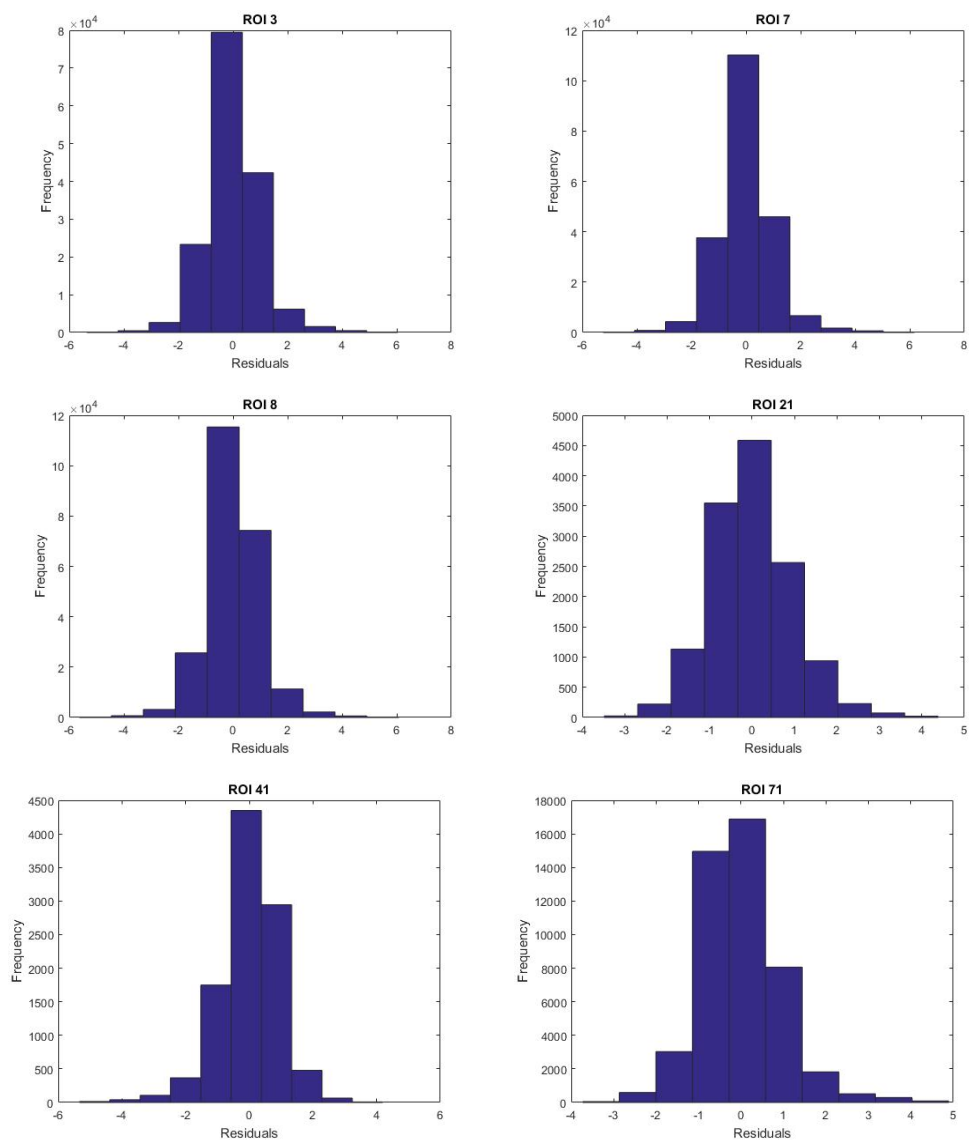


Figure 32: Histograms of the residuals in some regions of interest



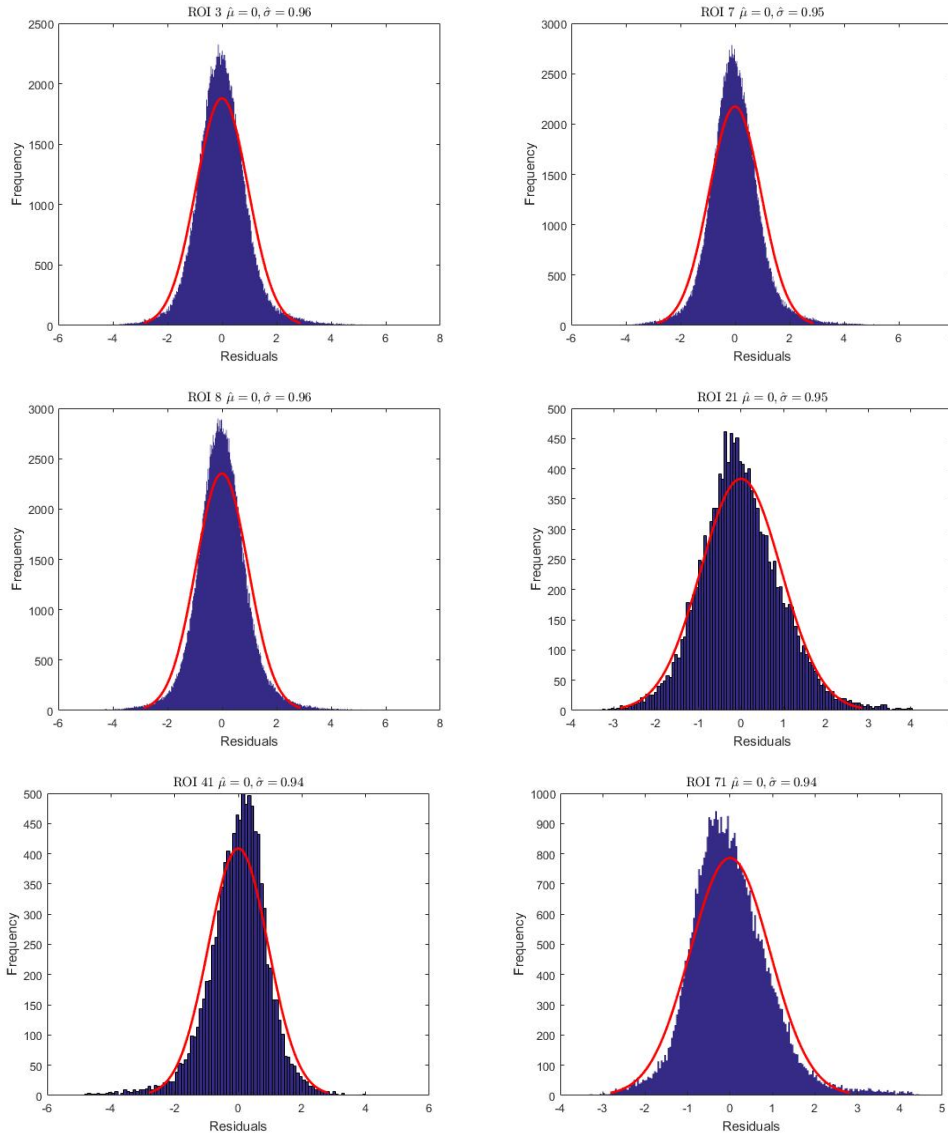


Figure 33: Histograms of the residuals in some regions of interest, with fitted normal distribution (red).

The plots in this section show that the model residuals are approximately standard normal, and uncorrelated suggesting a good model fit.

### 3.9.3 Temporal Correlation Between Residuals

In the previous section we looked at determining if each individual residual time series contained correlation or if they were just noise. Even though the results showed that the majority of voxels were not correlated, it would be interesting to investigate if the residual noise at each voxel is similar across voxels. In this section we will investigate if the residual time series between voxels are correlated. It is not feasible to test this for all pairs of voxels, therefore for each region of interest we will test 1000 random pairs of voxels for correlation using permutation tests.

For each pair of voxels we can calculate an observed correlation between the residuals at those voxels. By randomly permuting one (or both) of these vectors and recalculating the correlation we can obtain a new correlation value. Repeating this process many times will result in a distribution of correlation values. As we are calculating the values based on random permutations, the resulting distribution describes the correlations under the assumption that there is no order in the data. From this distribution we can compare the observed correlation to obtain a  $p$ -value of the likelihood of observing this correlation if there is no underlying order to the data. The  $p$ -values are obtained by counting the number of permuted correlations which are larger in magnitude than the observed one. This process will be carried out for all regions of interest, using 1000 permutations in each case for each pair of voxels. Figure 34 shows the results.

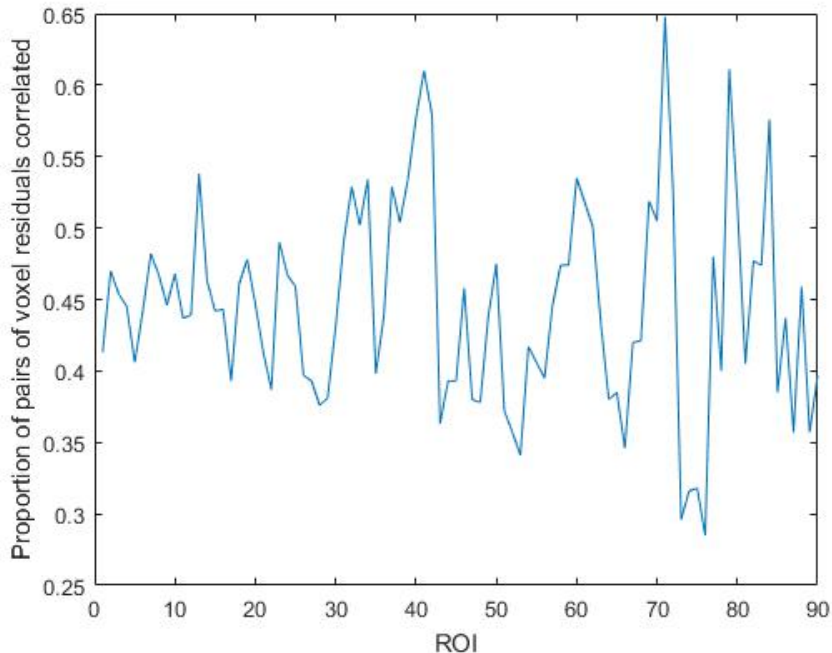


Figure 34: The proportion of pairs of voxel residuals which are correlated in each region of interest. In each region, 1000 pairs were assessed using a permutation test

From Figure 34 we can see that on average around 40–50% of the assessed pairs of voxels in each region were significantly correlated when assessing at the 5% level. Therefore, although we have seen that the voxel residuals are mostly noise, this shows that the residual noise at different voxels are correlated in a large number of cases.

Aside from testing for correlation between voxels within a region, we can also repeat this test for correlation of voxels between regions. One way of doing this is to look for correlation between regional means. Let  $h_v(t)$  be the residual at voxel  $v$  at time  $t$ . Then we define the regional mean for region  $i$  as

$$\bar{h}_i(t) = \sum_{v \in \mathcal{R}_i} \frac{h_v(t)}{|\mathcal{R}_i|} \quad (32)$$

where  $\mathcal{R}_i$  denotes region  $i$  and  $|\mathcal{R}_i|$  is the number of voxels in region  $i$ .

Figure 35 shows a histogram of p-values obtained when testing 1000 pairs of residual regional means for correlation using the permutation test.

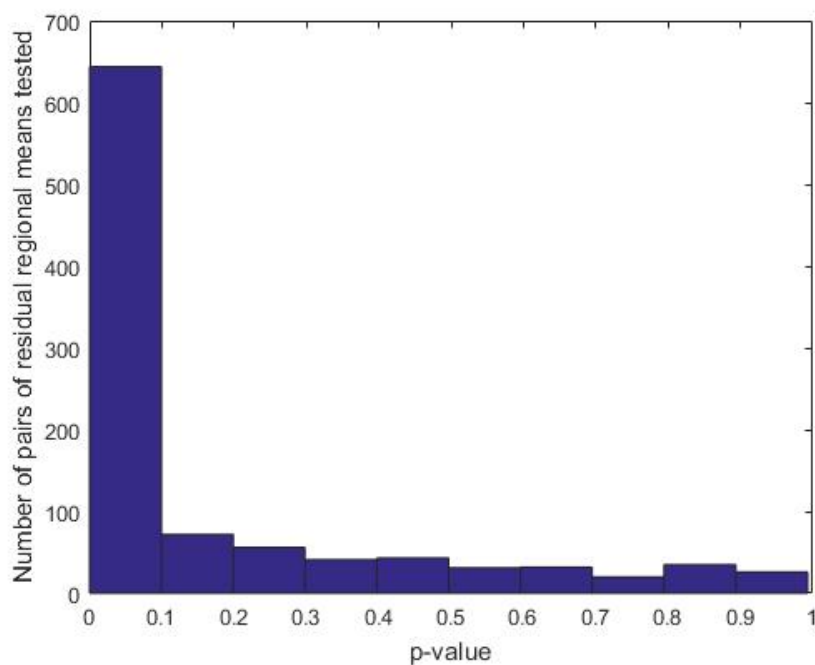


Figure 35: Histogram of p-values obtained when testing 1000 pairs of residual regional means for correlation using a permutation test.

From Figure 35 we can see a large number of significant p-values. In fact, 58% of the 1000 tested pairs of regional residual means had a p-value of less than 0.05 implying correlation.

In this section we have seen that a large proportion of the residual time series of voxels, and regional means are correlated. Although the residuals look noisy, there is correlation between this noise. This can be explained due to the fact that the residuals are spatially correlated, thus we would expect to observe correlation between pairs of voxel residuals.

## 4 Testing For Activation

### 4.1 Introduction

In this chapter we will look at testing the brain data for voxel activation. Following the work in chapter 3, as the covariates  $x_1$  and  $x_2$  of the model are the BOLD responses during the two different periods of activity in the experiment, the voxel activation test is the hypothesis test

$$\begin{aligned} H_0 : \beta_1 \text{ and } \beta_2 = 0 \\ vs \\ H_1 : \beta_1 \text{ or } \beta_2 \neq 0. \end{aligned} \tag{33}$$

All models described in chapter 3 assume normally distributed errors with unknown variance, thus one method of performing the test would be to test  $\beta_1 = 0$  and  $\beta_2 = 0$  independently using the usual t-test. However since we are testing two conditions simultaneously, it may be preferable to perform the activation test of both periods simultaneously, considering there is likely correlation between the two periods of activity. This can be achieved using the likelihood ratio test (LRT). The likelihood ratio test is used to compare two nested statistical models. Let  $L_g$  be the value of the likelihood of a general model evaluated at its maximum, and let  $L_s$  be the value of the likelihood of a smaller nested model evaluated at its maximum. Then the likelihood ratio test (expressed in terms of log-likelihoods) is given by

$$LRT = -2(\log(L_s) - \log(L_g)) \sim \chi_q^2 \tag{34}$$

where  $q$  is the difference in the number of parameters in each model. In the context of our problem, the larger general model will be that with both BOLD covariates (Equation (9)), and the smaller nested model is one without any BOLD response. Hence if we obtain a significant  $p$ -value using the  $LRT$ , then this will indicate that at least one of  $\beta_1$  or  $\beta_2$  is non-zero and required in the model, thus indicating voxel activation.

The rest of this chapter will be structured as follows. First some background information on multiple hypothesis testing will be given, then simulations will be used to explore correlation between the test statistics and to ensure the activation tests are working as expected, and finally testing will be applied to the brain data.

## 4.2 Multiple Hypothesis Testing

When performing multiple hypothesis tests care must be taken with regard to the type one error. If a single test is performed, the probability of rejecting the null hypothesis when it is true (reporting an active voxel when it is inactive) is

$$P(\text{Reject null}|\text{Null true}) = \alpha \quad (35)$$

where  $\alpha$  is some acceptable threshold level, usually set to 0.05. This is a false positive and so the probability of not reporting a false positive is  $1 - \alpha$ .

Suppose  $V$  of such independent tests are carried out, where  $V$  is the number of voxels, then the probability of reporting at least one false positive is

$$\begin{aligned} P(\text{At least 1 false positive}) &= 1 - P(\text{No false positives}) \\ &= 1 - (1 - \alpha)^V. \end{aligned} \quad (36)$$

If  $\alpha = 0.05$  and  $V = 211$  (the size of the smallest ROI), then the probability of obtaining at least one false positive is then 0.99998. In other words as the number of tests increase, the probability of falsely reporting active voxels becomes an almost certainty. This is a problem as even obtaining 5% false positives results in a lot of error if considering a large number of tests. For example a large region of 5000 voxels could result in 250 false positives, or the whole brain which consists of close to 200,000 voxels could result in reporting 10,000 voxels active when they are not. Therefore steps must be taken to reduce the type one error.

### 4.2.1 Bonferroni Correction

One basic method to reduce the type one error is the Bonferroni correction [23]. This method consists of changing the threshold for rejecting the null from  $\alpha$  to  $\alpha/V$ , and this can be shown to keep the probability of obtaining at least one false positive over all  $V$  tests to approximately  $\alpha$ . For example, using this correction with  $\alpha = 0.05$  and  $V = 211$  we have

$$\begin{aligned}
P(\text{At least 1 false positive}) &= 1 - P(\text{No false positives}) \\
&= 1 - \left(1 - \frac{\alpha}{V}\right)^V \\
&= 1 - \left(1 - \frac{0.05}{211}\right)^{211} \\
&= 0.0488.
\end{aligned} \tag{37}$$

Note that the probability obtained here is slightly below 0.05. The Bonferroni correction can be used on any data, independent or correlated, however it is known that for correlated data the Bonferroni correction can become extremely conservative [68]. This means that if the data is correlated and  $\alpha = 0.05$  the correction could lead to the probability of at least one false positive reducing to a value much smaller than 0.05. Furthermore, even for independent data, if  $V$  is large, the probability of rejecting the null for any individual test becomes extremely small. Thus the Bonferroni correction has the disadvantage of increasing the type two error. In other words it is much more likely to report inactivity even when the voxel is active.

#### 4.2.2 Benjamini-Hochberg Procedure

The Benjamini-Hochberg procedure [19] is another correction method which is now a popular method which is used to reduce type one error. The method is less conservative than the Bonferroni correction, however it has the advantage that it does not increase the type two error as extremely as the Bonferroni correction does. The Benjamini-Hochberg procedure can be used on independent or positively correlated test statistics [20] and is described as follows.

Suppose  $V$  hypothesis tests are performed. Order the  $p$ -values from low to high such that

$$P_1 < P_2 < \dots < P_V$$

Choose a desired level  $\alpha$  for which the overall number of false positives is to remain below then find the largest  $p$ -value which satisfies

$$P_k \leq \frac{k\alpha}{V}$$

and reject the null hypotheses corresponding to the  $p$ -values  $P_1, P_2, \dots, P_k$ .

### 4.2.3 Example Of How The Procedures Differ

In a subsequent section, simulated brain data will be used to observe how the Bonferroni and Benjamini-Hochberg procedures perform in terms of type one and two error. Here we will show a simple example where data is simulated with 1000 points such that the first 900 are from the  $N(0, 1)$  distribution and the remaining 100 are from the  $N(3, 1)$  distribution. The data are then tested to see if they are observations from a  $N(0, 1)$  distribution. The proportion of the first 900 data points which have a  $p$ -value of  $p < 0.05$  is the type one error, as we have simulated the data from the  $N(0, 1)$  distribution, though would reject this if testing at the 5% significance level. The proportion of the last 100 data points which have a  $p$ -value of  $p > 0.05$  when tested against the  $N(0, 1)$  distribution is the type two error, as we have simulated these data points from the  $N(3, 1)$  distribution, though would be confident were from the  $N(0, 1)$  distribution when testing at the 5% significance level.

In addition to looking at the type one and type two error without corrections, we will also apply the Bonferroni and Benjamini-Hochberg corrections to see how these impact the type one and type two errors. For example, in the case of the Bonferroni correction, we set the critical value for each individual test to be  $0.05/1000$ . As described in Section 4.2.1, this will keep the probability of obtaining at least one false positive to approximately 0.05. Table 18 shows the results.

Correction	Type 1 error	Type 2 error
None	0.044	0.16
Bonferroni	0.001	0.88
Benjamini	0.003	0.42

Table 18: Type 1 and type 2 error obtained when testing mixed data which is simulated from either the  $N(0, 1)$  or  $N(3, 1)$  distributions.

From Table 18 it can be seen that the Bonferroni correction does best in reducing the type one error. As noted in Section 4.2.1, if the number of tests is large, the correction gives very conservative results, and this comes at the cost of an increased type two error. The results show how the Benjamini-Hochberg procedure gives a reduced type one error (in comparison to no correction), without inflating the type two error as much as the Bonferroni correction.



When performing multiple tests it is worth considering if type one error or type two error is most important to the application. If type two error is not an issue, then the Bonferroni correction gives the lowest type one error. However, in general it is thought that in most applications that the Benjamini-Hochberg correction is the best option.

## 4.3 Brain Simulations

### 4.3.1 Correlation Between Test Statistics

Given that we are working with a spatial model (see Section 3.7), we would expect that test statistics for voxels will also be correlated and be dependent on the strength of the spatial correlation. In this section, this relationship will be investigated using simulated data. As in previous sections, data will be simulated for a smaller region of interest (ROI 41) containing 211 voxels. For each simulation spatial data will be generated, where all voxels will have the following parameter values:  $\beta_0 = 500, \beta_1 = \beta_2 = \beta_3 = 2, \sigma = 30, \rho_1 = 0.15$  and  $\rho_2 = -0.30$ . The spatial covariance function in Equation (29) will be used.

For each simulation we will estimate all parameters, and obtain a test statistic for the hypothesis  $H_0 : \beta_1 = 0$ . Hence we will obtain 211 test statistics for each simulation

$$\mathbf{t}_1 = (t_1, \dots, t_{211})$$

where

$$t_1 = \frac{\hat{\beta}_{11}}{sd(\hat{\beta}_{11})}$$

and  $\hat{\beta}_{11}$  is the estimate of  $\beta_1$  for the first voxel. To see how to correlation between test statistics change as the spatial parameter changes, another vector of length 211 will be formed of test statistics of neighbouring voxels within one unit distance

$$\mathbf{t}_2 = (t_1^*, \dots, t_{211}^*)$$

where  $t_1^*$  is the t-statistic corresponding to a voxel of one unit distance from the first voxel and so on. For each simulation  $corr(\mathbf{t}_1, \mathbf{t}_2)$  can be calculated and will tell us the correlation between neighbouring test statistics. This procedure will be repeated for 100 simulation and the mean will be taken. Furthermore, the procedure will be repeated for a set of values of  $\theta$  to see how

the correlation in test statistics changes in relation to the change in the spatial dependence on the field. Table 19 shows the results. A similar procedure can also be implemented for the t-statistics for the second activation period, ie testing if  $\beta_2 = 0$ .

$\theta$	Correlation between $\beta_1$ test stats	Correlation between $\beta_2$ test stats
0	-0.01	-0.01
0.1	0.09	0.08
0.2	0.18	0.18
0.3	0.26	0.27
0.4	0.35	0.34
0.5	0.43	0.42
0.6	0.48	0.51
0.7	0.56	0.56
0.8	0.61	0.63
0.9	0.68	0.66
0.99	0.67	0.68

Table 19: Correlation between the test statistics for  $\beta_1 = 0$  and  $\beta_2 = 0$  for neighbouring voxels of one unit distance separation. Results obtained using the mean of 100 simulations of 211 voxels in ROI 41.

From Table 19 it can be seen that as the spatial dependence increases, then so does the correlation between test statistics at neighbouring voxels as expected. One final thing of interest is the correlation between the test statistics for the estimates of  $\beta_1$  and  $\beta_2$  as the spatial parameter changes. To investigate this, the above simulation is used and the test statistics at all voxels are calculated. From this we can find the correlation as required. The result were that the correlation did not vary noticeably with a change in  $\theta$ , and remained at an almost constant level of approximately 0.25. Noticing that this is the value of the autoregressive parameter  $\rho$  in all simulations, the simulation was repeated, this time holding  $\theta$  constant and allowing  $\rho$  to vary, and this had no effect on the correlation between any of the test statistics.

In this section we have seen using simulated data, that test statistics are correlated between voxels, and as spatial dependence increases, so does this correlation. The correlation between the test statistics for  $\beta_1 = 0$  and  $\beta_2 = 0$  does not depend on the spatial parameter as expected, however the simulation showed that there exists a correlation between the statistics.

Therefore using the likelihood ratio test to test both simultaneously may be better than testing them individually.

### 4.3.2 Testing For Activation

In Section 4.2 we noted that for any given activation test, there will be a probability of detecting a false positive. If the threshold for the critical value of a hypothesis test is set to 0.05, then there is a 5% chance of reporting a false positive. In this section the objective is to simulate data and test for voxel activation, ensuring that we obtain the correct number of false positives. If this is achieved the simulations will also be used to show the impact of model misspecification on the number of reported false positives.

The first thing to note is that all simulations will be of temporal data only, as when we fit the spatial model, because of the sequential nature of it, the temporal model is used to estimate all mean parameters. In all simulations the temporal data will this time be generated with parameters set to  $\beta_0 = 500$ ,  $\beta_1 = 0$ ,  $\beta_2 = 0$ ,  $\beta_3 = 2$ ,  $\sigma = 30$ ,  $\rho_1 = 0.15$  and  $\rho_2 = -0.30$  at all voxels. Note that we use  $\beta_1 = 0$  and  $\beta_2 = 0$  to simulate inactive voxels, with the aim of determining how many were inferred to be active, thus obtain a false positive rate. If we fit the same model to the data as used to generate the simulated data, the false positive rate should be 5% when testing at the 5% level. As described in Section 4.1, the likelihood ratio test will be used to determine if voxels are active. Here two models will be fit, one which includes two BOLD responses, and one which includes neither. If the likelihood ratio test gives a significant test statistic then the larger model is favoured, implying at least one of the BOLD responses is required. As the generated data is such that  $\beta_1 = 0$ ,  $\beta_2 = 0$ , this is indicative of a false positive.

Before presenting results, we briefly note an issue that arose when investigating this. Under the conditions described above, and using the likelihood ratio test we found that almost all voxels were active. This is because the restricted likelihood was used in the log-likelihood ratio test. Recall that the proposed estimation method for the temporal model is to fit using REML to simultaneously estimate all parameters. The first step of REML is to remove the fixed effects of a model, therefore the likelihood ratio test cannot be used to compare the fixed effects of competing models when REML is used [5]. Hence for voxel activation tests, the parameters will be estimated using REML, and the log-likelihood from regular maximum likelihood estimation

will be evaluated at these values in order to perform the likelihood ratio test.

Table 20 shows the percentage of false positives obtained when generating temporal data for differing numbers of voxels and different lengths of time series. All tests carried out at the 5% significance level. The table shows the results obtained via the likelihood ratio test, and also the results obtained if testing each of the parameters  $\beta_1$  and  $\beta_2$  separately using the traditional t-test.

Number of Voxels	Length	% FP (LRT)	% FP $\beta_1$	% FP $\beta_2$
100	48	7	7	2
1000	48	7.7	6.2	5.3
10000	48	7.6	6.3	5.4
10000	96	6.5	5.3	5.4
10000	144	6.1	5.6	5.3

Table 20: Percentage of false positives obtained when simulating temporal inactive data using the likelihood ratio test (LRT) or the t-test to test  $\beta_1$  and  $\beta_2$  separately.

From Table 20 we can see that the percentage of false positives is a little high when testing at the 5% level with time series of length 48. This is likely due to the likelihood ratio test statistic being asymptotically chi-squared, thus as the time series length increases, we can see the false positive level decrease to the expected level. The t-test results obtained are much closer in all cases to 5%.

An interesting thing to investigate is the impact of model misspecification on the number of false positives obtained. Table 21 repeats the above simulation, however now an independent model will be fit to the temporal data.

Number of Voxels	Length	% FP (LRT)	% FP $\beta_1$	% FP $\beta_2$
100	48	26	17	13
1000	48	20.1	14.5	13.7
10000	48	17.4	13.6	12.1
10000	96	16.5	12.4	11.9
10000	144	15.6	11.4	11.4

Table 21: Percentage of false positives obtained when simulating temporal inactive data and fitting an independent model, using the likelihood ratio test (LRT) or the t-test to test  $\beta_1$  and  $\beta_2$  separately.

From Table 21 we can see that when the incorrect model is used, the number of false positives increases substantially. Therefore it is important to ensure that the model fit is as good as possible.

#### 4.4 Application To Brain Data

In this section we will investigate activation in the brain data. In particular we will find the number of active voxels using the temporal model in different regions of interest and testing at the 5% level. The numbers are then adjusted using the different corrections. Note that although the number of false positives will be correctly reduced using the Bonferroni or Benjamini-Hochberg procedures, the type two error will be inflated as seen in 4.3. Therefore the reported number of active voxels may be a lot lower than the true number of active voxels. Table 22 show results for some different regions of interest when the temporal model is fitted. In all cases the results obtained via testing both activity periods simultaneously using the likelihood ratio test, or individually using the usual t-test are given. Also the Benjamini-Hochberg procedure is applied to adjust for multiple hypothesis testing as described in Section 4.2.2.

ROI	Voxels	LRT	T-test $\beta_1$	T-test $\beta_2$	LRT (Benjamini)
3	3262	1177 (35%)	1174 (36%)	294 (9%)	587 (18%)
7	4338	1605 (37%)	1751 (40%)	304 (7%)	954 (22%)
8	4865	1314 (27%)	1193 (25%)	449 (9%)	389 (8%)
21	278	110 (40%)	117 (42%)	8 (3%)	64 (23%)
41	211	110 (52%)	127 (60%)	2 (1%)	95 (45%)

Table 22: Number of active voxels (and percentage) in different ROIs using the likelihood ratio test (LRT) to test both simultaneously, or the t-test to test individually. Also given is the result when corrected using the Benjamini-Hochberg procedure.

The most notable thing that can be seen from Table 22 is that for a subset of the regions of interest, there are far fewer activations detected in the second period of exercise in the experiment than the first period. This would imply that the response changes most significantly in the first period of exercise, and this type of behaviour will be seen again in later sections when using clustering or topological methods on the model residuals. Figure 36 shows the proportion of active voxels in all regions of interest.

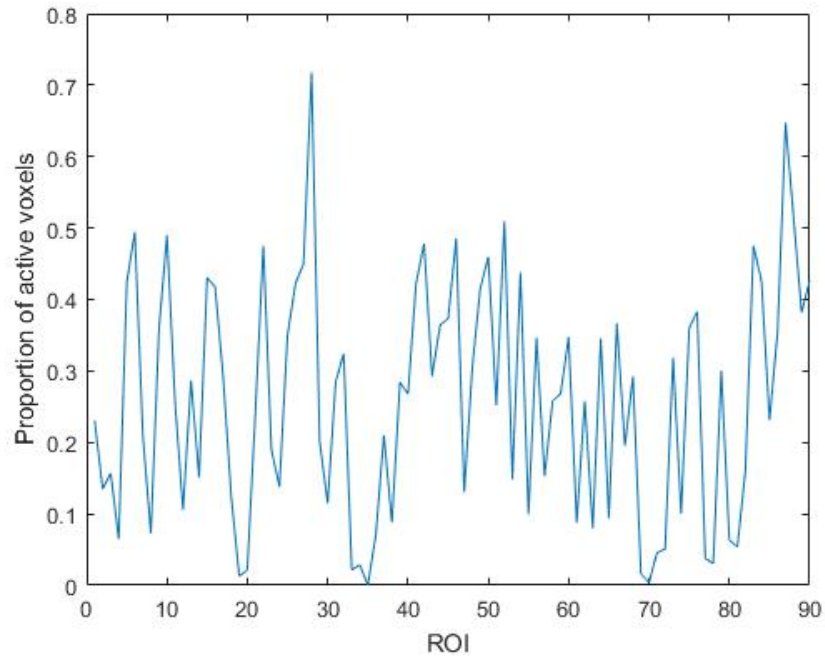


Figure 36: The proportion of active voxels (with correction) in all ROIs using temporal model

From Figure 36 we can see that the proportion of active voxels varies considerably across regions of interest, and Table 23 highlights some of the regions with the most and fewest activations.

ROI	Number Voxels	Percent Active
28	719	72
87	714	65
88	1090	51
52	2005	51
6	947	49
20	2176	2
69	1195	2
19	1953	1
70	714	0.4
35	472	0

Table 23: The lowest and highest percentage of active voxels in the ROIs using the temporal model and correction.

These regions of interest (see Table 63 in the Appendix) are a mixture of regions which are involved with movement and ones which are not. For example, the regions with the highest percentage of active voxels contain the Superior Frontal Gyrus (ROI 6) which makes up a substantial part of the frontal lobe which is involved in movement (see Section 2.2), however some of the other regions with a high percentage of active voxels are from the temporal lobe which is more associated with audio processing. The same can be said for the regions with the least percentage of active voxels. In the next chapters, we will move on to using the model residuals to investigate if any interesting features can be found using clustering then later topological methods.



## 5 Clustering

### 5.1 Clustering Of The Residual Regional Means

The model residuals are spatially correlated, as seen in the spatial modelling, and by calculating the correlation between pairs of voxel residuals in the previous section. In this section we will focus on investigating this spatial correlation further by using clustering methods to assess if there are any interesting spatial patterns in the residuals.

We begin by looking at the clustering of the regional residual means given in Equation (32). That is, for each region of interest, the mean of the residuals in that region will be used as being representative of that region. Therefore we will be clustering ninety time series, with the aim to see if there is any significant similarity or difference between regions.

#### 5.1.1 The Number Of Clusters

The most simple method of clustering is to use  $k$ -means clustering. To use  $k$ -means clustering, the number of clusters must be specified. There are many different methods to choose the number of clusters. The Calinski-Harabasz (CH) criterion [59] calculates a criterion score based on the distances between clusters,  $B(k)$ , and distances between objects within clusters,  $W(k)$ , and is given by

$$CH(k) = \frac{(N - k)B(k)}{(k - 1)W(k)} \quad (38)$$

where  $k > 1$  is the number of clusters.

The idea here is that a good clustering will have small within cluster distances and large between cluster distances (tight compact clusters that are well separated). The criterion is usually evaluated for a set of different values of  $k$ , and the one with the maximum value is the best, that is the one with the best ratio of between and within cluster distances. Figure 37 shows the Calinski-Harabasz criterion for the regional means. From this we see that the optimal number of clusters in this case is two.

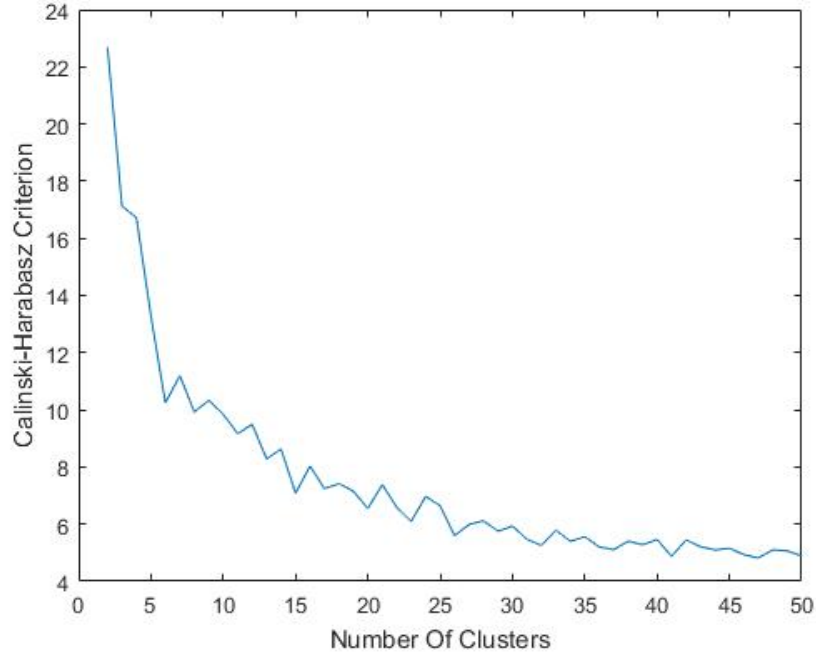


Figure 37: The Calinski-Harabasz criterion for different numbers of clusters using regional means.

The Calinski-Harabasz criterion here found that the optimal number of clusters is two (see Figure 37). However note that this criterion is not defined for one cluster. An alternative metric for determining the optimal number of clusters is the gap statistic [70] given by

$$G(k) = E[\log(W_k)] - \log(W_k) \quad (39)$$

where  $W_k$  is the average within cluster sum of squares about the cluster means defined by

$$W_k = \sum_{r=1}^k \frac{1}{2n_r} D_r$$

and  $D_r$  is the sum of all pairwise squared distances for all points in cluster  $r$  and  $n_r$  is the number of points in cluster  $r$ . The idea here is to choose the number of clusters which maximises the gap statistic. This occurs when the difference between the observed within region variability and the expected

within region variability (under the null distribution assumption that there is no clustering) is greatest [70]. Figure 38 shows the gap statistic for a set of different numbers of clusters applied to the regional residual means.

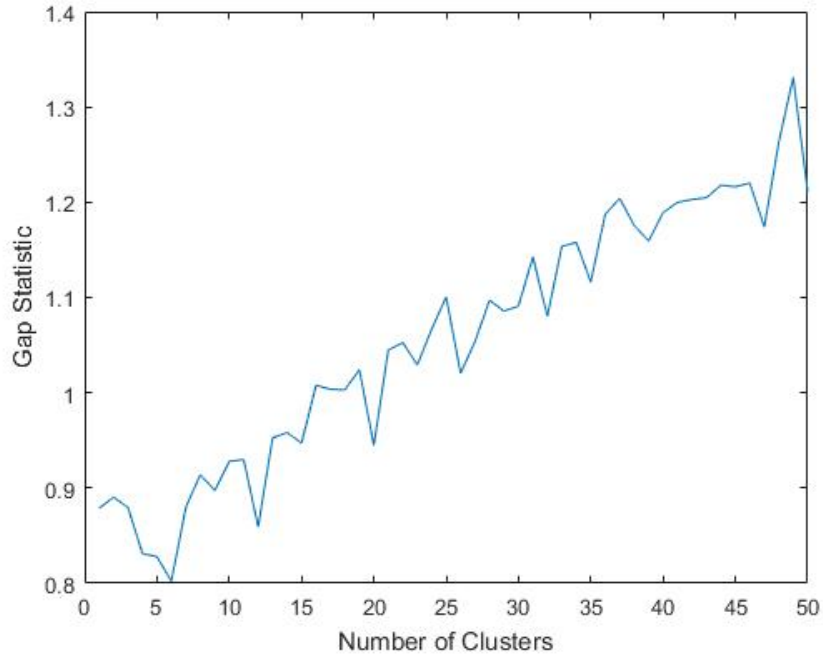


Figure 38: The gap statistic for different numbers of clusters

Although there is some suggestion from the Calinski-Harabasz criterion that two clusters is best, the gap statistic (Figure 38) suggests a very high number of clusters, here 49. Note that there are only 90 vectors being clustered, thus the gap statistic suggests very little evidence of clustering in the data. Therefore the Calinski-Harabasz criterion suggests very few clusters whilst the gap statistic suggests the opposite. Choice of number of clusters is clearly problematic for noisy time series data of this type. However, imposing two clusters did appear to give some interesting results (Figure 40), which will be further investigated.

### 5.1.2 Clustering Results

Using two clusters as suggested by the Calinski Harabasz criteria, Figure 39 shows a plot of the the residual times series' in the two clusters.

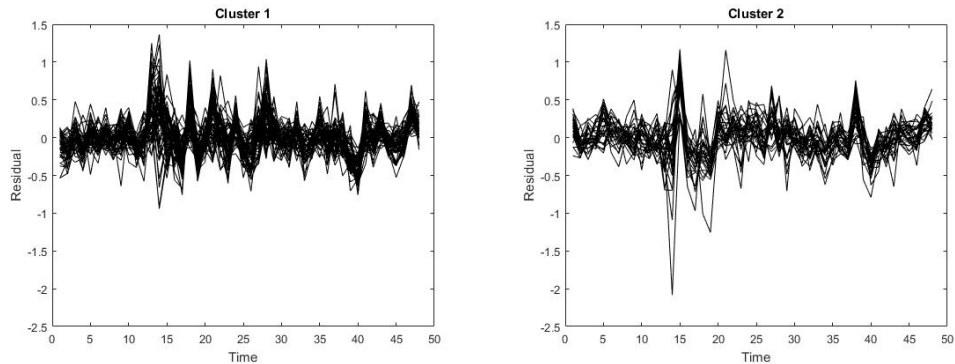


Figure 39: The two clusters of regional residual means

From Figure 39 we can see that the residuals are noisy in both cases. There are 60 of the residual regional means in the first cluster, and the remaining 30 in the second cluster. It does appear from the clustering, that the first cluster is much more noisy than the second cluster. Figure 40 shows the regional centroids coloured according to the clustering. Note that in this figure the points show the locations in the brain where the centroids are located.

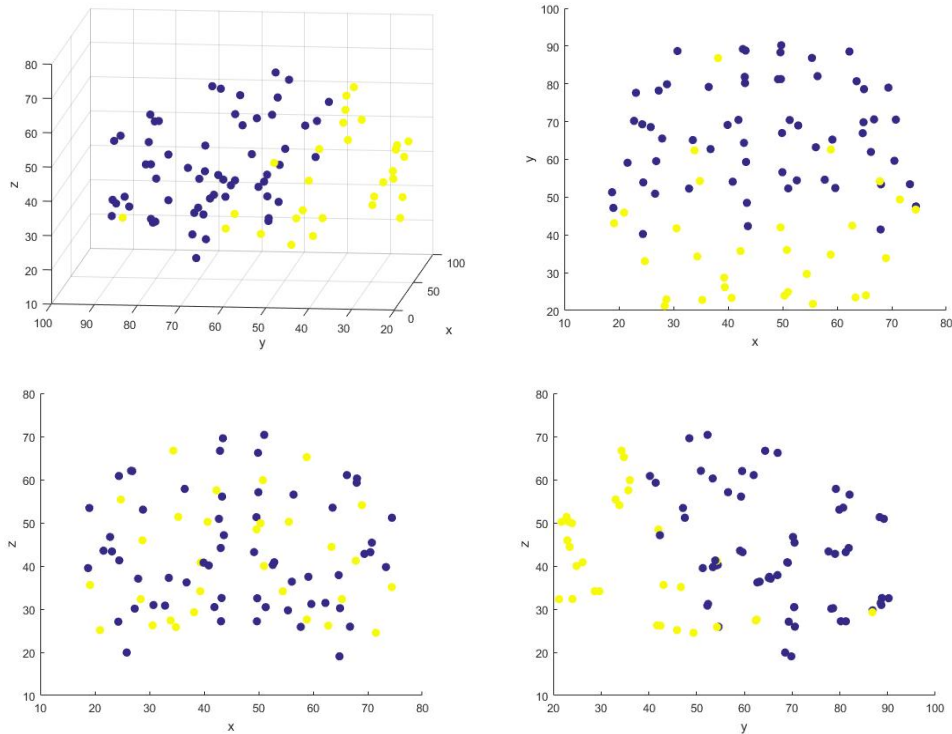


Figure 40: The regional centroids coloured according to the clustering of the regional residual means. Top left shows a 3D view of the clustering. Top right is 2D view in the  $xy$  plane, bottom left in the  $xz$  plane and bottom right  $yz$  plane.

From Figure 40 we can see that there is evidence of spatial clustering, that is that the residual means which are near each other seem to be assigned to the same cluster, and there seems to be a definite partition between the two clusters, certainly in three of the four plots.

### 5.1.3 Within Cluster And Between Cluster Correlation

To investigate the clustering, an interesting thing to check is the within cluster and between cluster voxel correlations. For each of the two clusters, the correlation between all pairs of regional means within each cluster was calculated, and then the correlation between all pairs between the two clusters. Let  $\mathcal{C}_k$  be the set of all regional means in cluster group  $k$  and let  $\delta_k$  be the set of all pairwise correlations between regional means in  $\mathcal{C}_k$ . Then

$$\bar{\delta}_k = \frac{\sum_{i,j \in \mathcal{C}_k} \text{corr}(\bar{h}_i, \bar{h}_j)}{\binom{|\mathcal{C}_k|}{2}} \quad (40)$$

is the mean of all pairwise correlations within cluster  $k$  and

$$SD(\delta_k) = \sqrt{\frac{\sum_{i,j \in \mathcal{C}_k} (\text{corr}(\bar{h}_i, \bar{h}_j) - \bar{\delta}_k)^2}{\binom{|\mathcal{C}_k|}{2} - 1}} \quad (41)$$

is the standard deviation of all pairwise correlations in cluster  $k$ . We can also define analogous expression for between cluster correlations. Let  $\gamma_{kk'}$  be the set of all pairwise correlations between regional means in  $\mathcal{C}_k$  and  $\mathcal{C}_{k'}$ . Then

$$\bar{\gamma}_{kk'} = \frac{\sum_{i \in \mathcal{C}_k, j \in \mathcal{C}_{k'}} \text{corr}(\bar{h}_i, \bar{h}_j)}{|\mathcal{C}_k| |\mathcal{C}_{k'}|} \quad (42)$$

is the mean of all pairwise correlations between cluster  $k$  and cluster  $k'$  and

$$SD(\gamma_{kk'}) = \sqrt{\frac{\sum_{i \in \mathcal{C}_k, j \in \mathcal{C}_{k'}} (\text{corr}(\bar{h}_i, \bar{h}_j) - \bar{\gamma}_{kk'})^2}{|\mathcal{C}_k| |\mathcal{C}_{k'}| - 1}} \quad (43)$$

is the standard deviation. Table 24 summarises the results for the clustering of the regional means.

Cluster	1	2
1	0.44 (0.21)	0.23 (0.17)
2	0.23 (0.17)	0.47 (0.21)

Table 24: The mean within cluster correlations  $\delta_k$  (and standard deviation), and mean between cluster correlation  $\gamma_{12}$  (and standard deviation) for the two clusters of the regional residual means.

From Table 24 it can be seen that the within cluster correlations are higher than the between cluster correlations. However note that the standard deviation is also fairly high in all cases.

#### 5.1.4 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) provides a method of optimally separating classes of objects. Using the clustering results, LDA can be used to assess the separation of the two clusters.

The general idea of LDA is to fit a line or surface (depending on the dimension of the data) to separate classes of points, such that the differences between the within class means are maximised and the within scatter is minimised [63]. More formally, let  $X_1 = (x_1^1, x_2^1, \dots, x_{l_1}^1)$  and  $X_2 = (x_1^2, x_2^2, \dots, x_{l_2}^2)$  be samples from two different classes, where the complete sample is defined by  $X = (x_1^1, x_2^1, \dots, x_{l_1}^1, x_1^2, x_2^2, \dots, x_{l_2}^2)$ . Then the linear discriminant is defined by the vector  $\omega$  which maximises

$$J = \frac{\omega^T S_B \omega}{\omega^T S_W \omega} \quad (44)$$

where

$$\begin{aligned} S_B &= (m_1 - m_2)(m_1 - m_2)^T \\ S_W &= \sum_{i=1,2} \sum_{x \in X_i} (x - m_i)(x - m_i)^T \end{aligned} \quad (45)$$

are the between and within class scatter matrices respectively, and  $m_i$  is the within class mean for class  $i$ . LDA can be used as a classification tool, or as a method of dimension reduction. Once a discriminant is fit, a new point can be classified, or assigned to a cluster, using its dependent variables, here its three dimensional coordinates. With respect to dimension reduction, the original points in  $n$  dimension space can be projected onto an  $m$  dimension surface (where  $m < n$ ) which separates the classes optimally. This may be useful in order to visualise classes which cannot be seen by inspection in high dimensions. In this section LDA will be used to both classify and to reduce dimension in order generate plots of the separation of the clusters in two dimensions.

Leave one out cross validation (LOOCV) can be used with LDA to assess the quality of the separation of the two clusters. The method is to leave one of the observations out (test data), and use the rest of the data as a training set to fit the linear discriminant. The test data is then assigned to a cluster based on which side of the linear discriminant its coordinates lies on. This can be repeated for all points. As we know the true cluster which all points are actually assigned to, we can then get a measure of accuracy of the LDA classifier. If there is an obvious separation between the two clusters, then this accuracy is expected to be high.

Figure 41 shows the residual regional means split into two clusters (top

left). This three dimensional set of points in two classes is then projected onto the optimal two dimensional surface (top right and bottom centre).

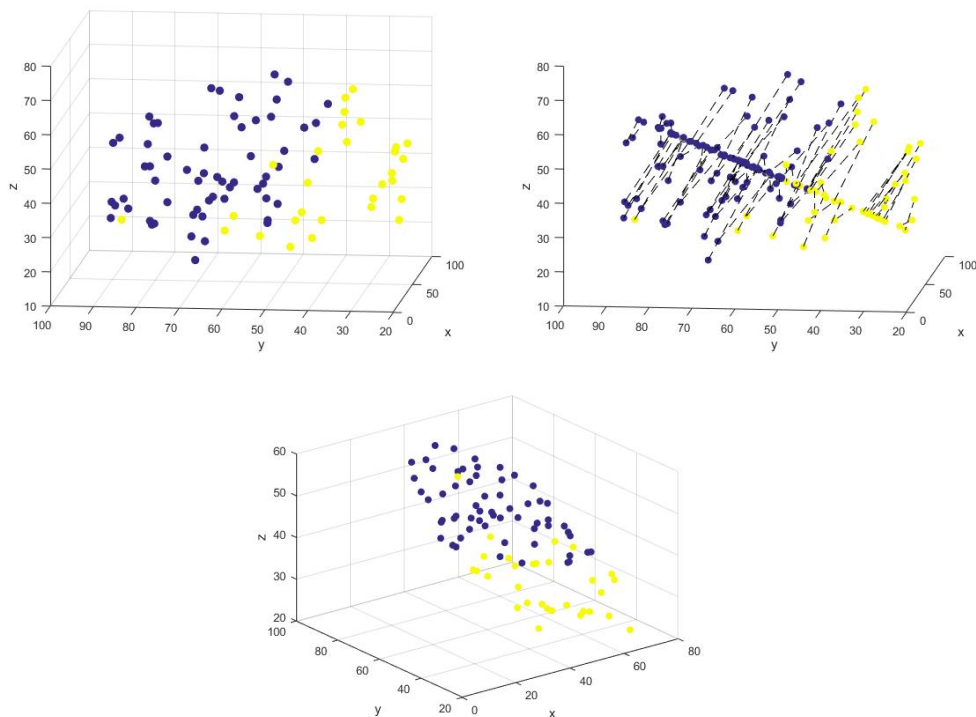


Figure 41: Clustering of regional residual means (top left), projection on to discriminant plane (top right) and the rotated discriminant plane (bottom centre).

Figure 41 shows that there is a clear separation between the two cluster groups. To measure the separation, LOOCV will now be used as described earlier. Table 25 shows the result from classifying each residual region mean, by using LDA to fit a discriminant to all other regional means, and assigning a class based on its location.



	Predicted Cluster	
Actual Cluster	1	2
1	55	5
2	5	25

Table 25: The number of residual regional means in each of the actual clusters, and the number in each of the clusters predicted using LOOCV.

Table 25 shows that out of the 60 residual region means in the first cluster, 55 are assigned to the first cluster using LOOCV and LDA. Similarly, out of the 30 residual regional means in the second cluster, 25 are assigned to the second cluster using LOOCV and LDA.

In order to summarise the results in Table 25, and in similar tables in later sections, three measures will be used; accuracy, sensitivity and specificity [64]. Consider the general results in Table 26.

	Predicted Cluster	
Actual Cluster	1	2
1	a	b
2	c	d

Table 26: The general number of regional means in each of the actual clusters, and the number in each of the clusters predicted using LOOCV.

First, let us define the accuracy of the classification as

$$Accuracy = \frac{a + d}{a + b + c + d}. \quad (46)$$

Hence the accuracy is the proportion of regional means which are correctly classified. Next define the sensitivity as the proportion of regional means in cluster one which are predicted to be in cluster one, ie

$$Sensitivity = \frac{a}{a + b}. \quad (47)$$

Finally define the specificity as the proportion of regional means in cluster two which are predicted to be in cluster two, ie

$$Specificity = \frac{d}{c + d}. \quad (48)$$

Applying these measure to the results in Table 25 results in an accuracy of 0.89, sensitivity of 0.92 and specificity of 0.83. It is important that multiple measures are used, as using accuracy alone can give misleading results (see Table 33 later). High results for all three measure gives evidence for strong separation between the two clusters.

## 5.2 Clustering Of Raw Data

### 5.2.1 Raw Regional Means

In the previous section we found that there appeared to be some spatial clustering of the residuals. In this section we will investigate if this pattern can be found by clustering various forms of the original raw data, rather than the residuals.

Figure 42 shows the clustering of the raw regional mean vectors  $\bar{y}_i(t)$  defined by

$$\bar{y}_i(t) = \sum_{v \in \mathcal{R}_i} \frac{y_v(t)}{|\mathcal{R}_i|}. \quad (49)$$

Note that two clusters have been chosen to compare the results with those obtained when clustering the residuals. Applying the C-H criterion to the raw mean vectors, the optimal number of clusters was 34. The actual clustering on the brain looks fairly similar to that seen when clustering the residual mean vectors. In Figure 42 the clustering appears to be due to the magnitude of the signal. This cannot be the case for the clustering of the residual regional means, as all residuals have mean zero.

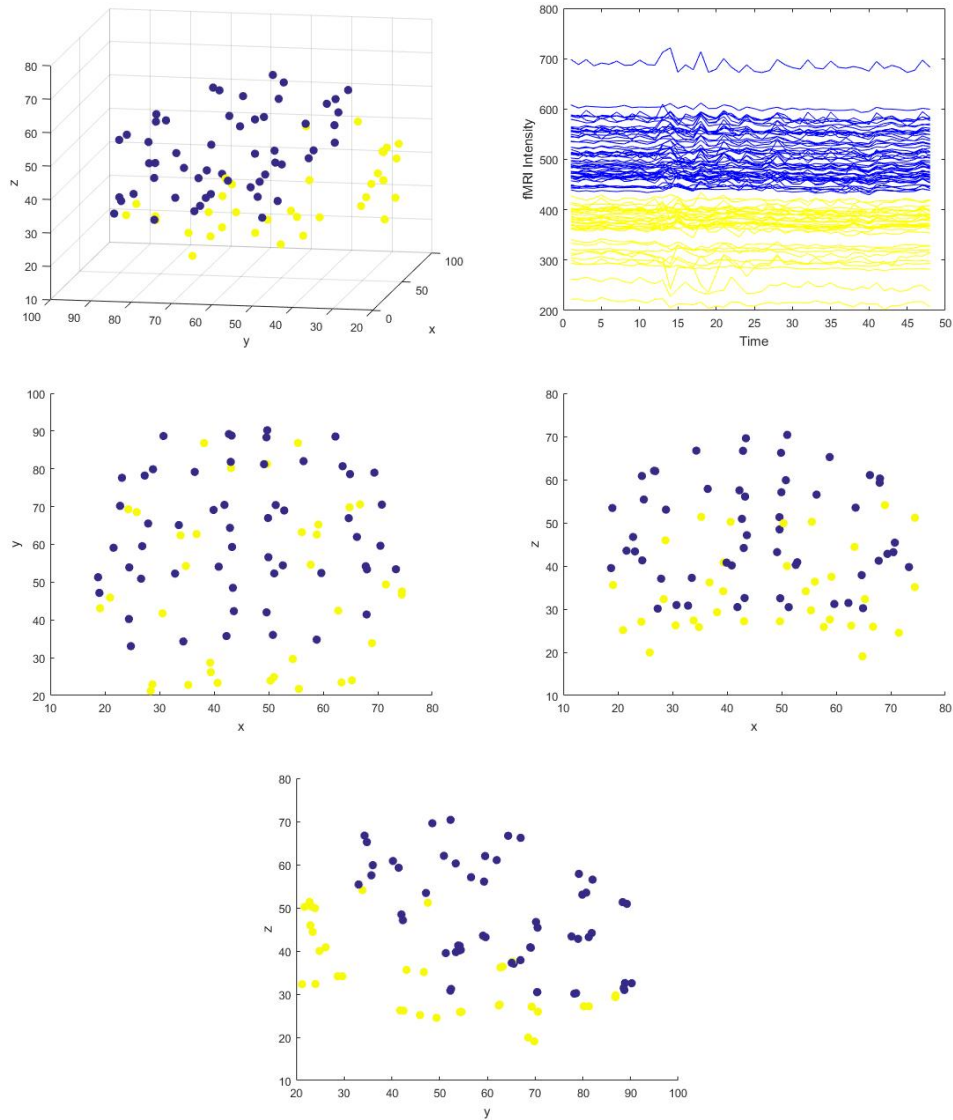


Figure 42: The two clusters of raw data regional means on the brain (top left), the raw data (top right) and the clusters in the x-y, x-z and y-z 2D planes.

Figure 43 shows the variance of each of the residual and raw regional mean vectors. Note that in the modelling the residuals are scaled such that they have unit variance. Investigating the model residuals it was found that

the variance of the residuals  $\mathbf{h}$  at each voxel was approximately equal to one as expected. However note that the regional means of the residuals  $\bar{h}_i$  will not have unit variance as seen in Figure 43.

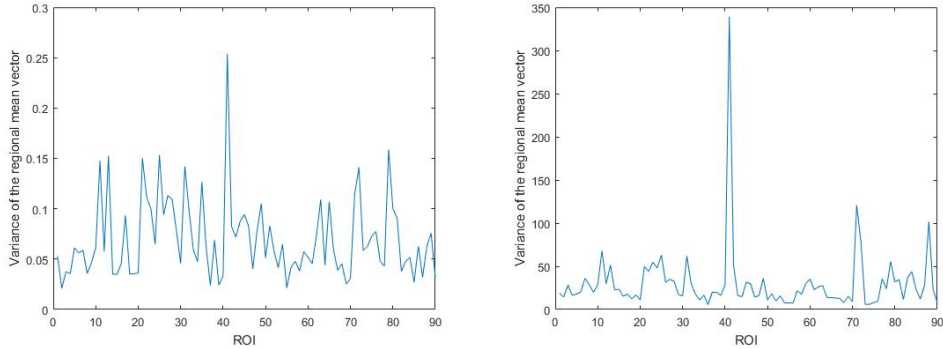


Figure 43: The variance of the residual regional means (left) and the raw regional means (right). Note the change of scale.

From Figure 43 we can see that there appears to be some evidence of correlation between the variance of the raw regional mean vectors and the variance of the residual regional mean vectors. In fact the correlation is 0.68 which may explain the similarity in the clustering of the two data sets.

## 5.2.2 Raw Centered Regional Means

Let  $\bar{r}_i(t)$  denote the centered regional means of the raw data defined by

$$\bar{r}_i(t) = \sum_{v \in \mathcal{R}} \frac{y_v(t) - \bar{y}_v}{|\mathcal{R}_i|} \quad (50)$$

where  $\bar{y}_v$  is the mean of the time series at voxel  $v$ . Figure 44 shows the result of clustering the raw data with the mean removed, and although we have specified two clusters here, the separation between the clusters is much less evident than the previous cases.

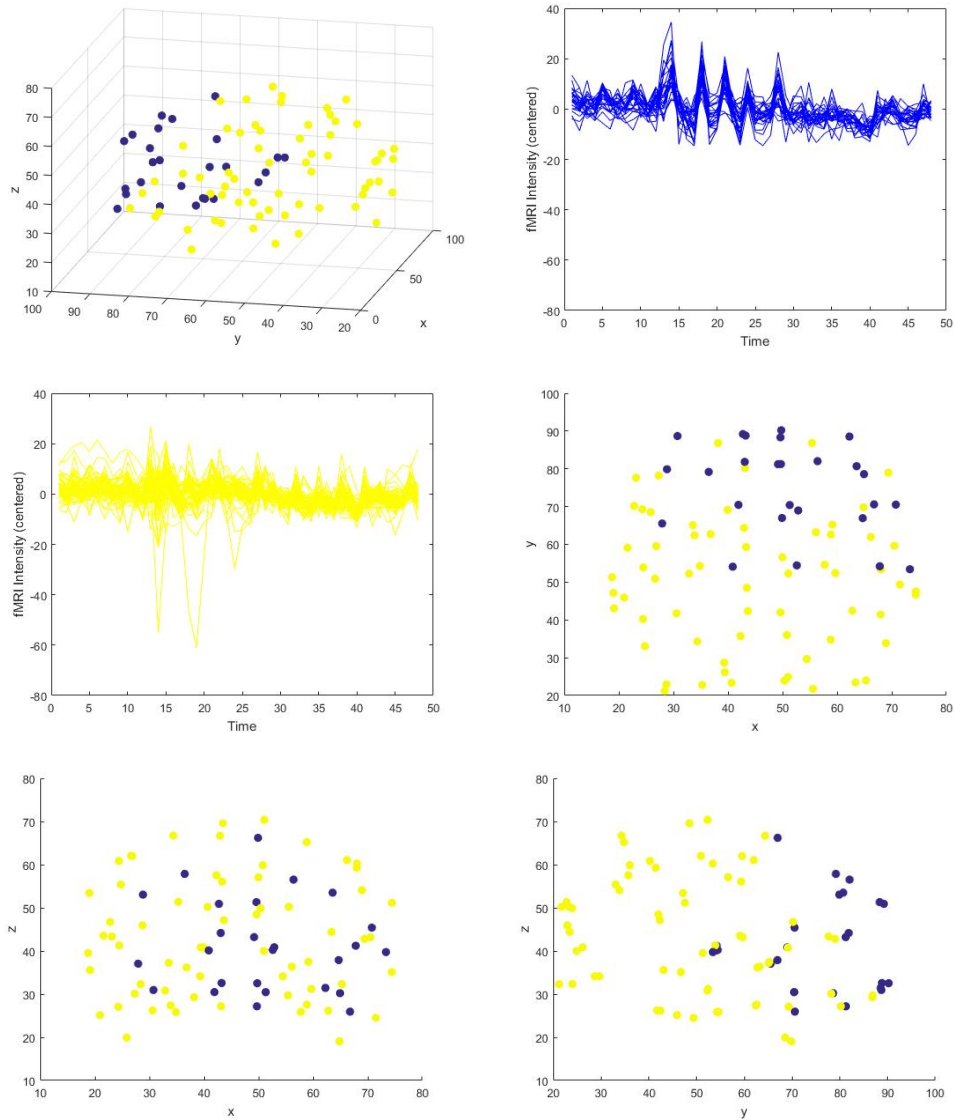


Figure 44: The two clusters of raw data means (centered) on the brain, the raw data and views from the x-y, x-z and y-z planes.

In Figure 44, the centered raw regional means were clustered with two clusters chosen. Using the C-H criterion the optimal number here is five. Figure 45 shows the CH criterion in this case. From this we can see only a marginal improvement of having five clusters over two, and further inves-

tigation found that 80 of the 90 mean vectors were in two of the clusters, therefore it would be preferable in this case to opt for two clusters.

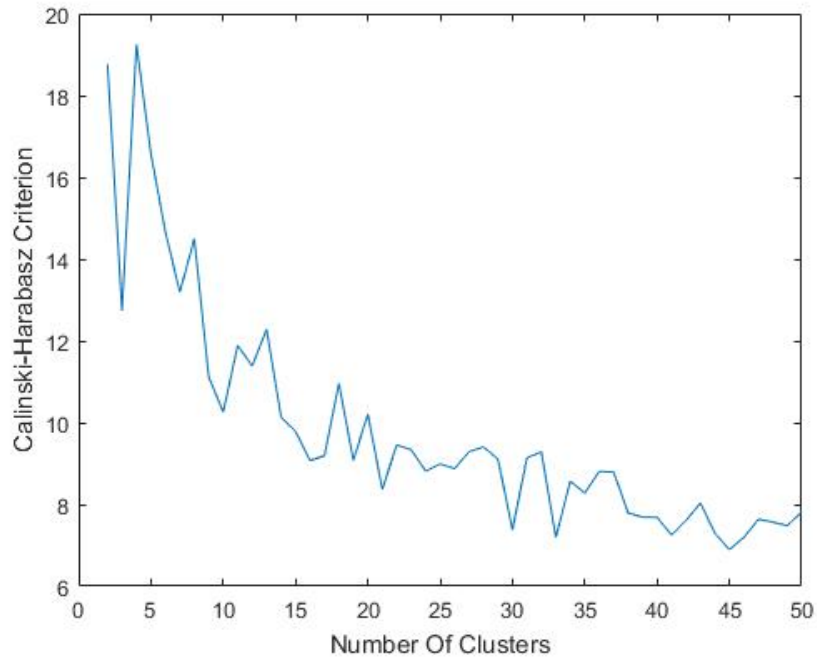


Figure 45: The CH criterion for the clustering of the centered raw regional means.

### 5.2.3 Clustering Using Number Of Voxels In Each Region

The next clustering to be considered will be the clustering of the regions of interest based on the number of voxels. Initially, the C-H criterion was used to calculate the optimal number of clusters. The result was 49 clusters (see Figure 46), many of which had a single element. Upon further investigation it was found that although region sizes vary from 211 to 4865 voxels, with median 1364, regions which had a difference of only a small number of voxels were put into different clusters. For example region two and three which have 3224 and 3262 voxels respectively were in the same cluster, though region one which has 3178 was put in a different cluster. Therefore, it appears that the C-H criterion is too sensitive in this instance. An alternative is to assume two clusters as previously.

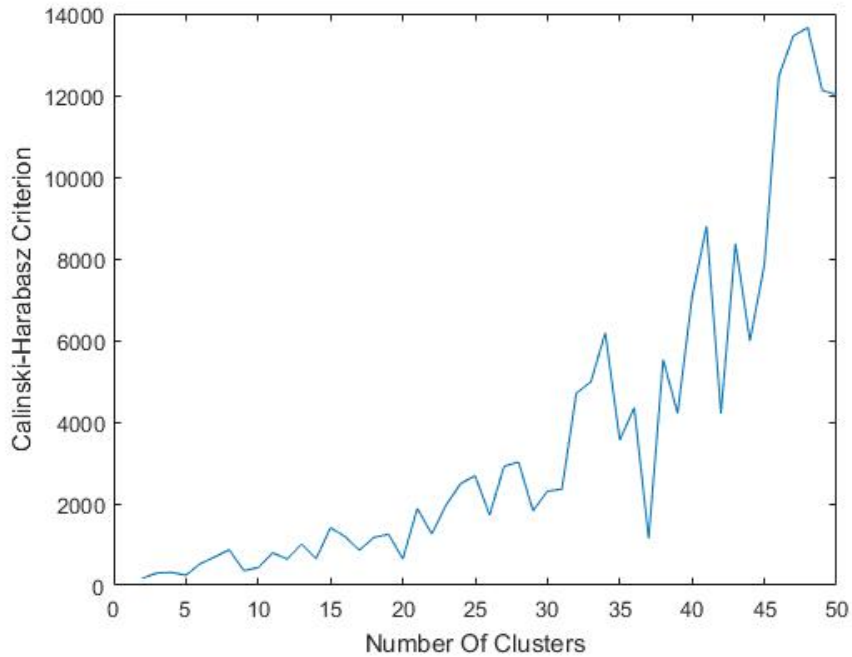


Figure 46: The Calinski-Harabasz criteria applied to the number of voxels in each region.

Figure 47 shows the clustering of the region sizes using two clusters. Here the clusters are of sizes 72 and 18. From the figure we can see that there is no real spatial pattern in the clustering. However the top right plot shows that the clustering is based on region size as expected.

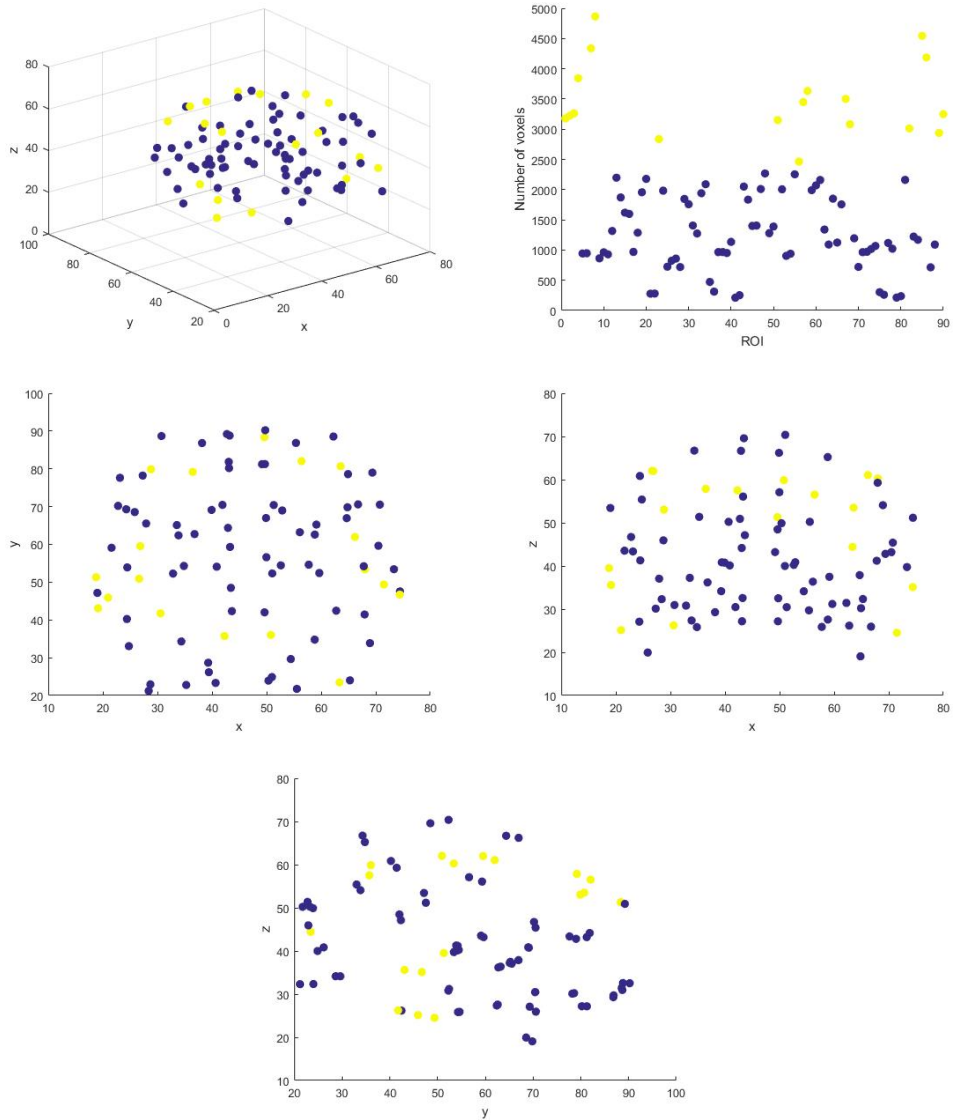


Figure 47: Clustering of number of voxels in each region using two clusters and the  $k$ -clustering algorithm. Top left shows the two clusters on the brain in 3D. Top right shows the number of voxels per region coloured according to cluster. The bottom three are the clusters on the brain in the x-y, x-z and y-z 2D planes.



#### 5.2.4 Clustering Using Mean (Over Time) Of Regional Means

The next clustering to be considered will be clustering based on the mean (over time) of each of the raw regional mean vectors. Let  $\bar{r}_i$  be the scalar mean over time of the regional mean for region  $i$  defined by

$$\bar{r}_i = \sum_{t=1}^{48} \frac{\bar{r}_i(t)}{48}. \quad (51)$$

Hence again we are clustering 90 scalars. If we allow the C-H criterion to choose the number of clusters, as before we obtain a large number of clusters, here 50. suggesting no real clustering. Figure 48 shows the results from forcing two clusters.

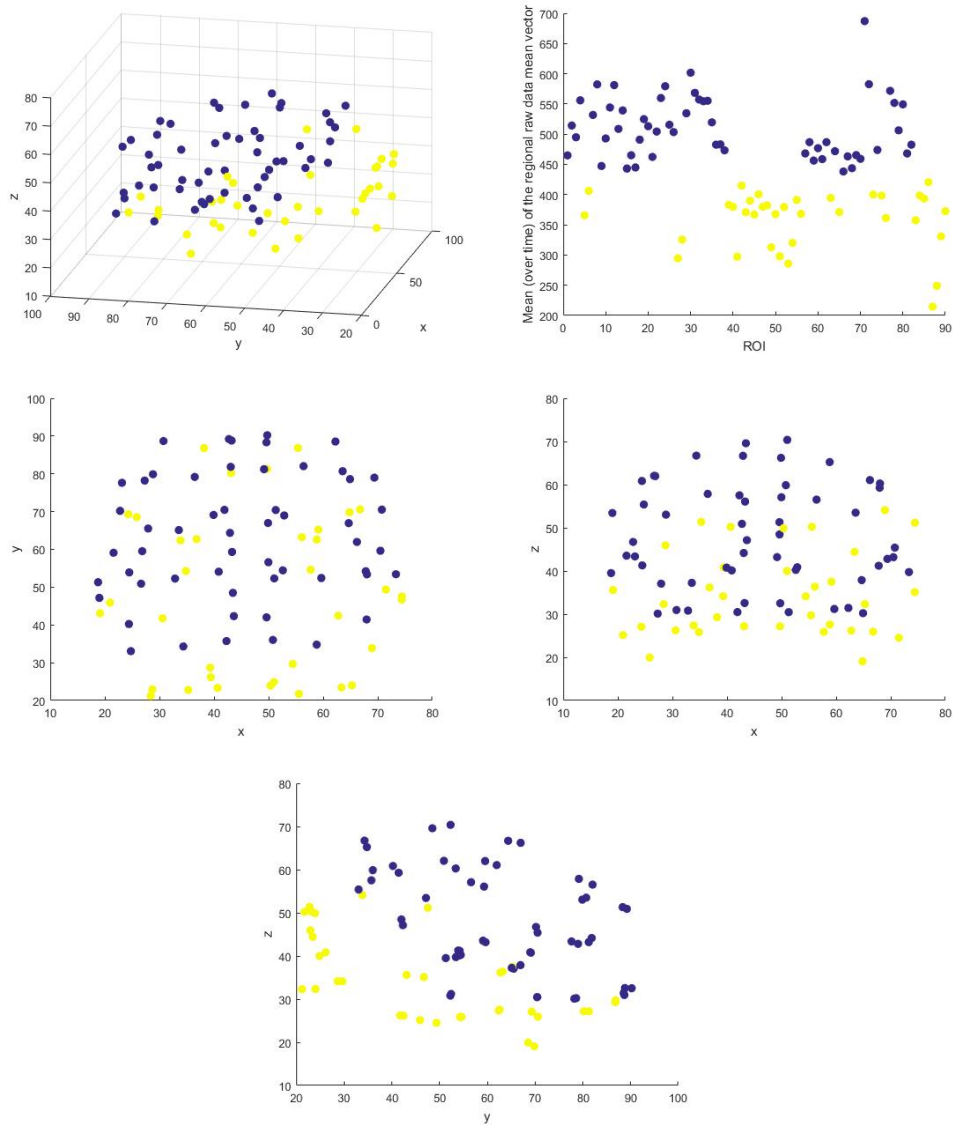


Figure 48: Clustering of the mean (over time) of the raw regional mean vectors. Top left shows the two clusters on the brain in 3D. Top right shows the scalar mean for each region coloured according to cluster. The bottom three are the clusters on the brain in the x-y, x-z and y-z 2D planes.

### 5.2.5 Clustering Using Standard Deviation Of Regional Means

The next clustering to be considered will be clustering based on the standard deviation of each of the 90 raw regional mean vectors defined by

$$SD(\bar{r}_i(t)) = \sqrt{\frac{\sum_{t=1}^{48} (\bar{r}_i(t) - \bar{r}_i)^2}{47}}. \quad (52)$$

Hence again we are clustering 90 scalars. If we allow the C-H criterion to choose the number of clusters, like before we obtain a large number of clusters, here 47, suggesting no real clustering. Figure 49 shows the results from forcing two clusters.

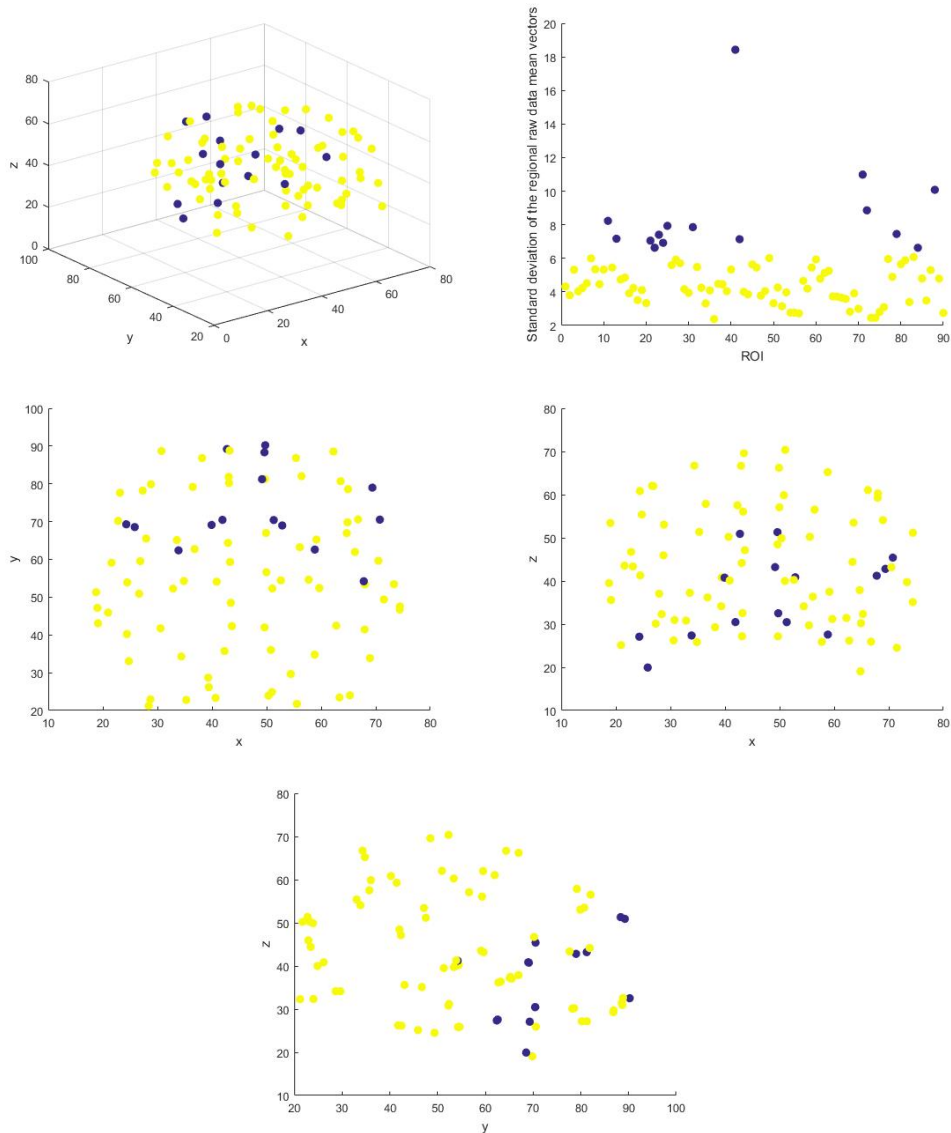


Figure 49: Clustering of the standard deviation of the raw regional mean vectors. Top left shows the two clusters on the brain in 3D. Top right shows the standard deviation for each region coloured according to cluster. The bottom three are the clusters on the brain in the x-y, x-z and y-z 2D planes.

From Figure 49 we can see that the clustering does separate out regions with high and low standard deviations (see top right) as expected, though

spatially there is no clear pattern.

### **5.3 Clustering Using Brain Anatomy**

In Figure 40, we can see that there is a clear separation of two areas of the brain when clustering the model residuals. In this section we will investigate if there are any relationships between the brain anatomy (see Section 2.2) and the clustering results.

Figure 50 shows the brain regional centroids divided into two to show those which appear on the left or right side of the brain, and Figure 51 shows the different regions divided into the four lobes of the forebrain.

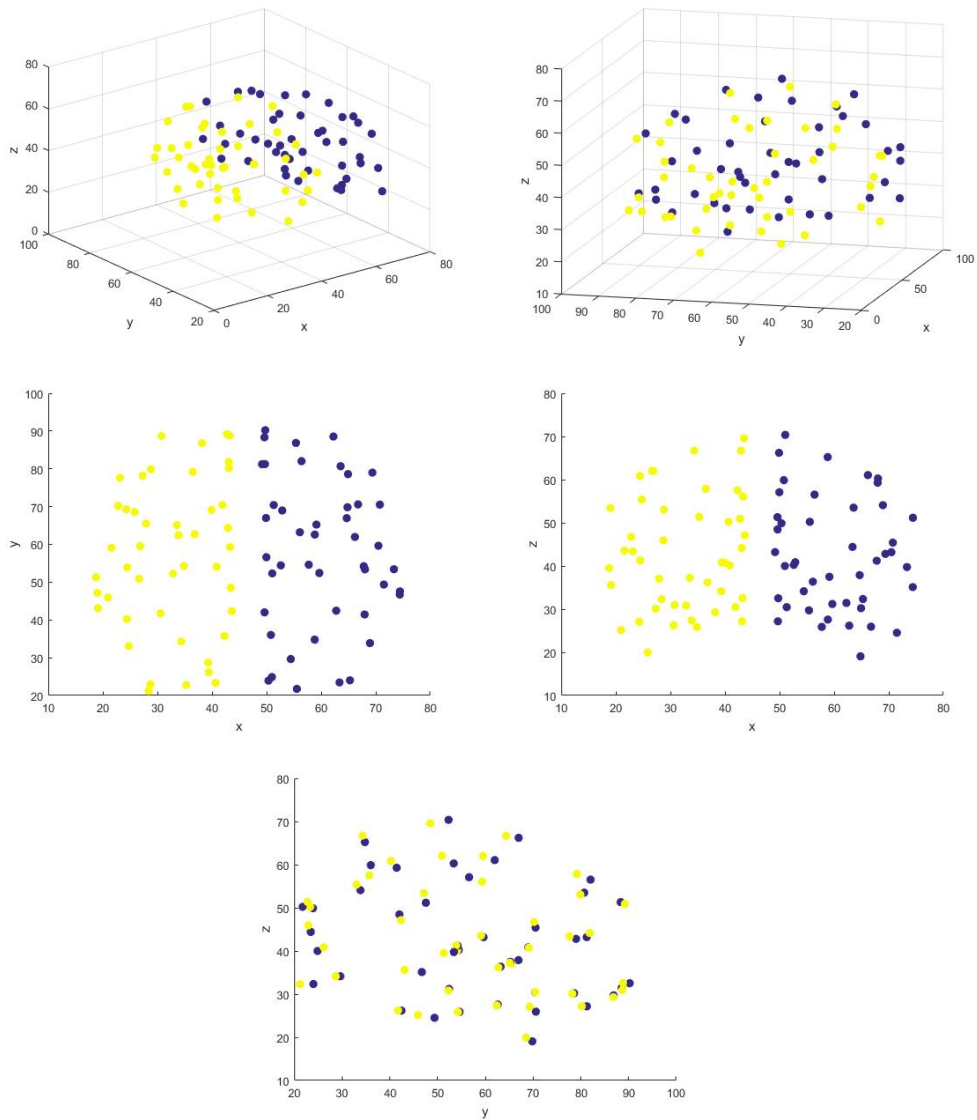


Figure 50: The regions of interest separated into left (yellow) and right (blue) sides of the brain. Top left figure shows the division, and the top right is the same plot orientated for comparison with other clustering plots. The remaining plots show the view from the various two dimensional planes.

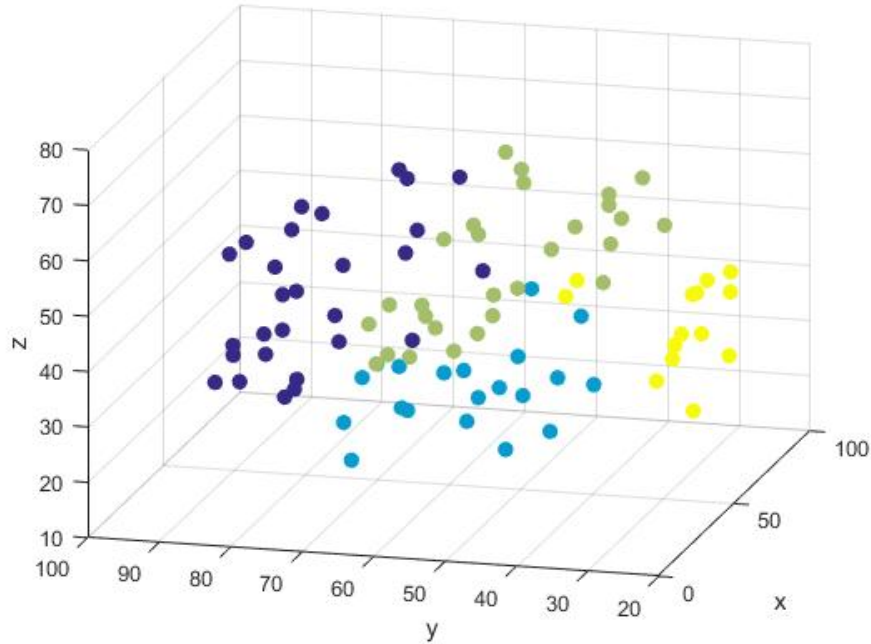


Figure 51: The regions of interest coloured according to the lobes of the forebrain (blue is frontal lobe, green is parietal lobe, yellow is occipital and light blue is temporal lobe).

Comparing Figure 51 with the clustering of the model residuals (Figure 40) it appears that the clustering could potentially be separating the temporal and occipital lobe from the frontal and parietal lobe. From Figure 50 it can be seen that the clustering is not based only upon the left and right side of the brain.

## 5.4 Comparing The Clusterings

### 5.4.1 Rand Index

One method of comparing the clusterings is to use the Rand Index. The Rand index is given by

$$RI = \frac{a + b}{a + b + c + d} \quad (53)$$

where  $a$  is now the number of pairs of points which are in the same cluster in both clusterings,  $b$  is the number of pairs which are in different clusters in both clusterings,  $c$  is the number of pairs which are clustered together in the first clustering, but clustered differently in the second cluster and  $d$  is the number of pairs which are clustered differently in the first clustering, but clustered together in the second cluster. The Rand Index takes a value between zero and one, with values closer to one implying a closer relationship between two clustering results. Table 27 shows the Rand index between the different clusterings that have been performed in this section.

	Number Voxels	$\bar{r}_i$	$SD(\bar{r}_i(t))$	$\bar{y}_i(t)$	$\bar{r}_i(t)$	$\bar{h}_i(t)$
Number Voxels	1	0.50	0.54	0.50	0.52	0.54
$\bar{r}_i$	0.50	1	0.50	1	0.53	0.66
$SD(\bar{r}_i(t))$	0.54	0.50	1	0.50	0.62	0.50
$\bar{y}_i(t)$	0.50	1	0.50	1	0.53	0.66
$\bar{r}_i(t)$	0.52	0.53	0.62	0.53	1	0.51
$\bar{h}_i(t)$	0.54	0.66	0.50	0.66	0.51	1

Table 27: The Rand Index comparing various clusterings.

From Table 27 it can be seen that the strongest relationship between clusterings is the comparison of the clustering of the raw regional means,  $\bar{y}_i(t)$ , and the residual regional means,  $\bar{h}_i(t)$ . This reflects the similarity seen in Figures 40 and 42.

A disadvantage of the Rand Index is that it does not take into consideration the possibility of elements being clustered together by chance, and also that it counts pairs which are grouped differently in both clusterings as evidence of similarity.

The following is an example which highlights these problems. Here two sets of independent standard normal data are generated, each set is of dimension 48 by 90 to mimic 90 regional mean vectors of length 48. Clustering analysis is carried out on each of these sets, and the Rand Index is calculated to compare the resulting cluster labels. Figure 52 shows the result of 1000 simulations.



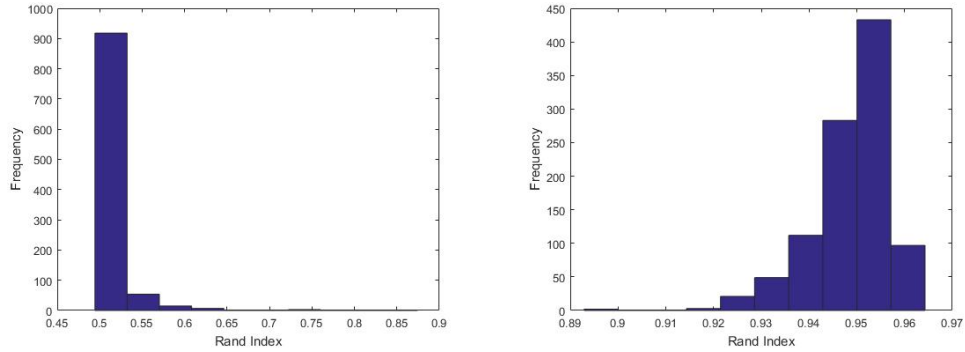


Figure 52: The Rand Index for 1000 simulations of sets of independent data and imposing two clusters (left) and 45 clusters (right) .

From Figure 52 we can see that when imposing two clusters, the Rand Index here has mean of 0.51 and standard deviation 0.03, yet when imposing a large number of clusters (here 45), the Rand Index has mean 0.95 and standard deviation 0.01. Although the Rand Index is one for identical clusterings we can see that for comparison of random clusterings the index is near 0.5 for two clusters. This is because of the random chance that pairs of points could be clustered together. From Figure 52 we can see that if we increase the number of clusters the similarity between the clusterings of random independent data is near one. This is because the value  $b$  in Equation (53), which counts pairs which are in different clusters in both sets of clusterings is used as evidence for similarity. In this example of high number of clusters, the cluster groups will be very small, and so it is highly likely for pairs of points which are not clustered together in the first clustering to not be grouped together in the second clustering, ie  $b$  will be large. Hence the Rand Index will say the two clusterings are very similar, even when the two sets of data are completely random and independent.

#### 5.4.2 Adjusted Rand Index

An alternative to the Rand Index which accounts for this is the Adjusted Rand Index (ARI) [52]. The ARI takes into consideration the expected value of the index under the assumption of randomness. The general form of the index [80] is

$$ARI = \frac{Index - ExpectedValue}{MaximumIndex - ExpectedValue}. \quad (54)$$

Therefore the ARI is of value zero if the expected value equals the index value, ie the ARI is zero if the clusterings we are comparing are random and independent of each other. Like the Rand Index, the maximum value for the ARI is one for when comparing identical clusterings, however the ARI can also take negative values if the expected value under randomness is greater than the index. The ARI assumes the generalised hypergeometric distribution as the model of randomness [52], and it can be shown that under this assumption the ARI can be expressed as

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}}{\frac{1}{2} [\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}] / \binom{n}{2}} \quad (55)$$

To explain this formula let  $X$  and  $Y$  be two clusterings of the same  $n$  data points into  $s$  and  $r$  cluster groups respectively. The  $n_{ij}$ ,  $a_i$  and  $b_j$  are from the contingency table

	$Y_1$	$Y_2$	$\dots$	$Y_s$	Sums
$X_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1s}$	$a_1$
$X_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2s}$	$a_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_r$	$n_{r1}$	$n_{r2}$	$\dots$	$n_{rs}$	$a_r$
Sums	$b_1$	$b_2$	$\dots$	$b_s$	

Table 28: The Adjusted Rand Index comparing various clusterings

Here  $n_{ij}$  is the number of occurrences that an element in clustering  $X$  is in group  $i$  and is in group  $j$  in clustering  $Y$ . Hence instead of counting the numbers of pairs together or not across two clusterings, the ARI counts the number of occurrences where an element (here a regional mean) has the same cluster label in both clusters, and then corrects for the possibility that this may happen by chance. Furthermore [69] gives a result for the variance of the ARI index as

$$Var(ARI) = \frac{\binom{n}{2}^2 \alpha}{\left(\binom{n}{2}^2 - [(a+c)(a+d) + (b+c)(b+d)]\right)^2} \quad (56)$$

where  $a, b, c, d$  are as defined in (53), and

$$\alpha = \frac{1}{16} \left( 2n(n-1) - \left( \frac{ef}{n(n-1)} \right)^2 + \frac{4(g-h)(m-h)}{n(n-1)(n-2)} + \frac{(e^2 - 4g + 2h)(f^2 - 4m + 2h)}{n(n-1)(n-2)(n-3)} \right) \quad (57)$$

where

$$\begin{aligned} e &= 2 \sum_{i=1}^r a_i^2 - n(n+1) \\ f &= 2 \sum_{j=1}^s b_j^2 - n(n+1) \\ g &= 4 \sum_{i=1}^r a_i^3 - 4(n+1) \sum_{i=1}^r a_i^2 + n(n+1)^2 \\ h &= n(n-1) \\ m &= 4 \sum_{j=1}^s b_j^3 - 4(n+1) \sum_{j=1}^s b_j^2 + n(n+1)^2. \end{aligned} \quad (58)$$

Table 29 shows the ARI between the different clusterings that have been performed in this section. The standard deviation for all comparisons as calculated using the above formula is 0.2 to one decimal place.

	Number	$\bar{r}_i$	$SD(\bar{r}_i(t))$	$\bar{y}_i(t)$	$\bar{r}_i(t)$	$\bar{h}_i(t)$
Number	1	-0.02	-0.08	-0.02	-0.01	0.04
$\bar{r}_i$	-0.02	1	-0.02	1	0.05	0.32
$SD(\bar{r}_i(t))$	-0.08	-0.02	1	-0.02	0.17	-0.03
$\bar{y}_i(t)$	-0.02	1	-0.02	1	0.05	0.32
$\bar{r}_i(t)$	-0.01	0.05	0.17	0.05	1	0.01
$\bar{h}_i(t)$	0.04	0.32	-0.03	0.32	0.01	1

Table 29: The Adjusted Rand Index comparing various clusterings of the data.

Table 30 shows the ARI between the various clusterings and some clusterings based upon the brain. Here left/right is a two cluster separating regions in the left and right side of the brain, and lobes is a two cluster with the temporal and occipital lobe in one region and the parietal and frontal lobe in another.

	Left/Right	Lobes
Number	-0.001 (0.20)	-0.004 (0.20)
$\bar{r}_i$	-0.01 (0.22)	0.38 (0.22)
$SD(\bar{r}_i(t))$	-0.002 (0.20)	-0.001 (0.20)
$\bar{y}_i(t)$	-0.01 (0.22)	0.38 (0.22)
$\bar{r}_i(t)$	0.02 (0.21)	0.01 (0.21)
$h_i(t)$	-0.01 (0.22)	0.30 (0.21)
Left/Right	1 (0.23)	-0.01 (0.22)
Lobes	-0.01 (0.22)	1 (0.22)

Table 30: The Adjusted Rand Index (and standard deviation) comparing various clusterings of the data with the brain anatomy clusterings.

From Table 29 it can be seen that the ARI is very close to zero when comparing most of the clusterings, indicating no evidence of a relationship between the clusterings. However, the raw regional means (uncentered) and the residual clusterings have an ARI of 0.32 with standard deviation 0.21, indicating some relationship between the clusterings. This reflects what has been seen in Figures 40 and 42. From Table 30 we can see that the ARI between the residual mean clustering and the partition of the brain into the left and right hand side is approximately zero, indicating an almost no relationship between the clusterings. However the ARI between the residual mean clustering and the two lobes clustering is 0.30, which is evidence of a relationship.

The ARI fixes the issue that the Rand Index had in terms of taking into consideration the possibility of pairs being grouped together by chance. As in Section 5.4.1, Figure 53 shows results from 1000 simulations of sets of random independent normal data, clustered using two or 45 clusters, and using the ARI to compare clusterings each time.

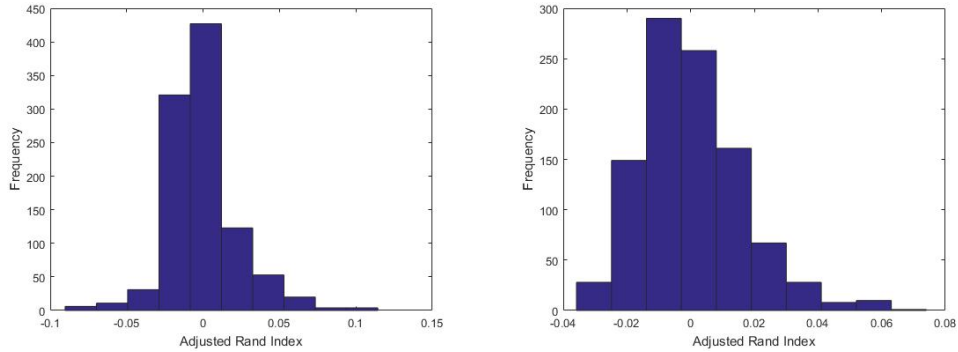


Figure 53: The Adjusted Rand Index (right) for 1000 simulations of sets of independent data and imposing two clusters (left) or 45 clusters (right).

Figure 53 shows that in both instances (small and large number of clusters), the ARI is near zero. Here the mean and standard deviation when using two clusters is 0 and 0.02 respectively, and for 45 clusters the mean and standard deviations is the same. Hence using the ARI we can be confident that a result near zero indicates that the clusterings are independent.

### 5.4.3 Association Test

Table 31 gives a more intuitive way of comparing the clusterings.

	Residuals Cluster 1	Residual Cluster 2	Total
F/P	48	8	56
T/O	12	22	34
Total	60	30	90

Table 31: Counts of the number of regions in the F/P (frontal/parietal lobes) or T/O (temporal/occipital lobes) that are in cluster one or two of the clustering of the residual regional means. Association p-value: 0.000001.

From Table 31 it can be seen that 48 of the 60 regions (80%) of the regions in the first cluster are parts of the frontal or parietal lobes, and 22 of the 30 (73%) regions in the second cluster are parts of the temporal or occipital lobes. This can be seen in Figure 51. Using this table, we can also test for association between the lobe groups and the residual clusters using the  $\chi^2$  test. Given two categorical variables, one with  $r$  levels and the other with

$c$  levels we can perform the hypothesis test with null  $H_0$  that there is no association between the two variables, against the alternate  $H_1$  that there is association, using the test statistic

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \quad (59)$$

where  $O_{i,j}$  is the number of observed samples which are of level  $i$  of the first category, and level  $j$  of the second category, and  $E_{i,j}$  is the expected frequency given by

$$E_{i,j} = \frac{n_i n_j}{n}$$

where  $n_i$  is the total number of observed samples for level  $i$  of the first category and  $n_j$  is the total number of observed samples for level  $j$  of the second category. The test statistic (8.4.2) can be tested using the  $\chi^2$  distribution with  $(r - 1)(c - 1)$  degrees of freedom. If the p-value is smaller than some threshold, here 0.05, this is evidence to reject the null and conclude that there is association between the two categories. Table 32 shows the association p-values when comparing various clusterings.

	Number	$\bar{r}_i$	$SD(\bar{r}_i(t))$	$\bar{y}_i(t)$	$\bar{r}_i(t)$	$\bar{h}_i(t)$	L/R	Lobes
Number	0	0.59	0.16	0.59	0.91	0.26	1	0.91
$\bar{r}_i$	0.59	0	0.29	0	$10^{-4}$	$10^{-7}$	0.52	$10^{-8}$
$SD(\bar{r}_i(t))$	0.16	0.29	0	0.29	0.004	0.23	0.40	0.85
$\bar{y}_i(t)$	0.59	0	0.29	0	$10^{-4}$	$10^{-7}$	0.52	$10^{-8}$
$\bar{r}_i(t)$	0.91	$10^{-4}$	0.004	$10^{-4}$	0	$10^{-4}$	0.06	0.005
$\bar{h}_i(t)$	0.26	$10^{-7}$	0.23	$10^{-7}$	$10^{-4}$	0	1	$10^{-7}$
Left/Right	1	0.52	0.40	0.52	0.06	1	0	1
Lobes	0.91	$10^{-8}$	0.85	$10^{-8}$	0.005	$10^{-7}$	1	0

Table 32: The association  $\chi^2$  p-values between various clusterings of the data

Table 32 shows some similarities and differences to the results obtained by comparing the clusterings using the ARI (see Tables 29 and 30). It can be seen that again there is a high similarity between the lobes of the brain and the residual clustering labels. Recall that the clustering of the raw uncentered data and the model residuals were similar, thus the figures in the tables above are similar. It can also be seen that there is no evidence of a relationship

between the clusterings of the residuals, raw and raw (centered) and the side of the brain which the region is on. However there are some differences between the association test results and ARI results. The association test has revealed more associations between the clusterings than the ARI test did. For example the ARI results between the model residuals and the raw centered data is 0.01 indicating no relationship, however the association p-value is  $10^{-4}$  which is strong evidence for association. This apparent contradiction in results could potentially be due to the high standard deviation of the ARI.

In this section we have seen that there is a relationship between the clustering of the residual means and the lobes in the brain, both by looking at the counts of the numbers of regions in each lobe, and by the ARI comparing the clusterings.

#### 5.4.4 Linear Discriminant Analysis

In this section, LDA will be applied to each of the clusterings to compare the separation of the two clusters in each case. Table 33 shows the accuracy of classification of each regional mean using LOOCV and LDA, details of which have been described in Section 5.1.4.

Clustering	Classification Accuracy	Sensitivity	Specificity
Number	0.80	1	0
$\bar{r}_i$	0.83	0.89	0.74
$SD(\bar{r}_i(t))$	0.80	0	0.96
$\bar{y}_i(t)$	0.83	0.89	0.74
$\bar{r}_i(t)$	0.83	0.65	0.91
$\bar{h}_i(t)$	0.89	0.92	0.83
Left/Right	1	1	1
Two Lobes	0.93	0.98	0.85

Table 33: The classification accuracy, sensitivity and specificity of the regional means using LOOCV and LDA.

From Table 33 we can see how well separated the two clusters are in each case. Using accuracy alone it appears all of the clusterings can be separated using LDA into two very distinct groups. However it can be seen that the clustering using the number of voxels in each region has zero specificity, meaning those regions which are actually in cluster two are never predicted

to be in group two using LOOCV. This highlights that these two cluster groups are not well separated. The clustering based on the left and right sides of the brain have perfect classification, specificity and sensitivity. This is to be expected as two groups provide a natural division between the regions that LDA can detect. Similarly the division of the brain into two pairs of two lobes provides a natural separation. From the other results it can be seen that the results which are generally best in terms of separation are the model residuals, and this reflects what can be seen in Figure 40. It is also interesting that although the clustering based on the raw uncentered data appears very similar to that of the residual data (Figure 42), all three of the measures in Table 33 are worse for the raw data than the residual data, indicating that the residual data clusters are better separated.

## 5.5 Clustering Simulations

Following the result of clustering the residual regional mean vectors which showed a clear spatial partition between the two clusters, in this section we will investigate further if this could be due to chance, or perhaps is a result of some aspect of the modelling. Here regional spatial data will be simulated and as before the spatial model will be fit, and then residual regional means will be clustered. The parameter values chosen in the simulation are the standard values used in Table 10. Figure 54 shows the results of two simulations using different values for the spatial parameter.

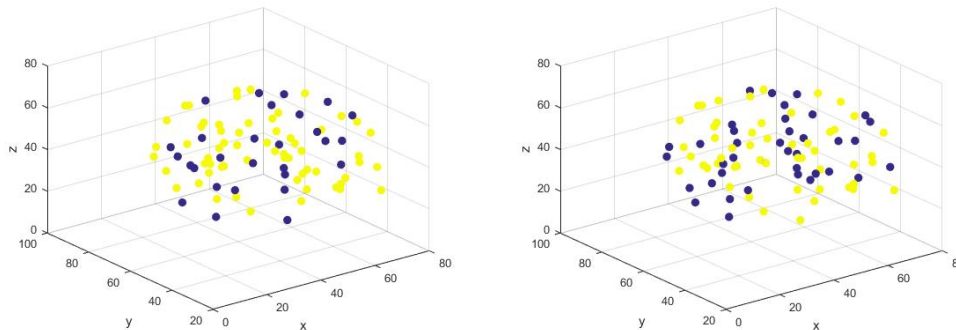


Figure 54: Clustering of regional residual means using simulated spatial data. Left has spatial parameter equal to 0.85 and right 0.25.

From Figure 54 we can see that for the different parameter values, there



is little change in the clustering, and there is no sign of the spatial clustering that was observed when applied to the brain data. It can often be difficult to see clusterings forming in three dimensional figure such as Figure 54.

Figure 55 shows another example of some simulation regional data which has be split into two clusters (top left), generated as described. This three dimensional set of points in two classes is then projected on to the optimal two dimensional surface (top right and bottom left) using LDA as described in Section 5.1.4. Finally the bottom right plot shows the projections alone. From here we can see clearly that there is no clear spatial separation between the two clusters.

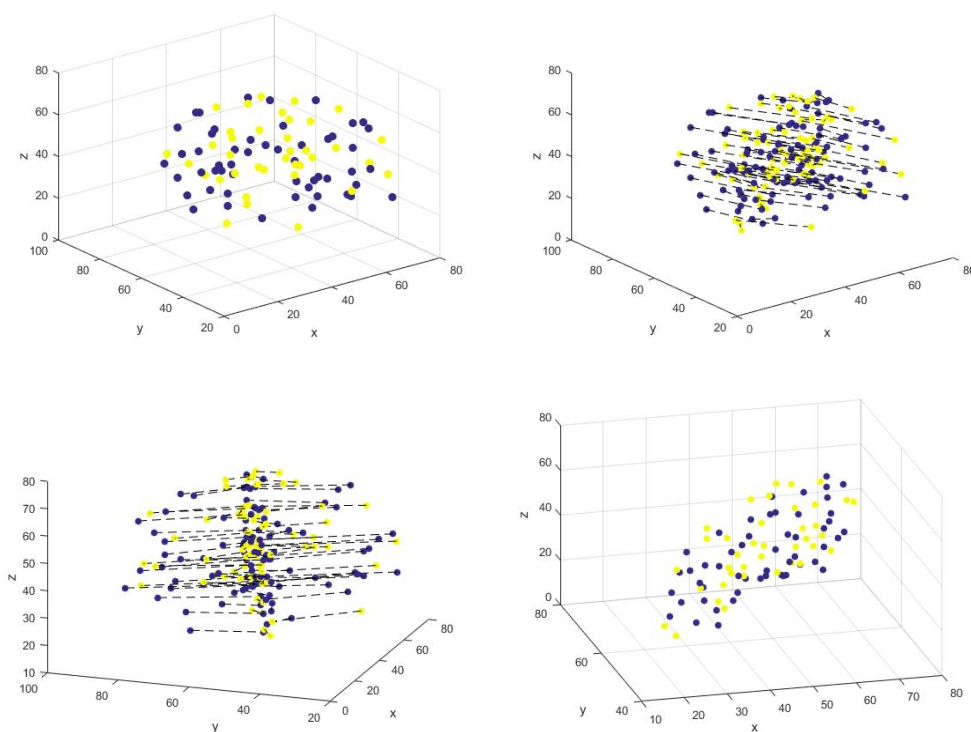


Figure 55: Clustering of regional residual means using simulated spatial data (top left) and projection on to discriminant plane (top right), rotated projections (bottom left), and the discriminant plane (bottom right).

From Figures 54 and 55 we can see no evidence of clustering on simulated brain data. Hence the clustering seen in Figure 40 is likely due to some aspect of the brain data, rather than only spatial correlation in the residuals.

## 5.6 Within Region Clustering

So far we have looked only at the clustering of the regional means. It may be interesting to investigate if voxels within a region are adequately represented by the regional mean for the purpose of clustering. In other words, would a voxel within a cluster be most similar to its own regional mean or to others, and therefore would it be within the same cluster group as its regional mean.

To investigate this we can perform a clustering analysis on the model residual as before. However this time, we will use a proportion (here 90%) of the voxels in each region of interest to calculate the regional mean, and then do the clustering. Following this we can use the remaining 10% of voxels to test if they would be clustered in the same group as its region's mean, or another group. To do this, we will check for each of the 10% of voxels which of the regional means it is closest to, and assign a cluster label to each based on this.

Figure 56 shows the clustering using 90% of all voxels, and from this we can see that the clustering is almost identical to that found when using all voxels to find regional means as expected.

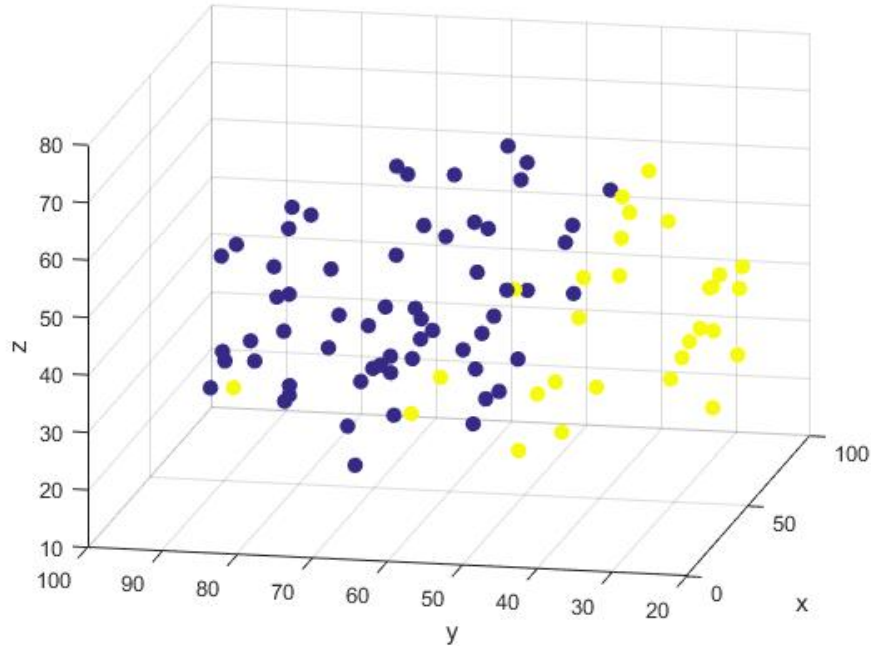


Figure 56: The two clusters formed using 90% of voxels in each region to produce the regional mean.

The remaining 10% of voxels (15178) were each then compared to each regional mean. It was found that 906 (6%) of these voxels were most similar to their own regional mean (using Euclidean distance), and 9932 (65%) were closest to a regional mean with the same cluster label as its own regional mean. We would expect a low percentage of voxels to be most similar to its own regional mean as the signals are noisy and many of the regional means are somewhat similar. The fact that 65% of the voxels had the same cluster label as its own regional mean suggests that there is a bias towards clustering voxels within a region of interest into the same cluster as its mean. However this is not overly convincing as there are only two clusters, so even if the 10% were assigned labels at random, we should expect to cluster them correctly 50% of the time.

## 5.7 Clustering Conclusions And Discussion

By clustering the model residuals we found a spatial partition of the regional residual means in Figure 40. Investigation of various clusterings of the raw data in Section 5.2 found that a similar clustering pattern could be found by clustering the raw uncentered regional means. Although the raw uncentered regional means were clustered according to the mean level (see Figure 42), this cannot be true for the model residuals as all model residuals have mean zero. However there appears to be a relationship between the variance of the raw regional means and the variance of the residual regional means that may explain the similarity in the clusterings (Figure 43).

Clustering using brain anatomy revealed that there was a similarity between the clustering of the residual regional means and clustering based upon two groups of lobes (one consisting of the frontal or parietal lobes, and the other of the temporal or occipital lobes) (see Section 5.3).

Various metrics were used to assess and compare the clusterings. The ARI and Association test (Sections 5.4.2 and 5.4.3) both confirmed the similarity between the model residuals, raw uncentered, and brain lobe clusterings. LDA with LOOCV was also used to assess the separation of the cluster groups for each clustering, and again this found that the residual, raw uncentered and brain lobe clusterings provided the best separation into two clusters.

In Section 5.5, spatial data was simulated and clustered using various values of the spatial parameters, in order to investigate if similar cluster results could be obtained. The results showed that there was no clear clustering of the simulated data, which provides evidence that the results obtained via the modelling is related to the brain data.

Although all clustering discussed here has been the result of  $k$ -means clustering, other clustering methods were investigated, in particular standard hierarchical [79] and spatial hierarchical clustering [50][46] were applied to the data. Hierarchical clustering allowed for different distance metrics to be investigated. Clustering is usually based on Euclidean distance between objects, however other metrics such as correlation or dynamic time warping [21] (which takes into account lagged correlation) could be used. Hierarchical spatial clustering also incorporated the spatial location (as well as similarity between voxels) to the clustering. This method was thought to strengthen the clustering given that the data is spatially correlated. However the alternative methods brought no new insights.

## 6 Topological Data Analysis

In Chapter 5 clustering was used on the raw brain data and the model residuals to investigate the spatial correlation in the data. Topological Data Analysis (TDA) is a relatively recent field of study which is based on analysing data from a topological perspective. In [33] a brief background on the field is given, with reference to several important papers which popularised the theory in the first few years of the 21st century. Most notable are the works of Edelsbrunner et al. (2002) [43], Zomorodian and Carlsson (2005) [85] and Carlsson (2009) [30] on persistent homology which is the foundation of most applications of TDA. The important thing to note about TDA is that it approaches data analysis from a different perspective to classical statistics and as a result it can often reveal different features in a data set. An example of this can be seen in [51] where TDA revealed differences in distributions when conventional marginal and correlation analyses did not.

In this section a brief introduction to some of the necessary theory of TDA will be given, and then this theory will be applied to the brain data to explore if TDA can reveal any distinguishing features in the data that the raw data and the clustering analysis could not.

### 6.1 Important TDA Concepts

Topology is a field of study which is concerned with properties of space which are preserved under continuous deformations. Intuitively this can be thought of as two shapes in space being topologically identical if one can be deformed into the other by stretching, crumpling or bending, but without splitting or gluing. More formally two functions  $f(x)$  and  $g(x)$  are said to be homotopic if there exists a function  $H(x, t)$  such that  $H(x, 0) = f$  and  $H(x, 1) = g$  [30]. If the second parameter of  $H(x, t)$  is thought of as time then this means that at  $t = 0$  we have the function  $f$  and as time increases the function smoothly changes to  $g$  at  $t = 1$ . If  $H$  exists then there exists a deformation from function  $f$  to  $g$ .

There are a number of properties of space which are topologically invariant, which means that the properties do not change under deformations of the space, and the analysis of such properties can reveal differences amongst different spaces. Betti numbers and persistence homology are two such concepts which are topologically invariant and will be the focus of the TDA analysis in this section. Detailed mathematical background on persistence

homology can be found in [33] and [85]. To describe the idea behind Betti numbers and persistence homology, an example will be given.

Consider a function defined on some space. Let  $z(x)$  be the value of the data at location  $x$ . Define a lower level set as

$$L(l) = \{x : z(x) \leq l\} \tag{60}$$

which is the set of locations in space such that  $z(x) \leq l$  for some chosen level  $l$ . Figures 57 and 58 show examples of different level sets on some brain data and some example data respectively, the first in two dimensions and the second in three. The brain data in Figure 57 is from brain slice number 42 and the example data in Figure 58 has been constructed to illustrate some key aspects of TDA clearly in three dimensions. The same principles apply to any data set, but the use of actual brain data in three dimensions does not produce figures which explain the process as clearly.

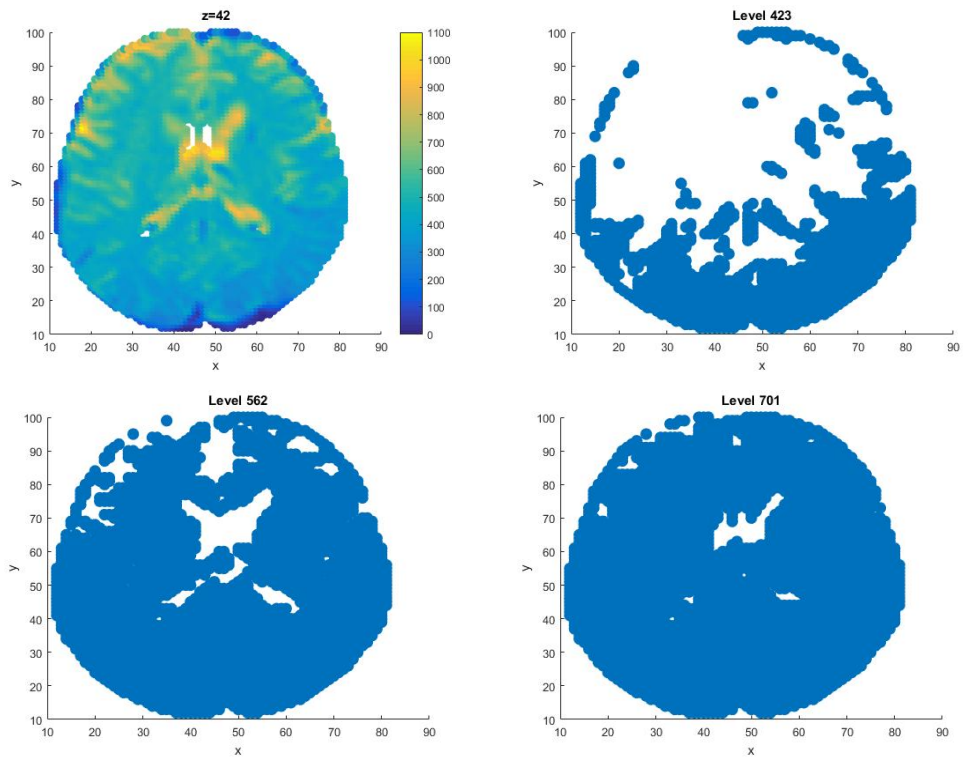


Figure 57: Some level sets of the brain data at  $t = 12$  on slice with  $z$ -coordinate 42. Top left shows the data and the other three show different level sets.

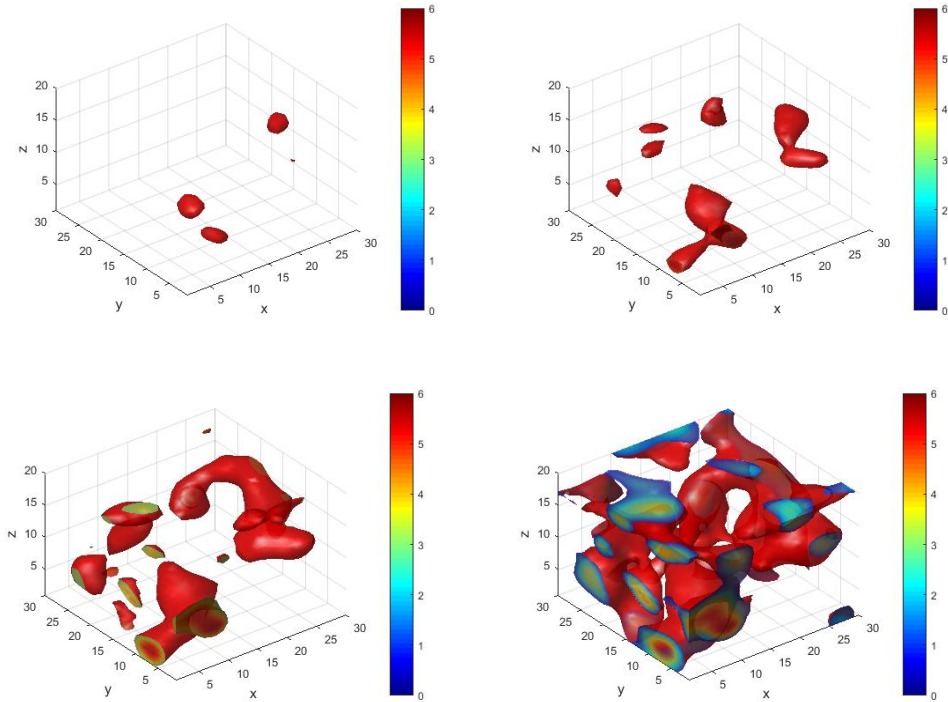


Figure 58: The level sets for a three dimensional example.

Figure 57 (top left) shows the brain data on a slice with  $z$ -coordinate  $z = 42$ . It is colour coded such that dark blue indicates low values at those points, and lighter yellow indicates high values. The other three plots show examples of level sets. As the level increases, more points are added to previous level sets, until at the highest level, the level set contains all points on the brain slice.

Figure 58 shows four examples of level sets on data in three dimensions. It can be seen that for a low level (top left) there are few locations where the value of the data is less than the level, and as the level increases, more and more locations appear. More formally, if we have  $n$  levels

$$l_1 < l_2 < l_3 < \dots < l_n$$

then

$$L(l_1) \subset L(l_2) \subset L(l_3) \subset \dots \subset L(l_n)$$



This process of subsetting the data using level sets is known as filtration, and persistent homology is concerned with analysing how topology evolves as we filter through the data using level sets. One quantity of interest which is topologically invariant is the number of components at each level. Here a simple example will be given to explain the notion of components.

Consider Figures 57 and 58. As the level increases, the size of the level set increases and eventually some of the locations are close enough together that we say they are connected. For the purpose of this work, where data is observed on a lattice, we define the condition for connectivity as lattice locations in the level set which share a lattice edge. In the most simple case of a level set which contains two points  $x_1$  and  $x_2$ , we say that they are connected and form a component if  $d(x_1, x_2) = 1$  where  $d$  is the distance between  $x_1$  and  $x_2$ . Let us denote this component  $\mathcal{C}_1 = (x_1, x_2)$ .

As the level increases, the level set grows larger. In our example suppose now that the level set increases and  $m$  new points  $x_i, i = 3, 4, \dots, m + 2$  are added to the two points in the previous level set. Consider each point in turn. If  $\min(d(x_3, x \in \mathcal{C}_1)) = 1$ , that is if  $x_3$  shares an edge with either  $x_1$  or  $x_2$ , then  $x_3$  is added to the component and so  $\mathcal{C}_1 = (x_1, x_2, x_3)$ . However if  $\min(d(x_3, x \in \mathcal{C}_1)) > 1$  a new component is formed and so  $\mathcal{C}_2 = x_3$ . For the remaining points, one of three things can happen. Firstly, the point may not share an edge with any points in previous components, thus forming another component. Secondly, the point may share an edge with point(s) in exactly one component set and so it will be added to that component. Finally, the point may share an edge with point(s) in more than one component set. In this case the components merge. The result of this process is that we have a set of components associated with this level.

As the level increases further and more points are added, the same rules apply for adding new points to existing or creating new components. However in the case where a new point shares an edge with a member of more than one existing component we add the rule that the component which was formed at the latter level joins the one which formed at the earlier level.

Figure 58 shows examples of components. The top left plot shows three connected components and top right shows six. The number of components at each level is topologically invariant and summarised by the first Betti number, denoted  $\beta_0$ . In two dimensions there is also a second Betti number  $\beta_1$  which is the number of holes/loops at each level, and in three dimensions a third Betti number  $\beta_2$  describes the number of voids at each level. In Figure 57, two-dimensional holes can be seen in all three level plots, the (white)

areas that are completely bounded by a single (blue) component. In Figure 58 a three dimensional hole can be seen in the bottom right plot, that is a set of connected locations with a tunnel in the middle. In contrast a void in three dimensional space is a set of locations which are not in the level set but all locations around them are. A real world example could be a hollow spherical object such as a tennis ball, and the empty space inside would be comparable to a void in three dimensional space. More precise mathematical definitions of these concepts can be found in [33].

Along with recording the number of components at each level, persistence homology requires also the birth and death times of the components. The birth time of a component is the level where a component first appears. As noted, if two components join as the level increases the oldest component survives and the newest component dies at that level. The result is that the first component to be born is the last to die. If calculating components in this manner using lower level sets as described, then it can be deduced that the first component refers to the global minimum in the data set, and other components refer to local minima. To highlight the idea of birth and death times, Figure 59 shows a simple example of some data consisting of 16 points on a four by four lattice, and Figure 60 shows some level sets of this data.

2.10	2.92	4.98	2.61
3.18	3.13	2.86	1.96
4.59	3.71	3.42	4.78
1.87	2.76	3.98	0.69

Figure 59: Example data set on a four by four lattice.

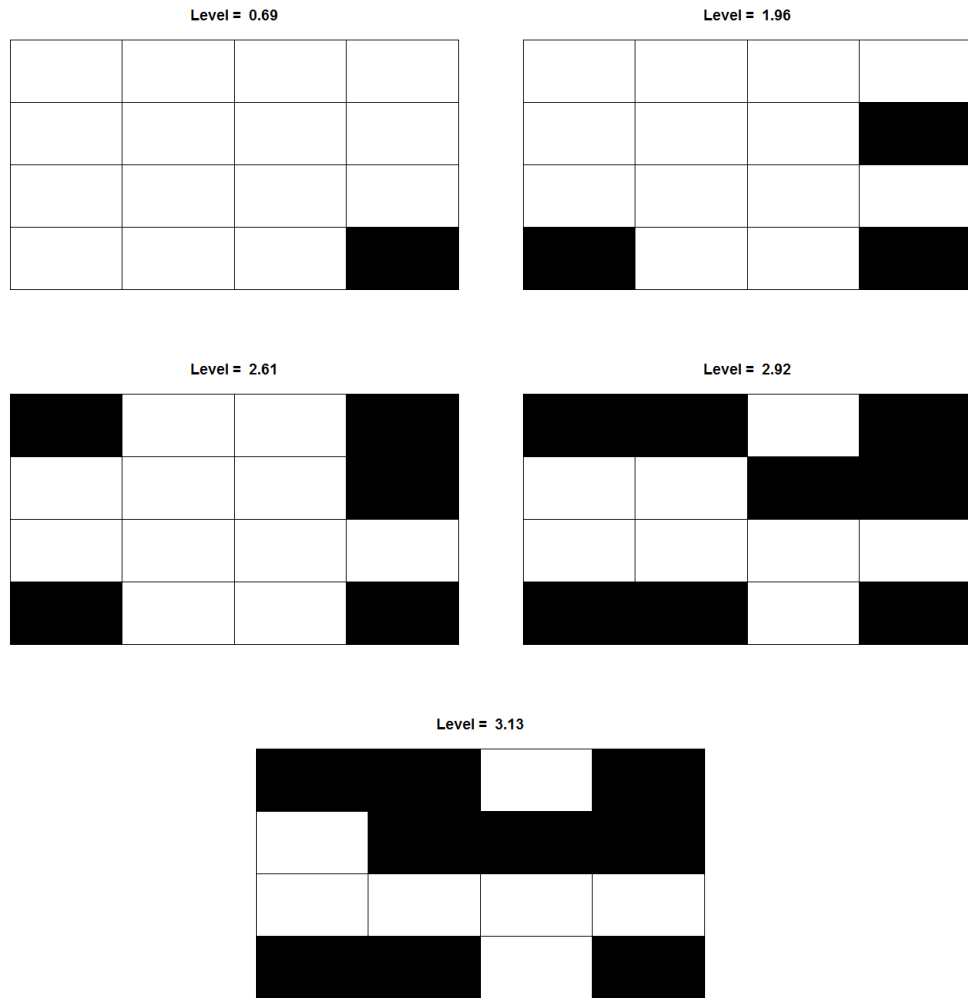


Figure 60: The level sets of some example data as the level increases.

Figure 60 shows the level sets at five different levels on the example data. The first plot shows the first component is born at level 0.69. This is the minimum of the 16 data points on the lattice as seen in Figure 59. Raising the level to 1.96, two new component have been born, thus there are now three components, so at this level  $\beta_0 = 3$ . By level 2.61, two more data points have been added to the level set. One forms a new component (top left point), however the other point is close enough to an existing component (seen at level 1.96, second row fourth column) that it joins, or becomes

connected to that component. At this level of 2.61, although there are five points in the level set, there are only four connected components hence  $\beta_0 = 4$ . At level 2.92 many more points are added to the four components, but no new components are formed as they are all close enough to an existing component to be connected. At level 3.13, one additional point is added to the level set. This point then connects the components first seen at level 1.96 (second row, fourth column) and level 2.61 (top left). At this level (3.13) it is said that the component that was born later dies, as it joins the component which was born at an earlier level. This example highlights how components are formed, including how birth and death times are calculated. An equivalent description of the components is the position of local minima on the field. Figure 60 shows the first component is born at level 0.69. This is the minimum of the 16 data points on the lattice. New components are born only if they are local minima on the field, and so are not connected to existing components. For a two dimensional field the holes can similarly be described using the local maxima of the field.

## 6.2 Persistence Diagrams

The history of components, holes and voids being born and dying as level sets change are often visualised using persistence diagrams, which plot birth against death levels, with a separate plot for each of the features. Figures 61, 62 and 63 show the persistence diagrams associated with the examples considered so far.

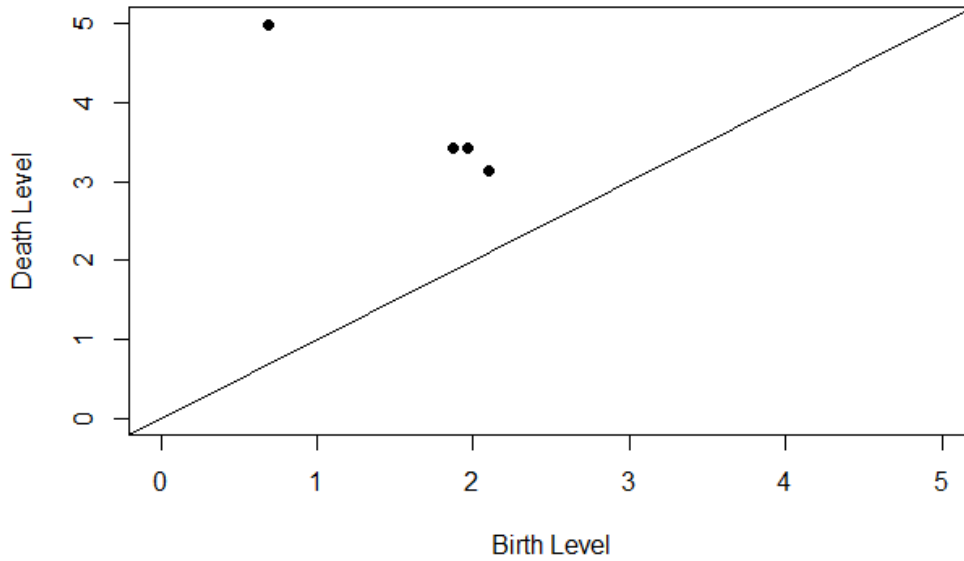


Figure 61: Persistence diagram of the components of the example four by four lattice of points.

Figure 61 shows the persistence diagram of the simple four by four lattice example data shown in Figure 59. The four points in this persistence diagram correspond to the four components described in Figure 60. As the data set is very small, there are no holes in this example. Next, Figure 62 shows the persistence diagram of the brain slice seen in Figure 57.

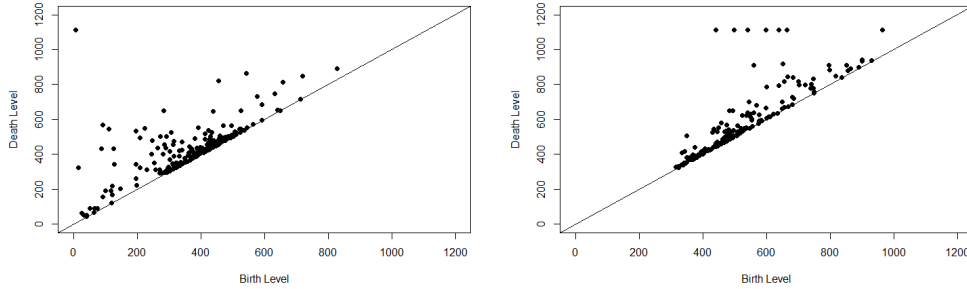


Figure 62: Persistence diagrams of the components (left) and holes (right) of the brain slice  $z = 42$  at  $t = 12$ .

In Figure 62 there is a separate persistence diagram for the components and holes of the two dimensional brain slice. Figure 63 shows the persistence diagrams for the three dimensional cuboid in example 58.

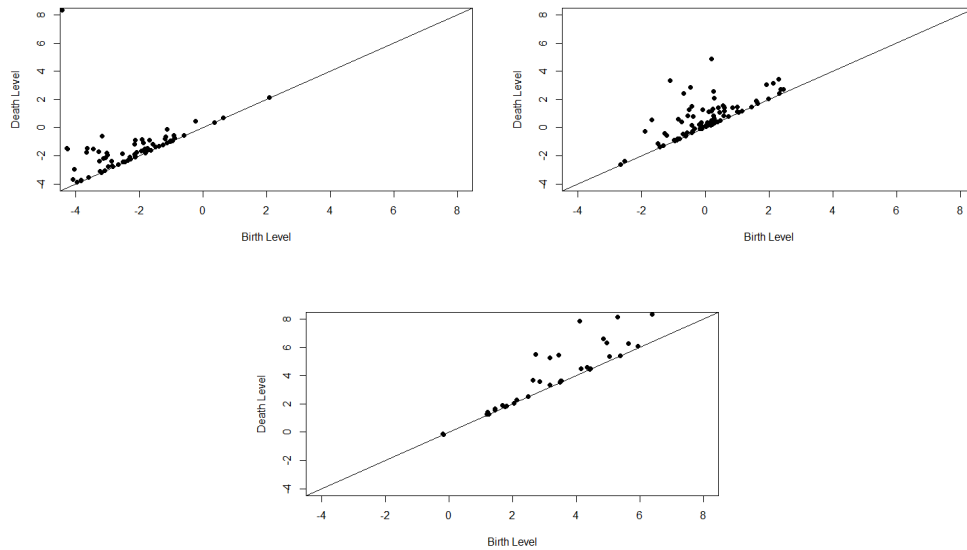


Figure 63: Persistence diagrams of the components (top left), holes (top right) and voids (bottom centre) of the three dimensional cuboid.

In Figure 63 there are three persistence diagrams, one for the components, holes and voids of the three dimensional cuboid. Finally, Figure 64 shows

the persistence diagram of the components of the full brain data set at  $t = 1$ , and will be used to explain how persistence diagrams can be statistically summarised.

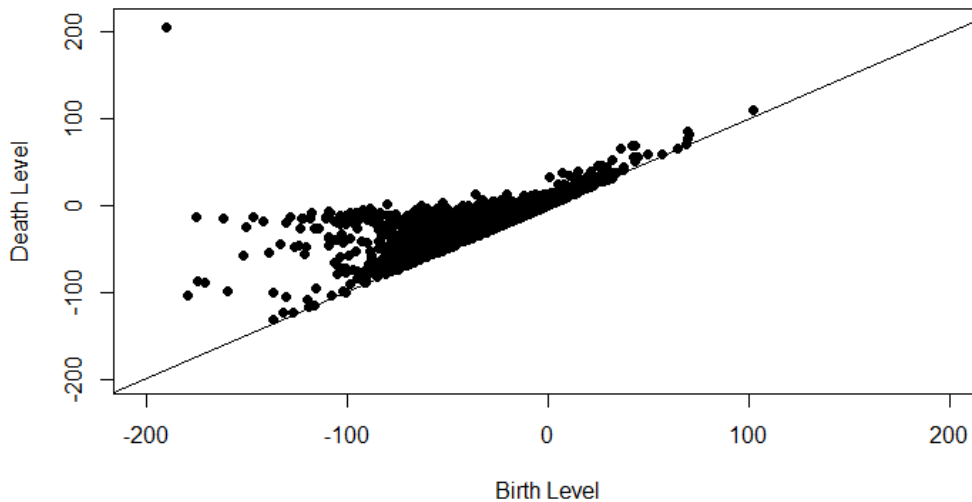


Figure 64: Persistence diagram of the components of the full data set at first time point.

In Figure 64 each point corresponds to a component of the full data set for the first time point. From the plot it can be seen that there is a diagonal for which no points lie below. This is because a component must be born before it dies. There is a point in the top left of the plot which corresponds to the birth and death of the first component as explained in Section 6.1. This point corresponds to the global minima of the data. Points on a persistence diagram which are very close to the diagonal correspond to components which die very soon after birth, and these correspond to short lived local minima. In terms of the data, these points correspond to noisy sections in the signal.

In [51] various statistics are used to summarise a persistence diagram. This is useful as the statistics can then be used to investigate differences between persistence diagrams. The method used in [51] to summarise a persistence diagram is as follows. First calculate the convex hull which encloses



a proportion of the points in the persistence diagram. The convex hull of a set of points is a convex polygon, ie all points in the set can be joined by a straight line which lies on or within the polygon. Precise mathematical definitions can be found in [4]. There are many known algorithms to calculate the convex hull of a set of points, for example see [48], [32] or [14]. The method used here is most similar to [14], and is given by the following algorithm, which splits the task by calculating both a lower and upper hull of the set, together making the complete convex hull. Let  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$  be the  $n$  coordinates of the points in the set. To calculate the lower convex hull:

1. Let  $P_1 = (x_j, y_j) = \min_{x_i}(x_i, y_i)$  be the coordinate with the smallest  $x$  value. This is guaranteed to be on the convex hull.
2. Calculate  $\theta_i = \arcsin\left(\frac{y_i - y_j}{\sqrt{(y_i - y_j)^2 + (x_i - x_j)^2}}\right)$  for  $i = 1, 2, \dots, n$  and  $i \neq j$ . This is the angle between  $P_1$  and all other points in the set.
3. Let  $P_2 = (x_k, y_k)$ ,  $k \leq n$  be the point which minimises  $\theta_i$ . This is the next point which lies on the lower convex hull.
4. Repeat steps 2 – 3 for each new point on the convex hull until  $P_l = (x_l, y_l) = \max_{x_i}(x_i, y_i)$ , then  $P_1, P_2, \dots, P_l$  are the coordinates of the vertices of the lower convex hull.

This process can be thought of as starting with the point with the lowest  $x$ -coordinate and moving in a counter clockwise direction to the point with the largest  $x$ -coordinate, passing through points which must lie on the convex hull. Similarly, to calculate the upper convex hull:

1. Start with  $P_l = (x_l, y_l) = \max_{x_i}(x_i, y_i)$  which is the coordinate with the largest  $x$  value.
2. Calculate  $\theta_i = \arcsin\left(\frac{y_i - y_l}{\sqrt{(y_i - y_l)^2 + (x_i - x_l)^2}}\right)$  for  $i = 1, 2, \dots, n$  and  $i \neq l$ . This is the angle between  $P_l$  and all other points in the set.
3. Let  $P_{l+1} = (x_m, y_m)$ ,  $m \leq n$  be the point which maximises  $\theta_i$ .

4. Repeat steps 2 – 3 for each new point on the convex hull until  $P_u = (x_u, y_u) = \min_{x_i} (x_i, y_i)$ , then  $P_l, P_{l+1}, \dots, P_u$  are the coordinates of the vertices of the upper convex hull. Note that  $P_u = P_1$ .

Then  $P = (P_1, P_2, \dots, P_l, P_{l+1}, P_{l+2}, \dots, P_u)$  are the vertices of the complete convex hull. The convex hull will give the overall shape of the distribution of points. Note that because outliers (such as the birth/death level of the first component) will have a large impact on the shape of the convex hull, often convex hulls are constructed using a subset of the points, and these are known as convex peels. Figure 65 shows the convex peels of the persistence diagram in Figure 64.

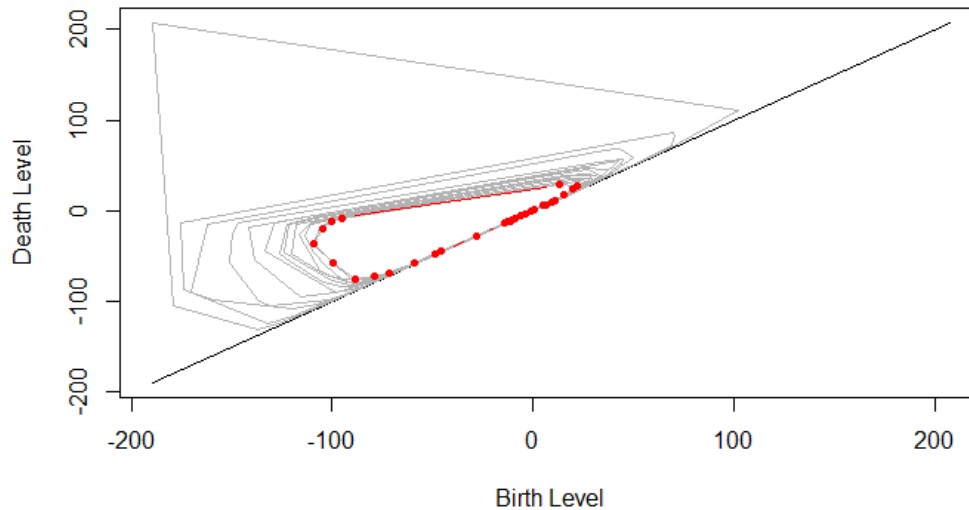


Figure 65: 95% convex peel (red dots) of the persistence diagram of the components of the full data set at the first time point.

Figure 64 shows in red the 95% convex peel, ie the convex hull constructed using 95% of the points. Here the 5% of points with the lowest  $x$ -coordinate were removed, hence the extreme points, such as the birth and death level of the first component will not impact the shape of the convex hull when making

comparisons between different persistence diagrams. The grey lines in Figure 64 show higher percentile convex hulls, which are shown for illustration.

Using the convex peel, a number of statistics can then be derived to summarise a persistence diagram. In [51] these statistics are as follows.

- The birth centroid,  $C_b$
- The death centroid,  $C_d$
- The area,  $A$
- The perimeter,  $P$
- The filamentarity,  $F$
- The number of points,  $N$

These statistics are derived from the final convex peel (red in Figure 65). As the convex peel is a polygon, there exist well known formula to calculate the centroid and area using the vertices of the polygon. The birth and death centroid are the birth and death level of the centroid of the convex peel. The perimeter of the convex peel can be found by calculating the distance between adjacent vertices and summing. The filamentarity is defined as

$$F = \frac{P^2 - 4\pi A}{P^2 + 4\pi A}. \quad (61)$$

If  $F = 1$  then the the points lie on a line, and if  $F = 0$  the convex hull makes up a perfect circle. Finally, the number of points refers to the number of points on the persistence diagram, which will vary for each time point, or for each field we consider.

An alternative function that can be used to describe the shape of a persistence diagram is the landscape function [27]. The landscape function first rotates the birth death points  $(b, d)$  to  $(x, y) = (\frac{b+d}{2}, \frac{d-b}{2})$  before fitting a piecewise linear tent function to each point. The  $k$ -th landscape function is then defined as the  $k$ -th largest of these tents. Full detail on how the landscape function is constructed can be found in [27]. Intuitively the landscape function can be thought of fitting a tent (triangle) function to each point, with the first landscape function being the trace of the outline of the peaks of all such tents. The second landscape function removes the peaks of the first, and traces out the remaining peaks, and so on. Figure 66 illustrates

this idea by showing an example of the first three landscape functions of the components persistence diagram at the first time point of the full, raw (centered) brain data.

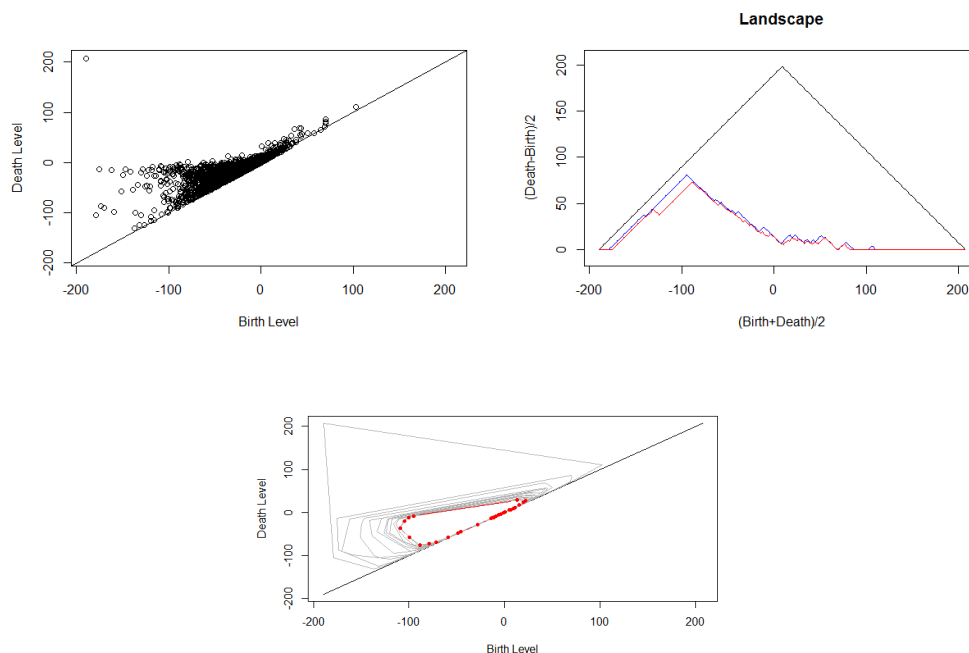


Figure 66: Persistence diagram of the components of full brain raw centered data at  $t = 1$  (top left), and first three landscape functions (black, blue, red respectively) (top right), and 95% convex peel (bottom centre).

From Figure 66, it can be seen that the tent function about the first component (seen in the top left corner of the top left figure) surrounds all points, and so we obtain the first landscape function, which in this case has one peak. Removing this point, and recalculating the second landscape function we obtain multiple peaks, as here each tent function is smaller as the points are closer to the diagonal, and so will not enclose all points. Hence tracing the outline of these tents will give a piecewise linear tent function with multiple peaks. In contrast to using a convex peel, the landscape function can capture more closely the shape of the persistence diagram. Figure 67 shows another example which highlights this more.

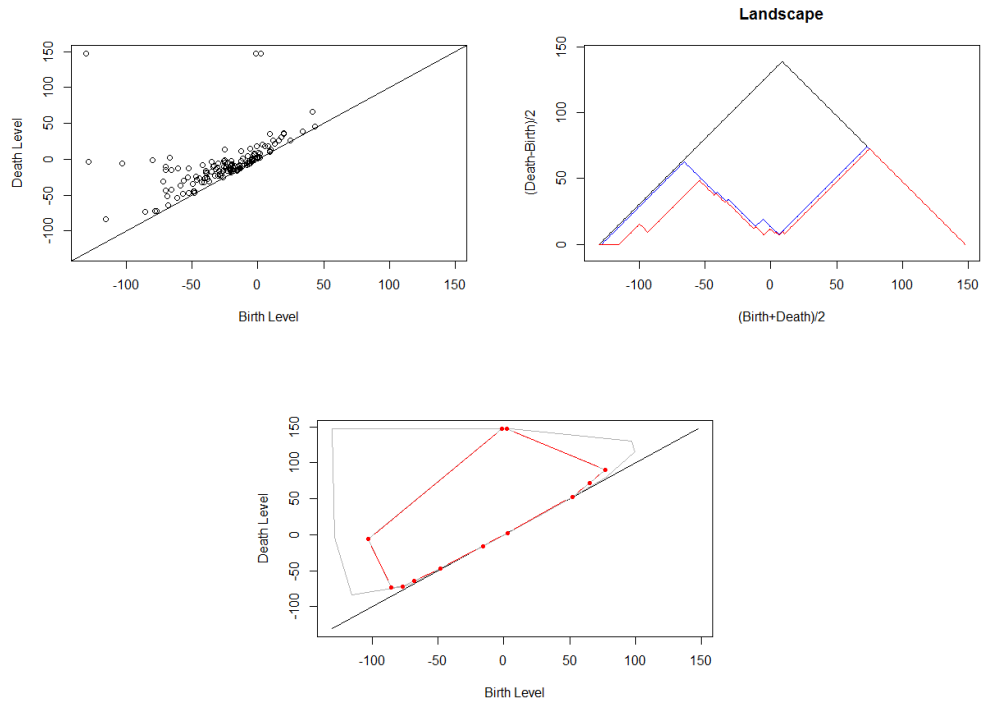


Figure 67: Persistence diagram of the components of ROI 3 raw centered data at  $t = 1$  (top left), and first three landscape functions (black, blue, red respectively) (top right), and 95% convex peel (bottom centre).

Figure 67 shows the persistence diagram, first three landscape functions and 95% convex peel for the components for ROI 3 at  $t = 1$ . It can be seen from the second and third landscape functions that there appears to be two clusters in the diagram which cannot be seen from the convex peel. Like with the convex peel, a set of summary statistics can be used to summarise the landscape functions.

Finally, the bottleneck distance is a quantitative way of comparing the similarity between persistence diagrams. This measure is based on finding the optimal mapping between points in two persistence diagrams. A formal description can be found in [44] with only an intuitive description given here. Let  $A$  and  $B$  denote two persistence diagrams containing  $n$  and  $m$  points respectively. In order to map points between  $A$  and  $B$  it is first necessary for both persistence diagrams to contain the same number of points. The first step of calculating the bottleneck distance is to add  $m$  points to persistence

diagram  $A$  and  $n$  points to  $B$ . The additional  $m$  points are the projections of all  $m$  points in  $B$  to the diagonal of  $A$ , with the same idea for adding  $n$  points to  $B$ . This results in both persistence diagrams containing  $n + m$  points. The additional points on the diagonals would in theory correspond to points which die at the same time they are born, hence they do not change the underlying structure of the field that the persistence diagrams represent.

The second step of calculating the bottleneck distance is to map each point in  $A$  to a unique point in  $B$  and calculate the distance between all mapped points. Let  $p_1$  denote the first choice of mapping from points in  $A$  to  $B$  and  $d_{1i}$  denote the distance between the  $i$ th mapped pair. Then the maximum such distance for the first possible mapping is

$$M_{p_1} = \max_i(d_{1i})$$

Repeating this procedure for all possible mappings results in a set of values  $M_{p_1}, \dots, M_{p_q}$  where  $q$  is the total number of possible mappings between the points. The bottleneck distance is then given by

$$BD(A, B) = \min_j(M_{p_j})$$

The bottleneck distance, along with the summary statistics derived from the convex peels and landscape functions, will be used to compare persistence diagrams and will be the focus of the next few sections.

## 6.3 Topology Of The Model Residuals - Full Brain

In this section, the topology of the model residuals for the full brain will be explored using the methods described in Section 6.2.

### 6.3.1 Convex Peels And Landscape Functions

Figures 68 and 69 shows how the vector  $(C_b, C_d, A, P, F, N)$  of summary statistics for convex peels and landscape functions respectively, change over time, when considering connected components. Similar results were obtained for holes and voids and can be found in the Appendix 10.2.

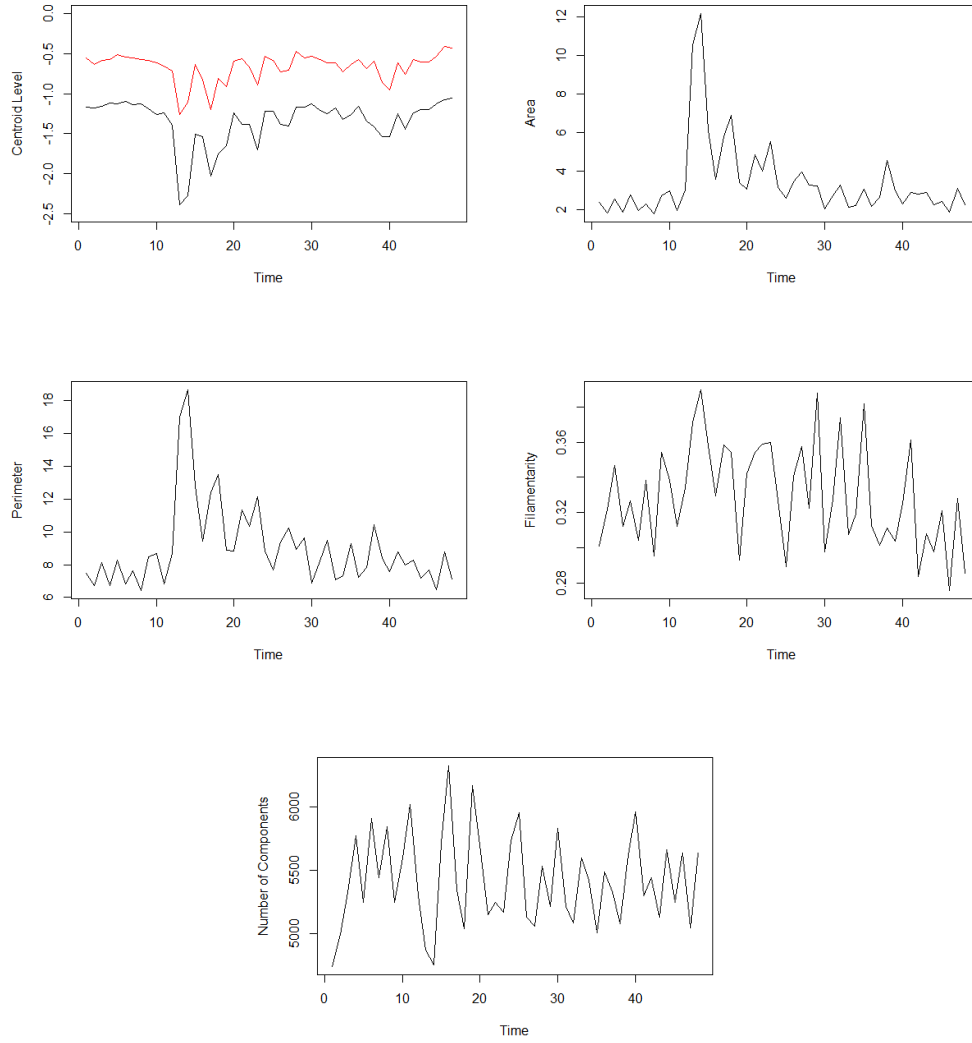


Figure 68: Summary statistics derived using 95% convex peels of persistence diagrams for components of the full brain model residuals. Top left shows the birth (black) and death (red) centroid levels over time, and the remaining plots the other summary statistics.

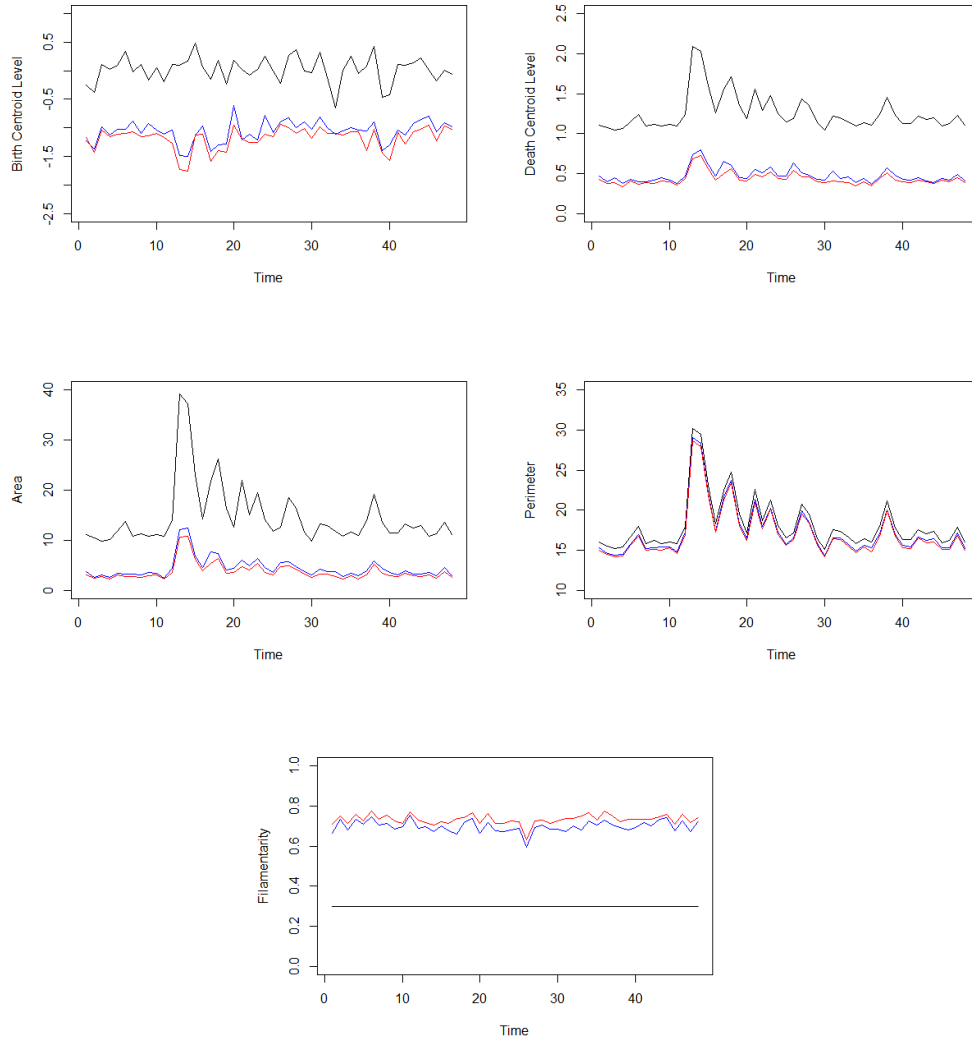


Figure 69: Summary statistics derived using the first three landscape functions (black, blue, red) of persistence diagrams of the components of the full brain model residuals.

As seen, the landscape and convex peels of the components show a large increase in various statistics at  $t = 12$ , the onset of activity. This indicates that the persistence diagram at this time, and so the topology of the field, changes dramatically. Figure 70 shows the mean and variance profile of the



full brain model residuals.

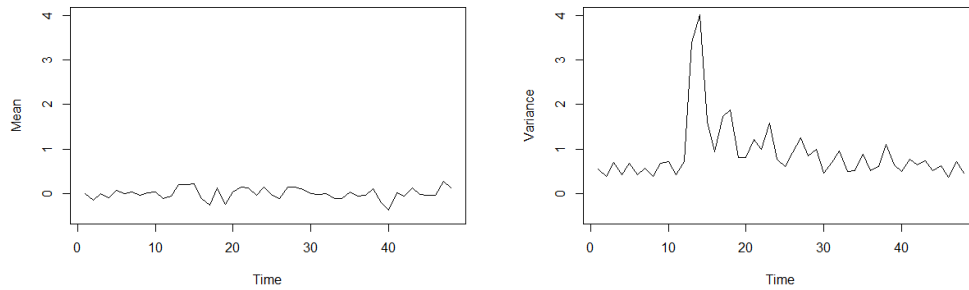


Figure 70: The mean and variance of the full brain model residuals.

From Figure 70 it can be seen that although each voxel has standardised residuals, the variance across voxels at each time point changes significantly at  $t = 12$ . This could explain the results seen so far in the persistence diagrams.

### 6.3.2 Bottleneck Distances

Figure 71 shows the bottleneck distances of the full brain model residuals at adjacent time points.

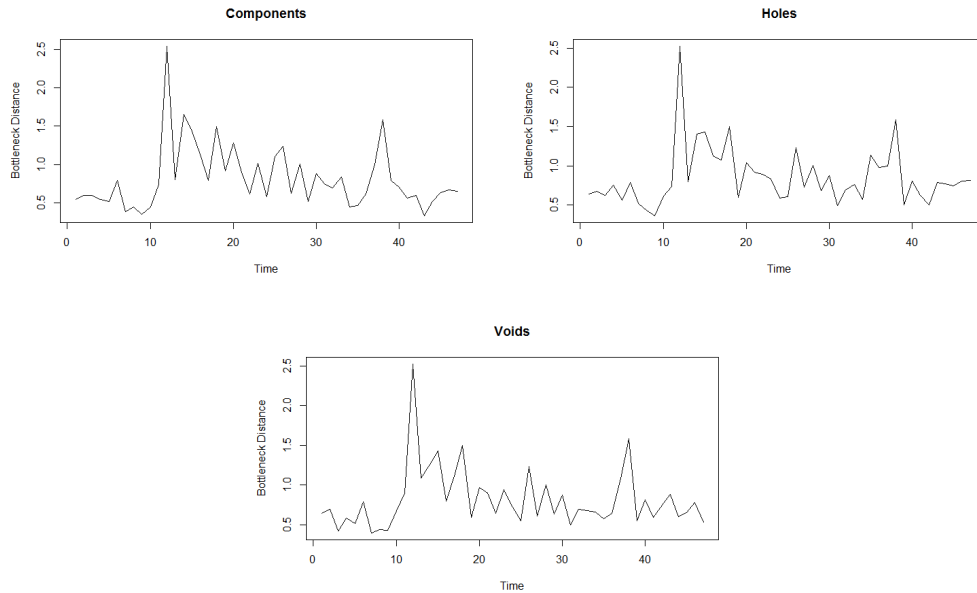


Figure 71: The bottleneck distances of persistence diagrams of the full brain at time  $t$  and  $t + 1$ .

As in Section 6.3.1, the bottleneck distances in Figure 71 show an increase in bottleneck distance at the same time as the onset of activity in the experiment.

## 6.4 Topology Of The Model Residuals - Full Brain Standardised

In this section, the model residuals (Equation (26)) will be standardised at each time point before exploring the topology.

### 6.4.1 Convex Peels And Landscape Functions

Figures 72 and 73 shows the summary statistics for the components derived from the convex peels and landscape functions. Similar results were obtained for holes and voids and can be found in the Appendix 10.3.

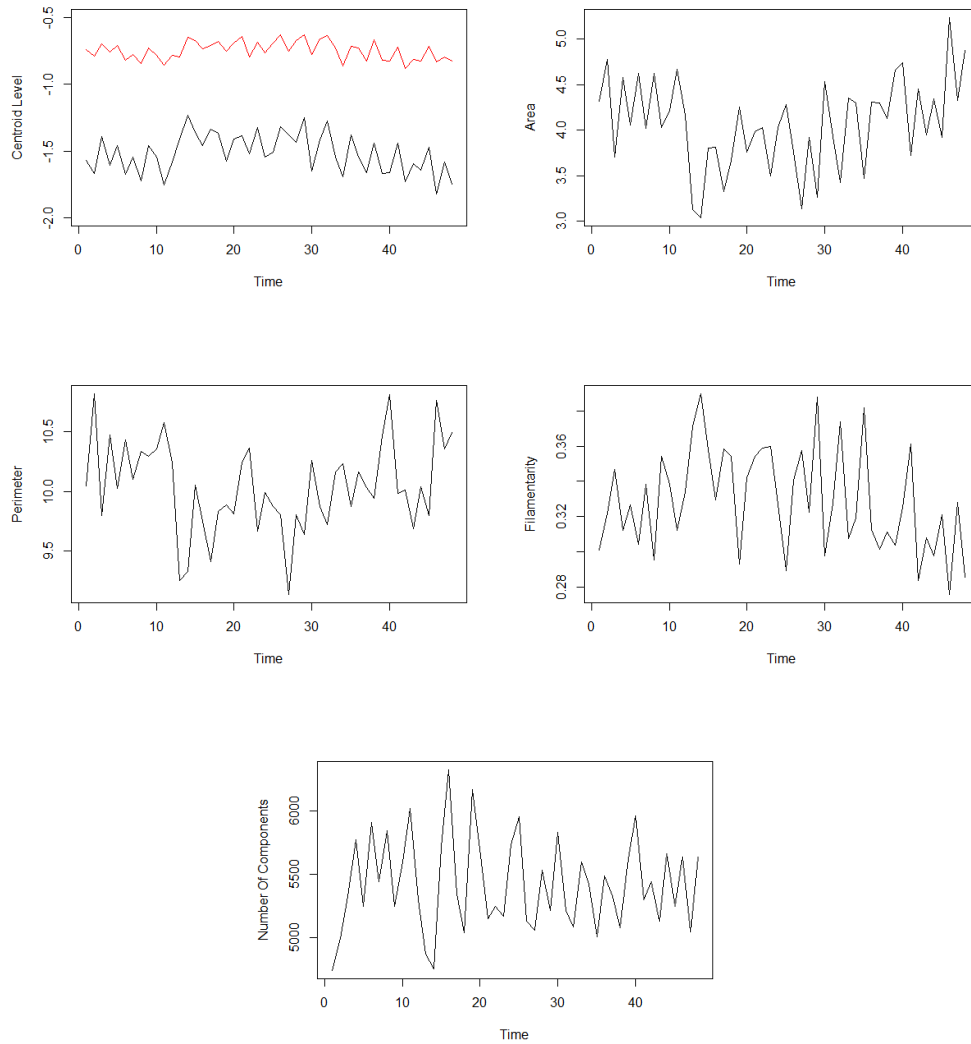


Figure 72: Summary statistics derived using 95% convex peels of persistence diagrams of the components of the full brain model residuals (standardised). Top left shows the birth (black) and death (red) centroid levels over time, and the remaining plots the other summary statistics.

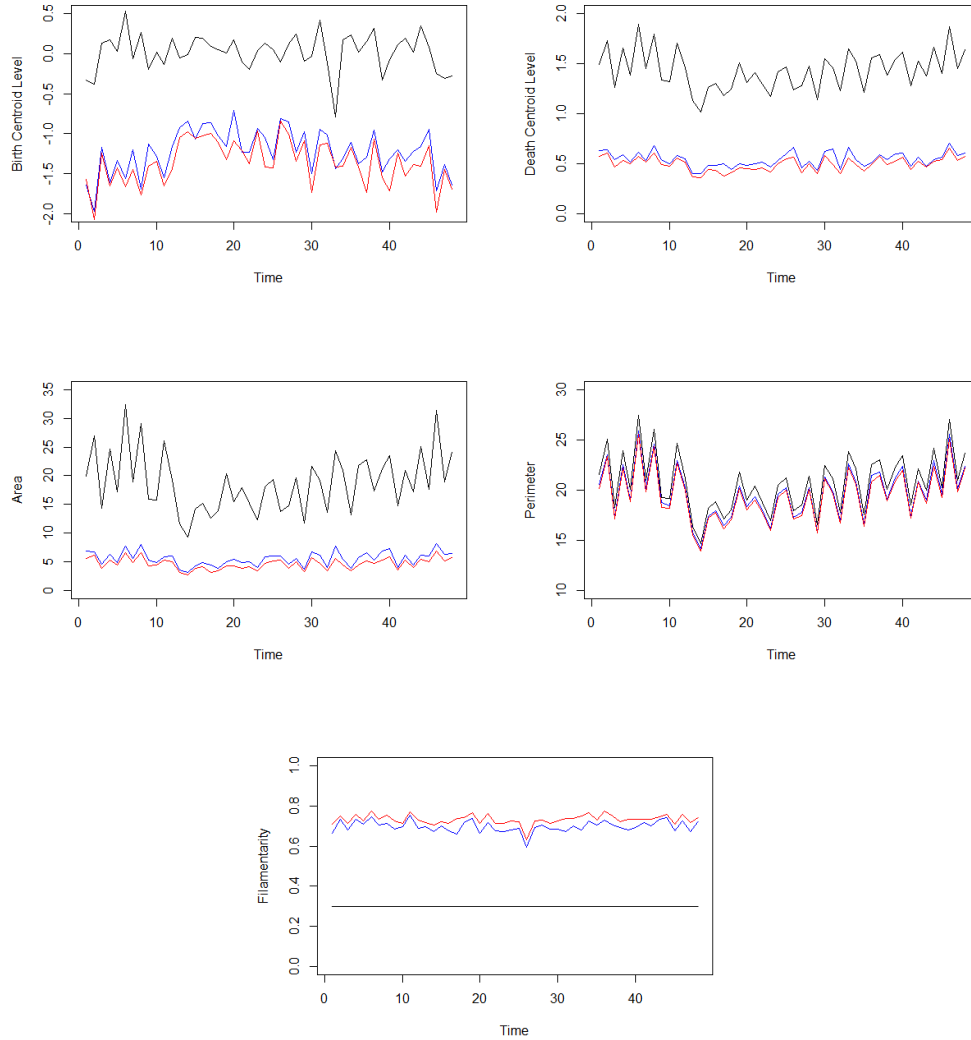


Figure 73: Summary statistics derived using first three landscape functions (black, blue, red) of persistence diagrams of the components of the full brain model residuals (standardised).

From Figures 72 and 73 it can be seen that there is still a marked decrease in the area and perimeter statistics at the onset of activity. Thus the topology is picking up change that is not explained by increased variance. One method of determining if there are any significant changes in the mean of a time series

is to apply a change point model to the series [42], [66].

These methods were implemented using the *changepoint* package in *R* and applied to the time series seen in Figures 72 and 73, but no significant change points were detected in any time series. However, as expected, significant change points could be detected for the non-standardised data shown in Section 6.3 . Figure 74 shows one example of this, where two change points were detected for the area summary statistic at  $t = 12$  and  $t = 23$ , the start and end times for the first period of activity. Also shown is the equivalent standardised time series where no change points were found.

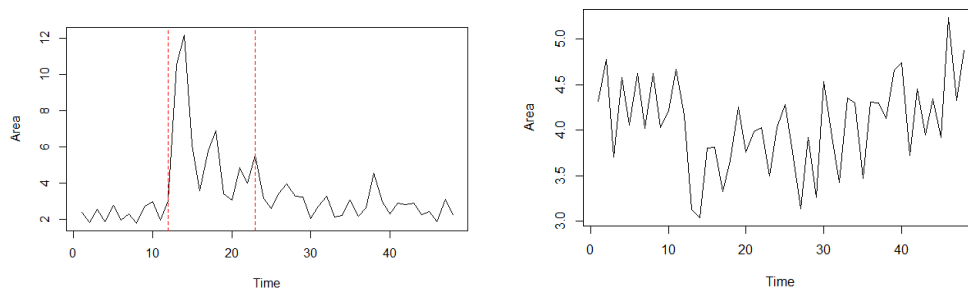


Figure 74: The area of the convex peel over time for the non-standardised (left), and standardised (right) full brain data set. Vertical red dashed lines show the change points in the time series.

#### 6.4.2 Bottleneck Distances

Figure 75 shows the bottleneck distances of the full brain standardised model residuals at adjacent time points.

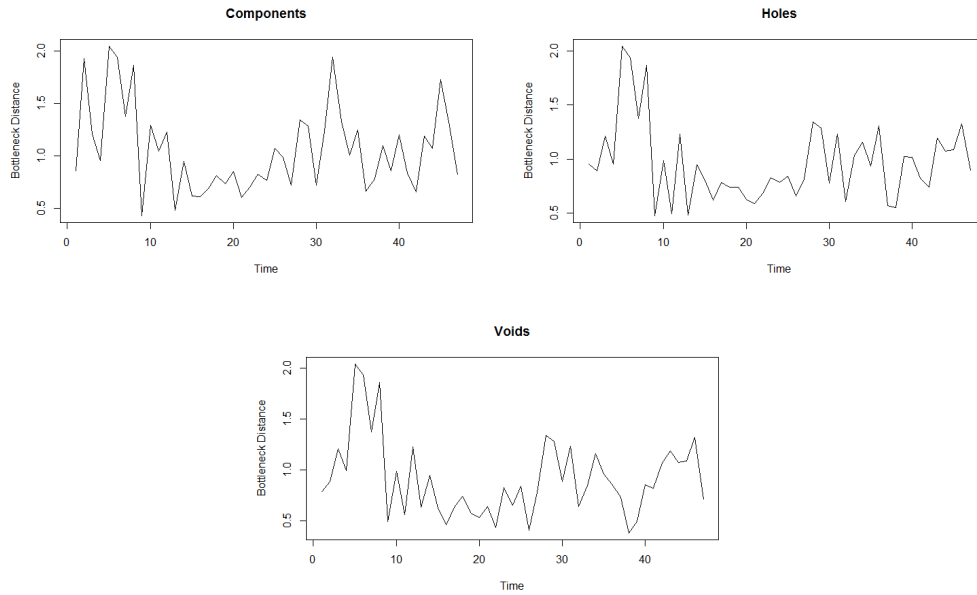


Figure 75: The bottleneck distances of persistence diagrams of the standardised full brain residuals at time  $t$  and  $t + 1$ .

Figure 71 does not show an increase in bottleneck distance at the same time as the onset of activity in the experiment. However for all three features, there is an increase in bottleneck distance between times five and six. Figure 76 shows the persistence diagrams at this time for components.

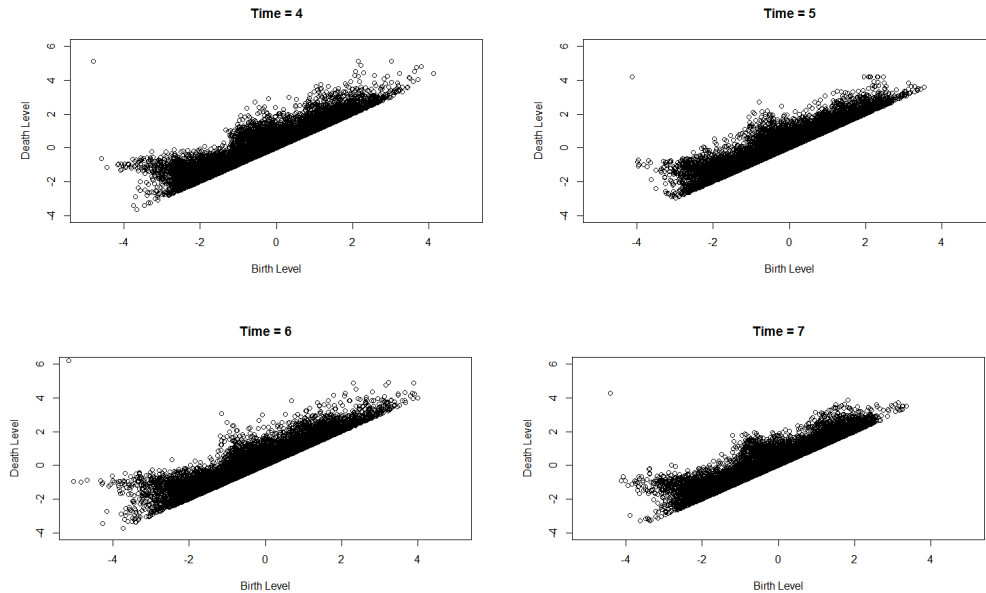


Figure 76: The persistence diagrams of the standardised full brain residuals at time  $t = 4$  to  $t = 7$ .

From Figure 76 we can see that the points on the persistence diagram at  $t = 5$  are somewhat more compact than those at  $t = 6$ , explaining the large difference in bottleneck distance between these two time points in particular.

## 6.5 Summary

In this chapter we have introduced some relevant background information on Topological Data Analysis. The most popular area of TDA is persistence homology which is concerned with the births and deaths of components, holes and voids of a field. As persistence diagrams are computed for a single field, it is useful to summarise the persistence diagram so that many diagrams can be compared, for example for comparing two different fields, or a field which changes over time. We have described two methods that can be used to achieve this; convex peels and landscape functions. From each of these we can obtain a set of summary statistics which can then be used to compare multiple persistence diagrams. These ideas were then applied to the model residuals of the full brain data set where it was found that the evolutions of summary statistics showed significant changes at the onset of activity (Figure

68). Upon investigation this seems mostly to be explained by changes in the variance in the data, and so the procedure was repeated on standardised model residuals. From this a change point analysis was applied to the time series vectors of each summary statistic and no real evidence was found for any significant change in the statistics over time.



## 7 TDA: The Number Of Components And Holes

In Section 6.1 it was noted that the components and holes are related to the local minima and maxima of the field respectively. In [51] the expected number of components and holes of a two dimensional stationary isotropic Gaussian random field was used to investigate if the number of components and holes of an observed field is consistent with a Gaussian random field. In [51] the fields being tested were observations on a square two dimensional lattice. In this section the aim is to extend this work to irregular lattices such as brain slices, and also to obtain results for three-dimensional brain data. Simulations will be used to investigate if any distinctions can be made between distributions which are all marginally normal with the same correlation structure on the brain slices, and how the correlation structure impacts the number of components and holes on the fields. Finally, we will calculate the number of components and holes for brain slices and the full brain to see if they are consistent with a Gaussian field.

### 7.1 Expected Number Of Components And Holes For A Gaussian Random Field

In this section we will give brief background to the theory presented in [51].

Given a standardised two-dimensional Gaussian random field  $z(x)$ , let  $N_0$  and  $N_1$  be the number of components and holes respectively. The components and holes can also be described using the local minima and maxima of the field. Here we will concentrate on local maxima of the field. Due to the symmetry of the Gaussian distribution, the same results can be applied to local minima.

Let  $z_1 = z(x_1)$  be the value of the field at location  $x_1$  and  $z_2 = z(x_2)$  be the  $k$ -vector of field values of the immediate neighbours of  $x_1$ . As  $z(x)$  is standardised and Gaussian we can write

$$\begin{aligned} z(x) = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} &\sim N \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{21} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix} \right) \\ &= N \left( \begin{pmatrix} 0 \\ 0_k \end{pmatrix}, \begin{pmatrix} 1 & r^T \\ r & R \end{pmatrix} \right) \end{aligned} \tag{62}$$

where  $0_k$  is a  $k$ -vector of zeroes,  $r$  is the  $k$ -vector of correlations between  $z(x_1)$  and  $z(x_2)$ , and  $R$  is the  $k$  by  $k$  correlation matrix of  $z(x_2)$ .

A local maximum occurs at  $x_1$  if  $z_2 < z_1 1_k$ . To calculate this probability let us first consider the conditional distribution  $z_2|z_1$ . In general  $z_2|z_1 \sim N(\bar{\mu}, \bar{\Sigma})$  where

$$\begin{aligned}\bar{\mu} &= \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(z_1 - \mu_1) \\ \bar{\Sigma} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.\end{aligned}\tag{63}$$

In the case where the field is standardised, and using the notation in (62) we have

$$z_2|z_1 \in N(rz_1, R - rr^T).\tag{64}$$

Let  $\phi^{(p)}(x; \mu, \Sigma)$  denote the  $p$ -dimensional multivariate Gaussian probability density function and  $\Phi^{(p)}(x; \mu, \Sigma)$  the corresponding cumulative distribution function. Then the conditional density for  $z_2$  given  $z_1$  can be written as  $\phi^{(k)}(z_2; rz_1, R - rr^T)$  and so

$$\begin{aligned}p_1 &= Pr(z_2 < z_1 1_k) = \int_{z_1} Pr(z_2 < z_1 1_k | z_1 1_k) Pr(z_1 1_k) dz_1 \\ &= \int_{z_1} \left( \int_{z_2 < z_1 1_k} \phi^{(k)}(z_2; rz_1, R - rr^T) dz_2 \right) \phi^{(1)}(z_1; 0, 1) dz_1 \\ &= \int_{z_1} \Phi^{(k)}(z_1 1_k; rz_1, R - rr^T) \phi^{(1)}(z_1; 0, 1) dz_1 \\ &= \int_{z_1} \Phi^{(k)}((1_k - r)z_1; 0_k, R - rr^T) \phi^{(1)}(z_1; 0, 1) dz_1.\end{aligned}\tag{65}$$

A general skew normal distribution can be used [49], [16], [15], [18] to show that

$$\Phi^{(q)}(D\mu; \nu, \Delta + D\Sigma D^T) = \int_y \Phi^{(q)}(Dy; \nu, \Delta) \phi^{(p)}(y; \mu, \Sigma) dy.\tag{66}$$

Using this result with  $p = 1$ ,  $y = z_1$ ,  $\mu = 0$ ,  $\Sigma = 1$ ,  $q = k$ ,  $D = 1_k - r$ ,  $\nu = 0_k$  and  $\Delta = R - rr^T$  gives

$$p_1 = \Phi^{(k)}(0_k; 0_k, R + 1_k 1_k^T - r 1_k^T - 1_k r^T). \quad (67)$$

Therefore to find the probability that  $z_1$  is a local maxima, all that is required is to evaluate the zero mean,  $k$ -dimensional cumulative Gaussian density at  $0_k$ , given  $r$  and  $R$ . Finally, if we sum over all locations  $x_j$  we find that  $E[N_0] = E[N_1] = \sum_j p_j$ .

As noted, in order to calculate  $E[N_0]$  and  $E[N_1]$  we require  $r$  and  $R$ , which depend on the point and its immediate neighbours. As noted in [51] immediate neighbours of a point in two dimensions could be defined by a shared edge, or by either a shared edge or vertex. We will define two locations to be neighbours only if their associated voxels have a shared edge. In two dimensions, this means that a point may have up to a maximum of four neighbours. Figure 77 show all possible neighbour configurations on a two dimensional irregular lattice assuming isotropy.

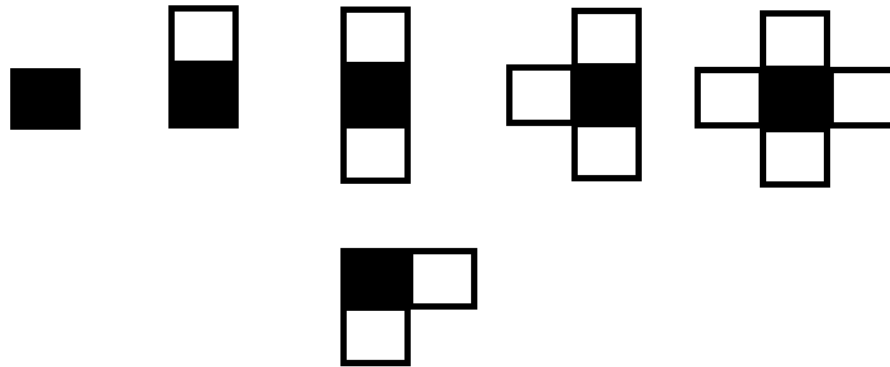


Figure 77: The various two dimensional neighbourhood configurations. Black is point of interest, and white are neighbours to that point. From left to right: zero to four neighbour configurations.

Assuming that the field is stationary, the expected number of maxima,  $E[N]$ , is then given by

$$E[N] = \sum_i |P_i| f(R(P_i), r(P_i)) \quad (68)$$

where  $P_i$  is point type  $i$ , and  $f(R(P_i), r(P_i))$  is the probability that point type  $i$  is a maximum, which depends on local correlations  $R$  and  $r$  as defined above. Point type here is used to distinguish points on the lattice which have a different number of neighbours. Hence to calculate the expected number of maxima, we require the total number of points for each of the different point types, and the probability of each point type being a maximum.

We can also find an expression for  $Var(N)$ . First let  $I_i$  be an indicator, such that  $I_i = 1$  if there is a maxima at position  $x_i$  and  $I_i = 0$  otherwise. Therefore  $N = \sum_i I_i$ . Hence  $Var(N) = Var(\sum_i I_i) = \sum_i \sum_j Cov(I_i, I_j) = \sum_i \sum_j (E[I_i I_j] - E[I_i]E[I_j])$ . To evaluate this expression, we need to consider separately three possible arrangements of neighbours.

Let  $x_{11}$  and  $x_{12}$  be two positions on the field, with  $x_{21}$  the  $k_1$ -vector of neighbours of  $x_{11}$  and  $x_{22}$  the  $k_2$ -vector of neighbours of  $x_{12}$ .

The first possible arrangement of neighbours is such that  $x_{21}$  and  $x_{22}$  are distinct, ie the case where  $x_{11}$  and  $x_{12}$  have no common neighbours. In this case  $(z_1, z_2)$  where  $z_1 = (z_{11}(x_{11}), z_{12}(x_{12}))$  and  $z_2 = (z_{21}(x_{21}), z_{22}(x_{22}))$  is Gaussian with zero mean and variance matrix

$$\begin{pmatrix} R_{11} & R_{12}^T \\ R_{12} & R_{22} \end{pmatrix}.$$

In a similar derivation to that given above, Equation (66) can be used to find that

$$E[I_1 I_2] = \Phi^{(k)}(0_k; 0_k, R_{22} - R_{12} R_{11}^{-1} R_{12}^T + D R_{11} D^T) \quad (69)$$

where  $k = k_1 + k_2$  and  $D = J - R_{12} R_{11}^{-1}$  where

$$J = \begin{pmatrix} 1_{k_1} & 0_{k_1} \\ 0_{k_2} & 1_{k_2} \end{pmatrix}.$$

The second possible arrangement of neighbours is such that  $x_{11}$  and  $x_{12}$  have some common neighbours. In this case  $R_{22}$  is singular, however  $R_{22} - R_{12} R_{11}^{-1} R_{12}^T + D R_{11} D^T$  is in general not singular, hence the same result applies.

The final arrangement of neighbours is such that  $x_{11}$  is itself a neighbour of  $x_{12}$ . In this case  $E[I_1 I_2] = 0$  as it is impossible for both locations to be local maxima, and so  $Cov(I_1, I_2) = -E[I_1]E[I_2]$ .

Using the above results, the variance of the number of maxima can now be calculated in principle. However due to the size of the covariance matrices involved, this becomes computationally expensive. Therefore an estimate of the variance will be used in this work, where we will ignore covariances between points separated by a distance where the covariances are so small they become negligible. Define  $\mathcal{N}_i = \{j : |x_i - x_j| < \delta_0\}$  then we estimate the  $Var(N)$  by

$$\widehat{Var}(N) = \sum_i \left\{ E[I_i](1 - E[I_i]) + \sum_{j \in \mathcal{N}_i} Cov(I_i, I_j) \right\} \quad (70)$$

for a sensible choice of  $\delta_0$ , here three lattice distance units.

The results in this section show how to calculate the expected number of minima and maxima of any Gaussian field. As noted in Chapter 6, the minima refer to the number of components, and the maxima to the number of holes in two dimensions or voids in three dimensions. We have no theoretical result for the expected number of holes on a three dimensional field.

## 7.2 Simulations

In this section, we will use simulations to investigate if distinctions can be made between various distributions all marginally normalised with the same correlation structure. The motivation is to determine if topological features can distinguish between the fields, where analysis of first and second order properties would not. Also of interest is how the number of components and holes are affected by the underlying correlation structure of the field.

We will begin by introducing the various distributions and parameters that will be considered. Next we will replicate the work on the lattice in [51], showing that there is a difference in number of components and holes when the lattice is of size 256 by 256, before investigating a smaller lattice of similar size to the brain slices. Finally, we will move away from the regular lattice, and see if the results are applicable to irregular lattices such as a brain slice, or a full three dimensional brain.

### 7.2.1 Simulation Parameters

To begin, let us consider the five distributions as described in [51].

1. Let  $z(x)$  be a GRF with  $N(0, 1)$  margins with the Matérn correlation function

$$\text{Corr}(z(x), z(x + d)) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu}|d|/\eta)^\nu K_\nu(\sqrt{2\nu}|d|/\eta) \quad (71)$$

where  $K_\nu$  is a modified Bessel function of the third kind. Note that the Matérn reduces to the exponential correlation function if  $\nu = 0.5$ . This is our first distribution.

2.  $\chi_1^2$ . Let  $z_1$  be a GRF with Matérn correlation then  $z^* = z_1^2$  is  $\chi_1^2$ . Marginally transform using  $z = \Phi^{-1}(F_{\chi_1^2}(z^*))$  where  $\Phi(\cdot)$  and  $F_{\chi_1^2}(\cdot)$  are the  $N(0, 1)$  and  $\chi_1^2$  cumulative distribution functions, respectively to give our second distribution. Hence  $z$  is marginally Gaussian but  $z(x)$  is not a Gaussian random field.
3.  $\chi_3^2$ . Let  $z_1, z_2$  and  $z_3$  be three independent GRFs, each with the same correlation structure. Then  $z^* = \sum_{i=1}^3 z_i^2$  is  $\chi_3^2$ . Marginally transforming using  $z = \Phi^{-1}(F_{\chi_3^2}(z^*))$  leads to the third distribution.
4.  $T_3$ . Use four GRFs with the same Matérn correlation and then

$$z^* = \frac{z_1}{\sqrt{\sum_{i=2}^4 z_i^2/3}}$$

is  $T_3$ . Transform using  $z = \Phi^{-1}(F_{T_3}(z^*))$  to give the fourth distribution.

5.  $F_{3,3}$ . Use six GRFs with the same Matérn correlation and then

$$z^* = \frac{\sum_{i=1}^3 z_i^2/3}{\sum_{i=4}^6 z_i^2/3}$$

is  $F_{3,3}$ . Transform using  $z = \Phi^{-1}(F_{F_{3,3}}(z^*))$  to give the fifth and final distribution.

By construction, the marginal distributions are all  $N(0, 1)$ . The Gaussian Matérn parameters that will be considered in simulations for model 1 are  $(\nu, \eta) = (0.5, 2), (0.5, 5), (1, 5), (0.5, 20)$ . Exponential two was chosen because, as we will see in later sections, this is similar to the correlation on the brain slices. Exponential 20 is chosen so that we can replicate results in

[51], and exponential five is chosen as an intermediate step to see the impact of increasing the value of  $\eta$  from two. Also included is the Matérn with  $\nu = 1$  and  $\eta = 5$ . The parameter  $\nu$  is important for small scale correlations, and increasing  $\nu$  slows down the decay of small scale correlations. For models 2 – 5 parameters (root) will be chosen such that after the transform, the correlations (target) are very similar to those in model 1. The root parameters will be estimated using a numerical search for each of the four sets of parameters.

Table 34 shows an example where the Matérn parameters for model 1 are  $(\nu, \eta) = (0.5, 2)$ . The table shows estimates obtained using a numerical search of the root Matérn parameters required for models 2 – 5 such that the target correlations are very close to those of model 1. Also given is the mean and standard deviation of the correlations at various distances for the various fields using the specified parameters, obtained from 50 simulated fields.

			Distance			
Model	$\nu$	$\eta$	1	$\sqrt{2}$	2	3
Gauss	0.5	2	0.607	0.493	0.368	0.223
$\chi_1^2$	0.8	4	0.618 (0.02)	0.505 (0.03)	0.371 (0.03)	0.220 (0.03)
$\chi_3^2$	0.5	4.4	0.592 (0.02)	0.483 (0.02)	0.367 (0.02)	0.232 (0.02)
$T_3$	0.8	1.9	0.610 (0.02)	0.481 (0.03)	0.341 (0.03)	0.187 (0.03)
$F_{3,3}$	0.6	4.3	0.596 (0.02)	0.480 (0.03)	0.359 (0.03)	0.217 (0.03)

Table 34: The (exact) exponential correlations for the Gaussian field, and the mean and standard deviation of correlations at various lengths from 50 simulations of various fields using the parameters given.

From Table 34 it can be seen that the estimated parameters return correlations which are very close to the ones used on the GRF. Table 35 shows the numerical estimates for the parameters obtained when considering other correlation structures that will be used. In all cases the target correlations are very close to the those for model 1.

Model	$\nu$	$\eta$	$\nu$	$\eta$	$\nu$	$\eta$	$\nu$	$\eta$
Gauss	0.5	2	0.5	5	1	5	0.5	20
$\chi_1^2$	0.8	4	0.8	9.4	1.9	8.8	0.74	41
$\chi_3^2$	0.5	4.4	0.5	11.8	1.8	8	0.54	42
$T_3$	0.8	1.9	0.5	7	1.3	5.5	0.58	22
$F_{3,3}$	0.6	4.3	0.6	10.8	1.1	9.4	0.54	50

Table 35: The numerical estimates for the Matérn parameters for various distributions such that the field correlations are very similar.

### 7.2.2 Images Of The Fields On Two Dimensional Lattices

In Section 7.2.3, simulations using the distributions and parameters stated in Section 7.2.1 will be used to see if any distinctions can be made between the various distributions based on number of components and holes. This will be done for two different sized lattices. One of size 256 by 256 to replicate the work in [51], and the second will be a smaller lattice of size 65 by 65 which contains a similar number of points to that of a brain slice. First, we give examples of the types of pattern that are produced by the different distributions. We concentrated on models 1, 2 and 4, with examples of models 3 and 5 given in the Appendix Section 10.4. Figures 78-81 give two examples for each model, on 65 by 65 lattices, and for the four choices of correlations.



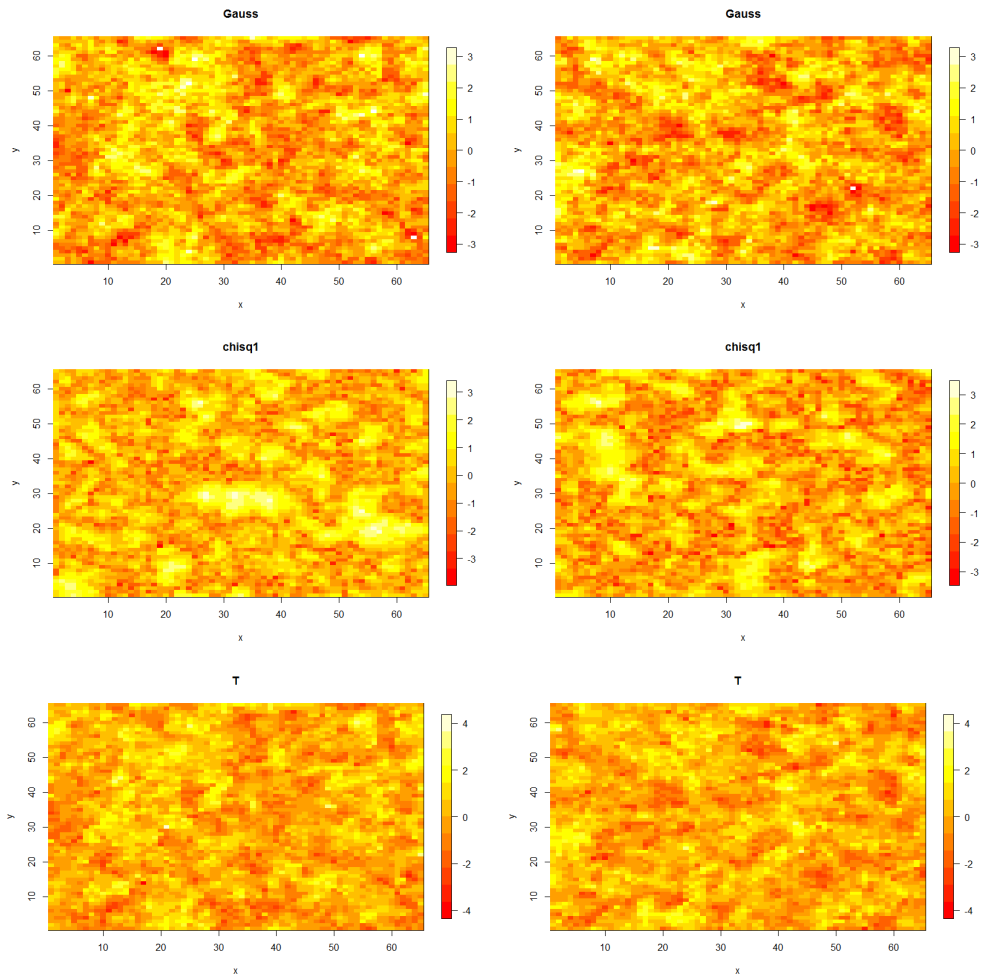


Figure 78: Heatmap images of simulated data from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 2)$ .

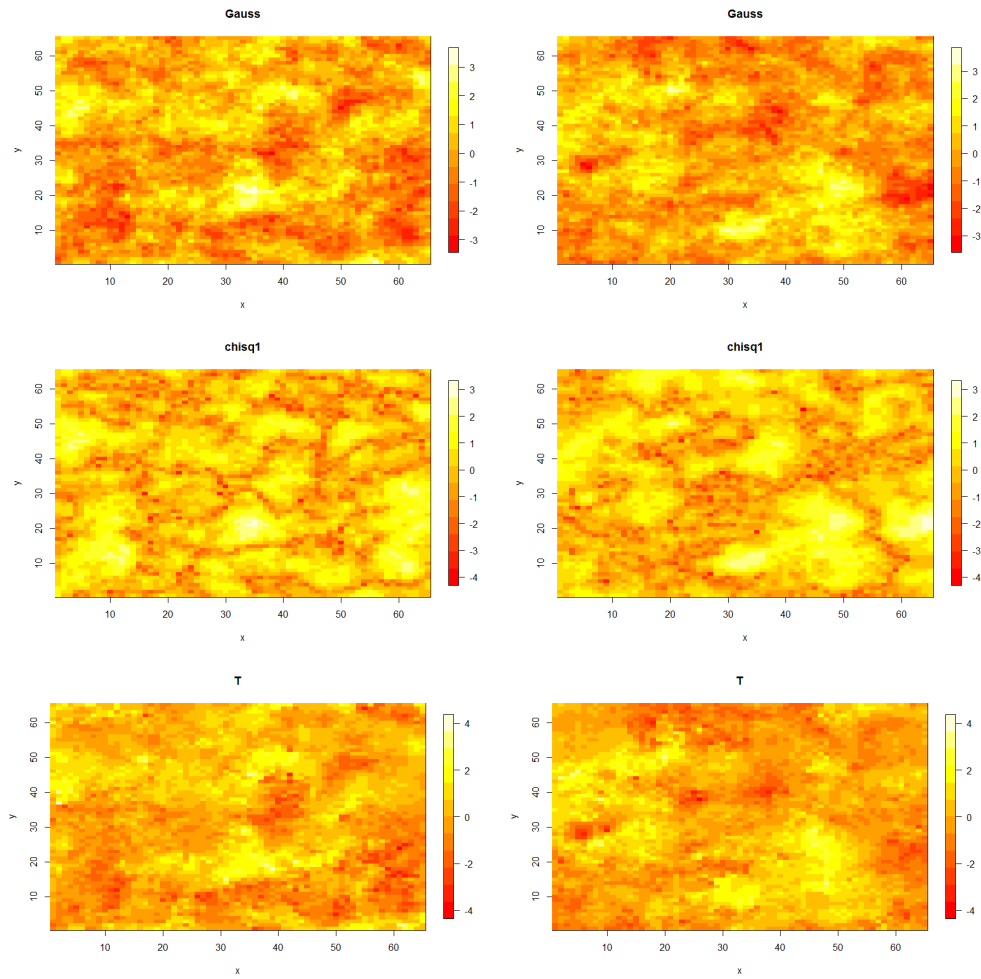


Figure 79: Heatmap images of simulated data from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 5)$ .

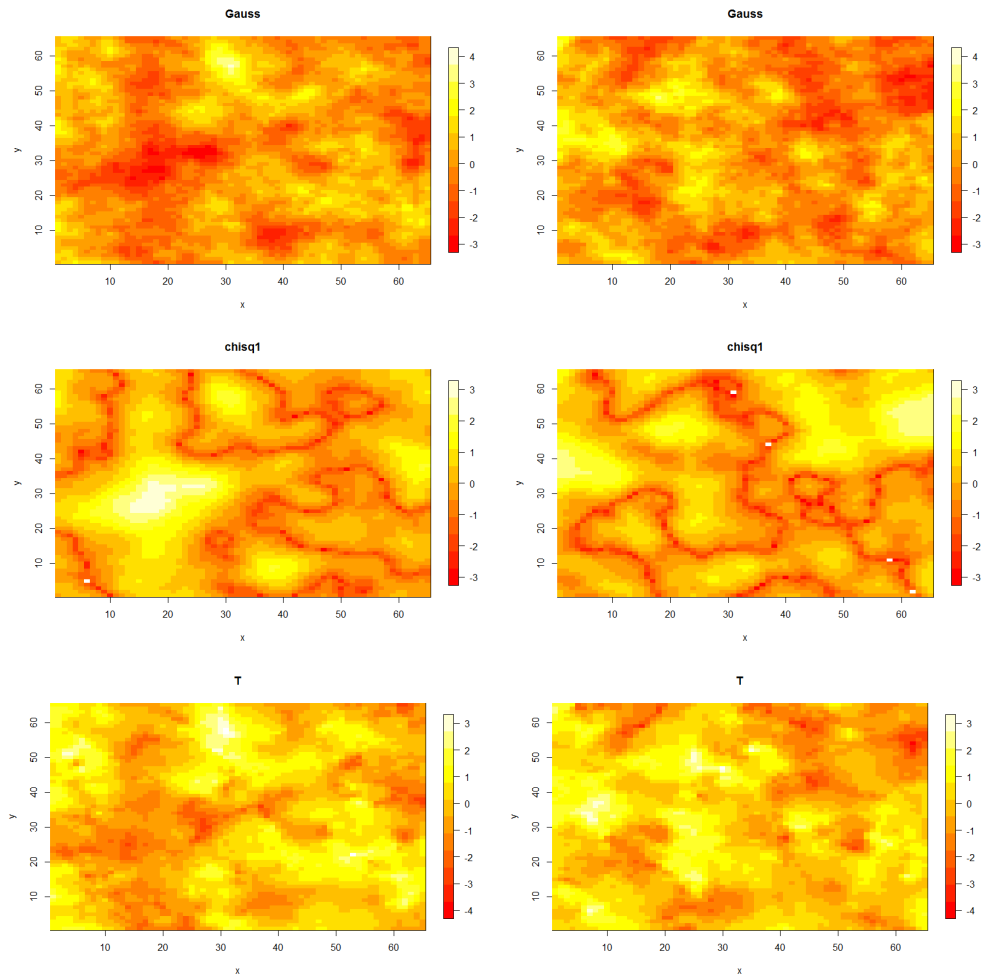


Figure 80: Heatmap images of simulated data from the various distributions with target correlation parameters  $(\nu, \eta) = (1, 5)$ .

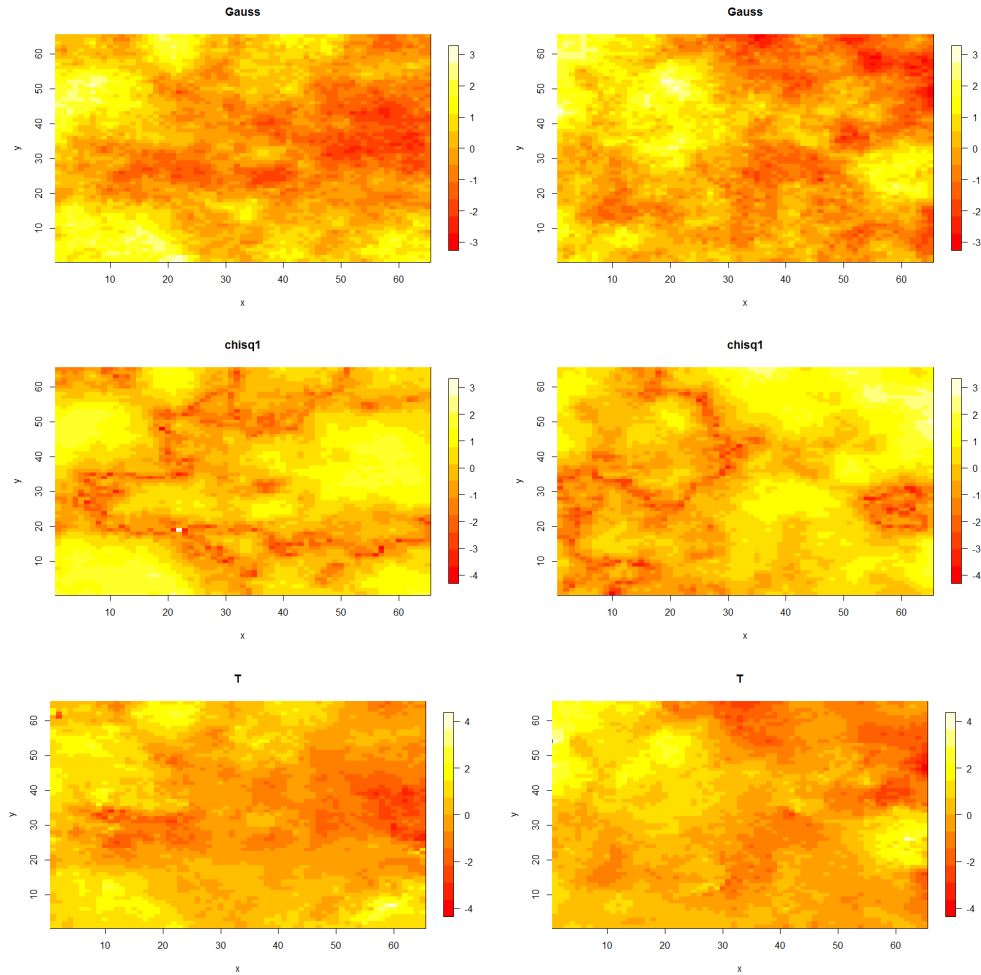


Figure 81: Heatmap images of simulated data from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 20)$ .

From Figures 78 - 81, we can see some of the consequences of changing the correlation parameters. For example, increasing  $\eta$  results in fewer red points (small values), implying fewer components. Also note that for small  $\eta$ , in particular when  $\eta = 2$  there is little visual difference between the fields considered. However as  $\eta$  increases we can see differences. Of particular interest is the case where  $(\nu, \eta) = (1, 5)$ , where the  $\chi_1^2$  simulations are very different from the rest, where long thin lines of small values can be seen in the images. Also note the two simulations on this field are quite different

with respect to the locations of these lines. These images will be referred back to in future sections where some of these interesting features will be explored and explained.

### 7.2.3 Simulations On Regular Two Dimensional Lattices

In this section we will simulate data on lattices to investigate if the expected number of components and holes for a Gaussian field using the theory in Section 7.1 is consistent with simulations, and also to see if distinctions can be made between distributions which are all marginally normal with the same correlation.

Tables 36 and 37 show the mean and standard deviation of the number of components and holes from 50 simulations for each of the five distributions and each of the four correlations considered. Also given in the table is the expected number of components/holes for the Gaussian field and approximations of the standard deviation using the theoretical result. Results are also shown as boxplots in Figures 82-85.

Distribution	$(\nu, \eta) = (0.5, 2)$		$(\nu, \eta) = (0.5, 20)$	
	Components	Holes	Components	Holes
Gauss Theory	9377 (64)	9377 (64)	7786 (66)	7786 (66)
Gauss	9379 (69)	9370 (68)	7782 (72)	7795 (71)
$\chi_1^2$	8884 (96)	7769 (74)	5558 (180)	5179 (124)
$\chi_3^2$	9644 (63)	9356 (71)	7365 (72)	7318 (79)
T	8290 (51)	8291 (62)	6956 (77)	6953 (85)
F	8944 (79)	8961 (53)	7364 (94)	7359 (81)

Table 36: The mean and standard deviation of the number of components and holes from 50 simulations of each of the five different distributions on 256 by 256 lattices.

Distribution	$(\nu, \eta) = (0.5, 5)$		$(\nu, \eta) = (1, 5)$	
	Components	Holes	Components	Holes
Gauss Theory	8371 (65)	8371 (65)	4605 (35)	4605 (35)
Gauss	8377 (68)	8372 (55)	4609 (66)	4604 (58)
$\chi_1^2$	6910 (100)	6025 (77)	2643 (77)	1145 (40)
$\chi_3^2$	8508 (74)	8366 (73)	4250 (65)	4057 (56)
T	8379 (84)	8372 (59)	3209 (49)	3211 (48)
F	7570 (91)	7568 (66)	3526 (74)	3528 (67)

Table 37: The mean and standard deviation of the number of components and holes from 50 simulations of each of the five different distributions on 256 by 256 lattices.

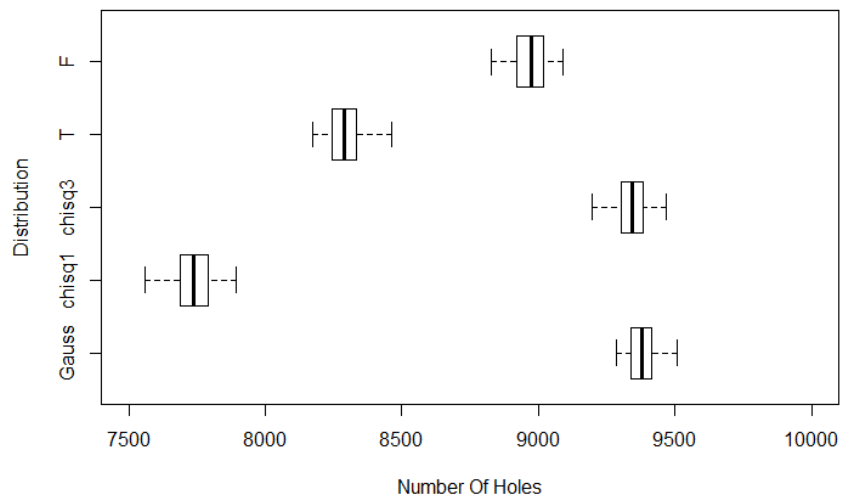
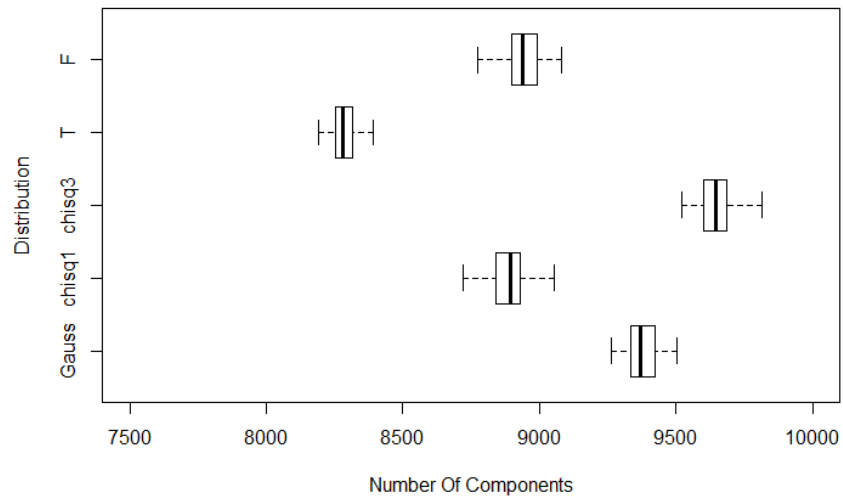


Figure 82: Boxplots of the number of components (top) and holes (bottom) from 50 simulations of each of five distributions on 256 by 256 lattices, all with target correlation  $(\nu, \eta) = (0.5, 2)$ .

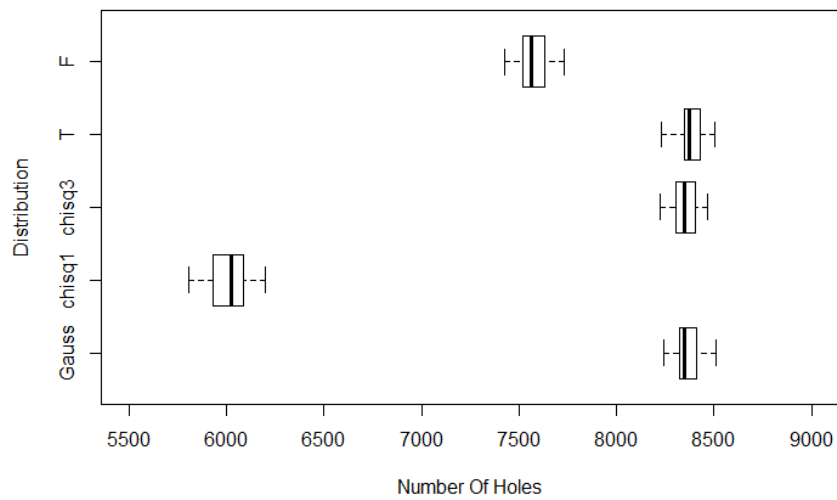
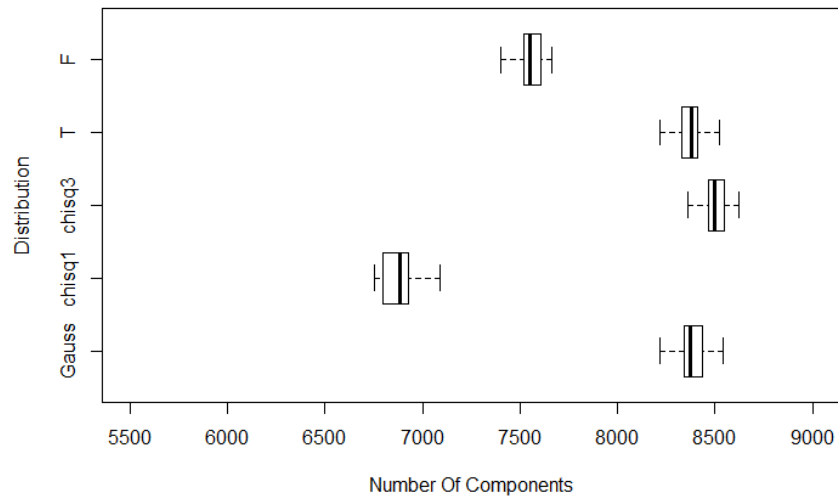


Figure 83: Boxplots of the number of components (top) and holes (bottom) from 50 simulations of each of five distributions on 256 by 256 lattices, all with target correlation  $(\nu, \eta) = (0.5, 5)$ .



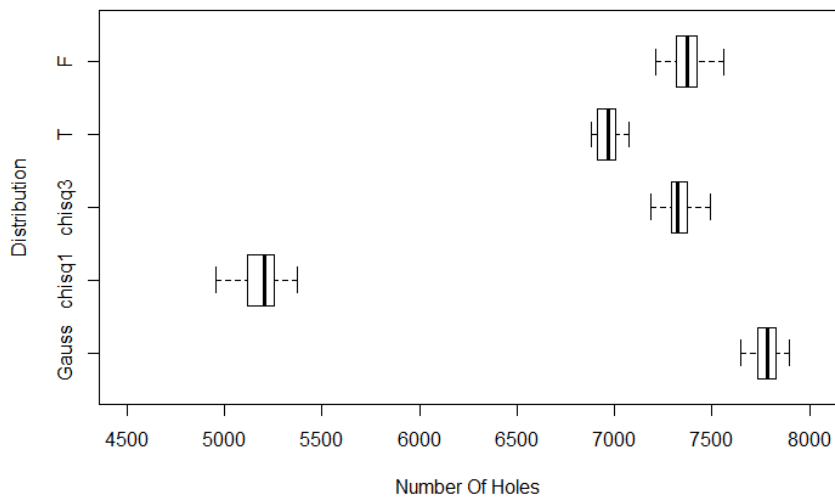
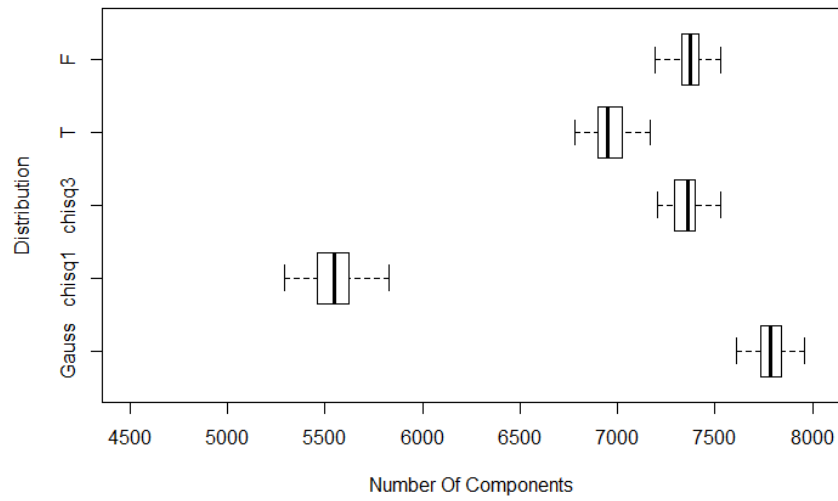


Figure 84: Boxplots of the number of components (top) and holes (bottom) from 50 simulations of each of five distributions on 256 by 256 lattices, all with target correlation  $(\nu, \eta) = (0.5, 20)$ .

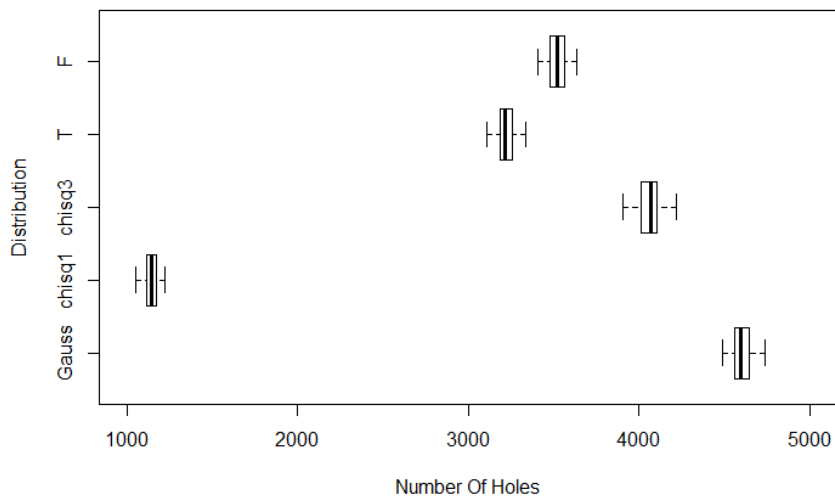
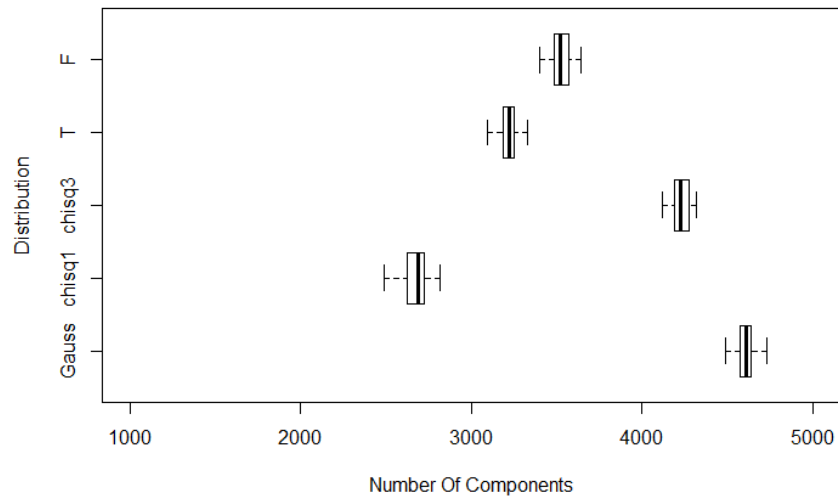


Figure 85: Boxplots of the number of components (top) and holes (bottom) from 50 simulations of each of five distributions on 256 by 256 lattices, all with target correlation  $(\nu, \eta) = (1, 5)$ .

The boxplots show that the number of components and holes on the 256

by 256 lattice can distinguish between the various distributions for some correlation parameters. In particular when the local correlation between points on the lattice is high, for example in the cases where the target parameters are  $(\nu, \eta) = (0.5, 20)$  or  $(\nu, \eta) = (1, 5)$ , the number of components and holes on the Gaussian field is different from all other distributions. However for the other cases, where local correlations are small the differences are not as clear. For the case where  $(\nu, \eta) = (0.5, 2)$  differences between the Gauss and all other distributions can be seen in the number of components, but there is overlap between the Gauss and  $\chi_3^2$  field in the number of holes. Hence there is asymmetry in the number of components and holes for the  $\chi_3^2$  field, and this can also be seen in the  $\chi_1^2$  field. For the case where  $(\nu, \eta) = (0.5, 5)$  distinctions can only be made between the Gaussian,  $\chi_1^2$  and  $F$  fields.

Figure 86 explores this further by fixing the target correlation parameter  $\nu = 0.5$  and investigating the difference in components and holes of the different distributions as  $\eta$  changes.

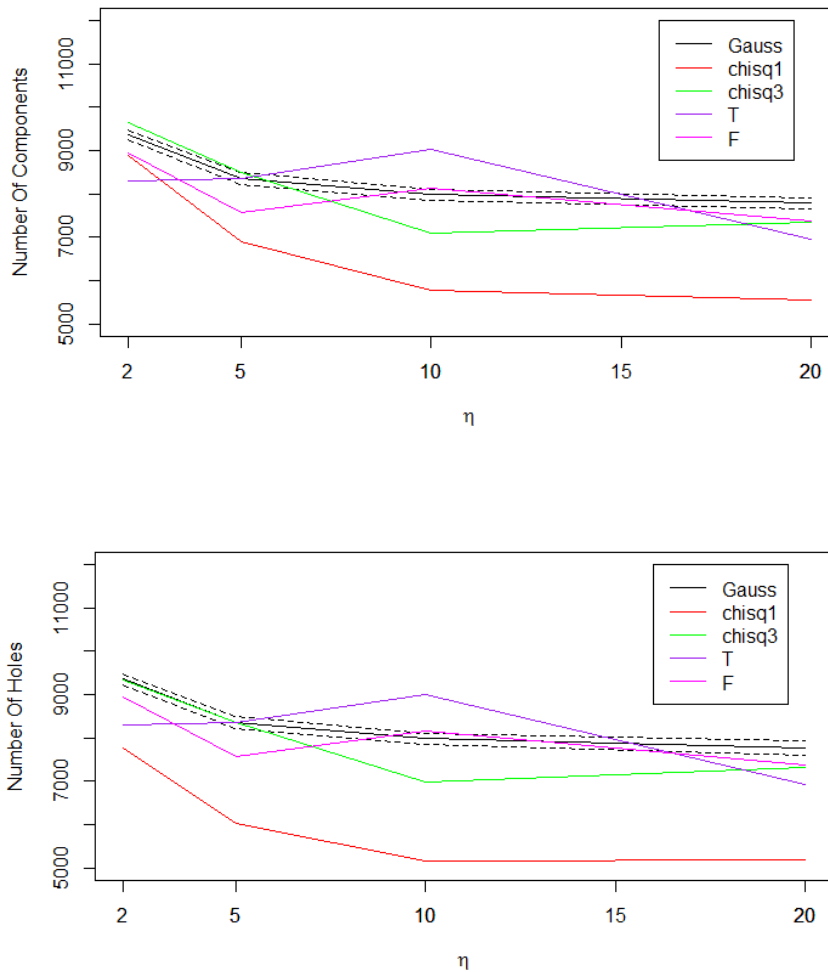


Figure 86: The mean number of components and holes (and two standard deviation interval for Gauss - black dashed), obtained from 50 simulations on a 256 by 256 lattice. The Matérn parameter  $\nu = 0.5$  for the Gaussian field in all cases and  $\eta$  varies. Parameters for other fields were chosen such that all fields have correlation structures that are indistinguishable.

Figure 86 shows that as  $\eta$  increases, the number of components and holes decrease in most cases. This is to be expected, and it will be shown later in Section 7.3.2 (Figure 112) that for a Gaussian field, as lag one correlation

increases the number of extrema decreases. The figure also shows that for  $\eta = 20$  there is a distinction between the Gaussian field and the other fields as shown in [51]. However note that this distinction is present but less extreme as  $\eta$  decreases (and so the small lag correlation decreases and hence there are more extrema). Finally the figure also shows that there is asymmetry in the number of components and holes for the  $\chi_1^2$  field for all values of  $\eta$ . However the difference in number of components and holes on the  $\chi_1^2$  field is much more noticeable when  $\eta$  is small. Here when  $\eta = 2$ , the difference in extrema for the  $\chi_1^2$  field is 1115, and  $\eta = 20$  is 379.

Figure 87 repeats the above, however now all Gaussian fields have Matérn parameter  $\nu = 1$ . The effect of increasing this parameter is to increase the local correlation, hence it can be seen that this leads to a decrease in components and holes.

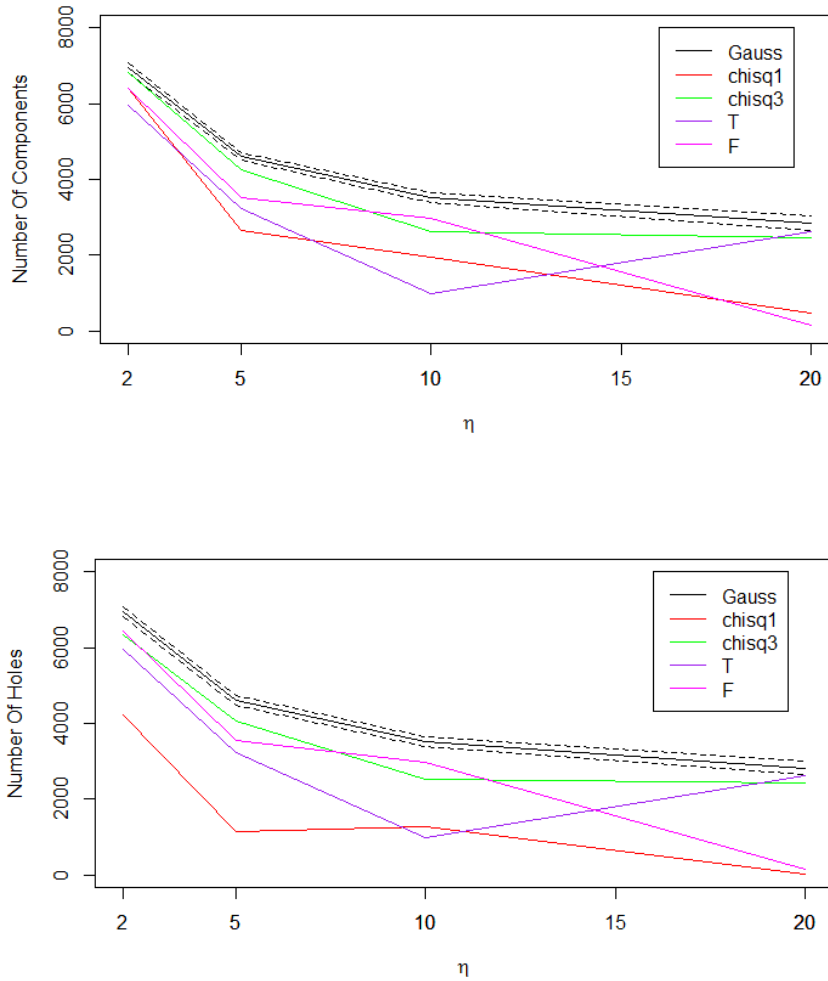


Figure 87: The mean number of components and holes (and two standard deviation interval for Gauss - black dashed), obtained from 50 simulations on a 256 by 256 lattice. The Matérn parameter  $\nu = 1$  for the Gaussian field in all cases and  $\eta$  varies. Parameters for other fields chosen such that all fields have same correlation structure.

Tables 36 and 37, and Figures 86 and 87 show that on the 256 by 256 lattice, the number of components and holes on the Gaussian are significantly different from the rest, except for the case where  $(\nu, \eta) = (0.5, 5)$ , where in

this case the results from the  $T$  distribution is very similar to that of the Gaussian.

The largest brain slice has 4338 points, hence we will repeat the above for a 65 by 65 lattice, to determine if distinctions can still be made on a smaller field. Tables 38 and 39 show the mean and standard deviation of the number of components and holes from 50 simulations for each of the five distributions and each of the four correlations considered. Also given in the table is the expected number of components/holes for the Gaussian field and approximations of the standard deviation using the theoretical result. Results also shown as boxplots in Figures 88-91.

Distribution	$(\nu, \eta) = (0.5, 2)$		$(\nu, \eta) = (0.5, 20)$	
	Components	Holes	Components	Holes
Gauss Theory	616 (17)	616 (17)	513 (17)	513 (17)
Gauss	615 (15)	615 (15)	512 (18)	511 (17)
$\chi_1^2$	580 (16)	513 (18)	358 (31)	340 (24)
$\chi_3^2$	631 (16)	615 (16)	488 (19)	484 (17)
T	544 (15)	547 (14)	462 (19)	458 (16)
F	590 (14)	589 (14)	485 (19)	487 (17)

Table 38: The mean and standard deviation of the number of components and holes from 50 simulations of each of the five different distributions on 65 by 65 lattices.

Distribution	$(\nu, \eta) = (0.5, 5)$		$(\nu, \eta) = (1, 5)$	
	Components	Holes	Components	Holes
Gauss Theory	551 (17)	551 (17)	308 (10)	308 (10)
Gauss	549 (19)	549 (13)	303 (12)	310 (14)
$\chi_1^2$	455 (24)	396 (23)	177 (22)	82 (11)
$\chi_3^2$	558 (19)	552 (18)	287 (16)	276 (15)
T	547 (14)	552 (20)	220 (13)	216 (13)
F	496 (19)	500 (18)	241 (12)	241 (16)

Table 39: The mean and standard deviation of the number of components and holes from 50 simulations of each of the five different distributions on 65 by 65 lattices.

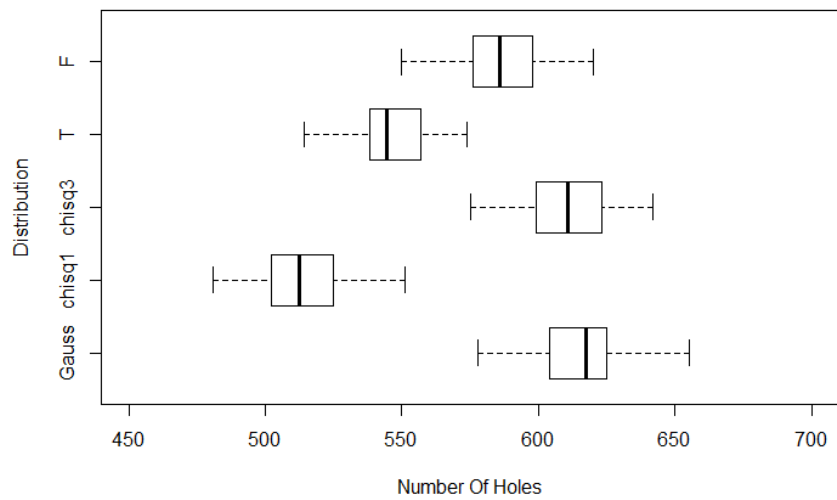
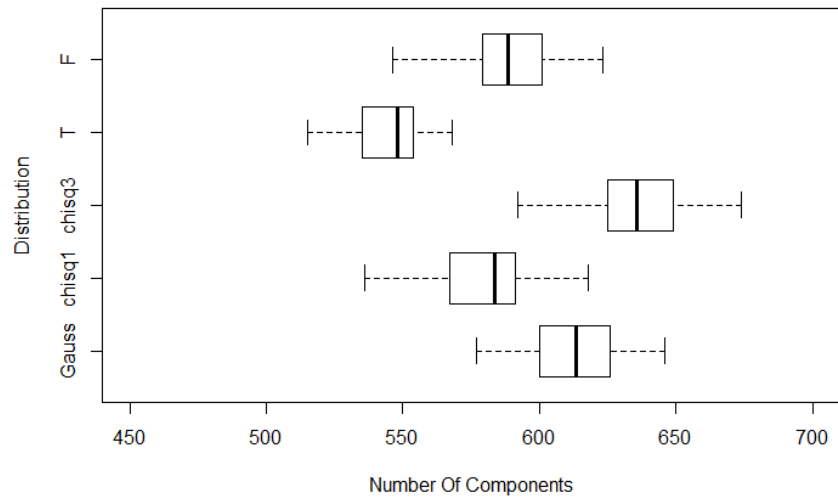


Figure 88: Boxplots of the number of components (top) and holes (bottom) from 50 simulations of each of five distributions on 65 by 65 lattices, all with target correlation  $(\nu, \eta) = (0.5, 2)$ .



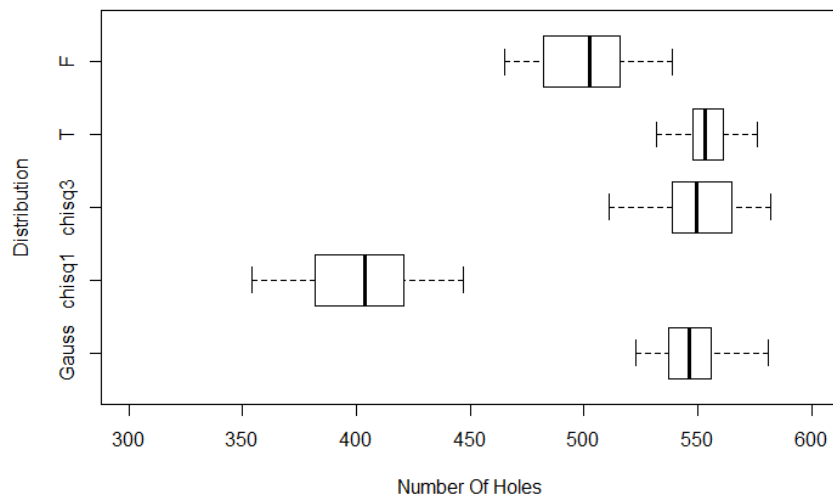
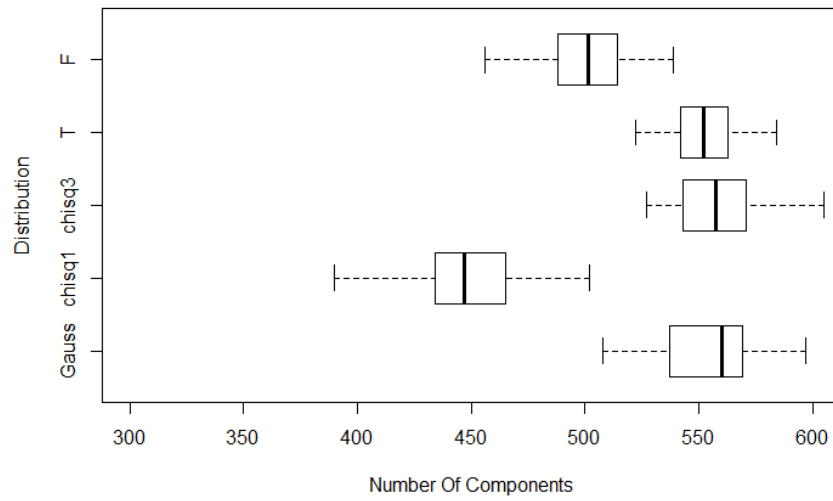


Figure 89: Boxplots of the number of components (top) and holes (bottom) from 50 simulations of each of five distributions on 65 by 65 lattices, all with target correlation  $(\nu, \eta) = (0.5, 5)$ .

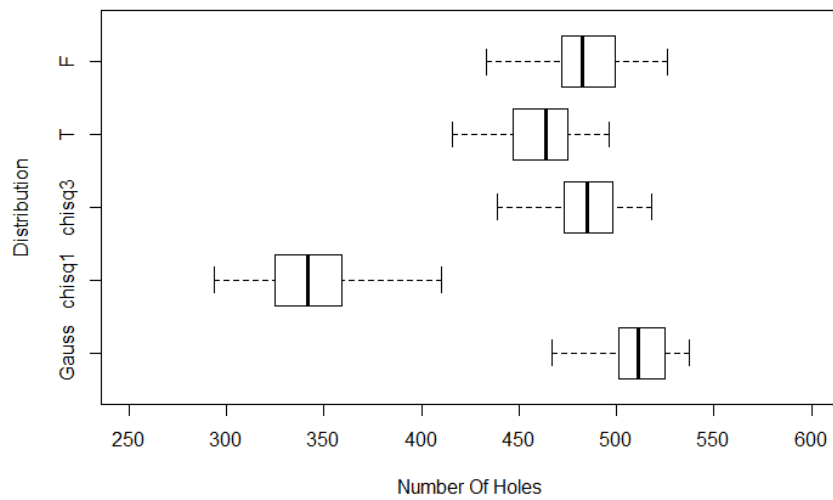
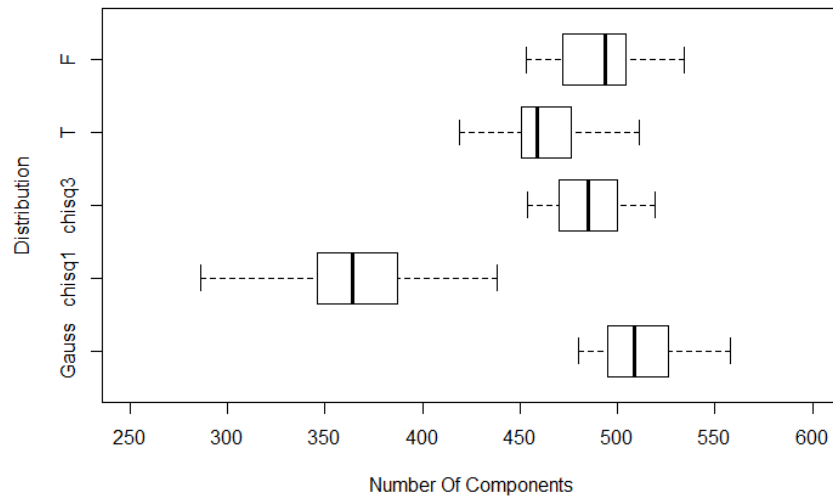


Figure 90: Boxplots of the number of components (top) and holes (bottom) from 50 simulations of each of five distributions on 65 by 65 lattices, all with target correlation  $(\nu, \eta) = (0.5, 20)$ .

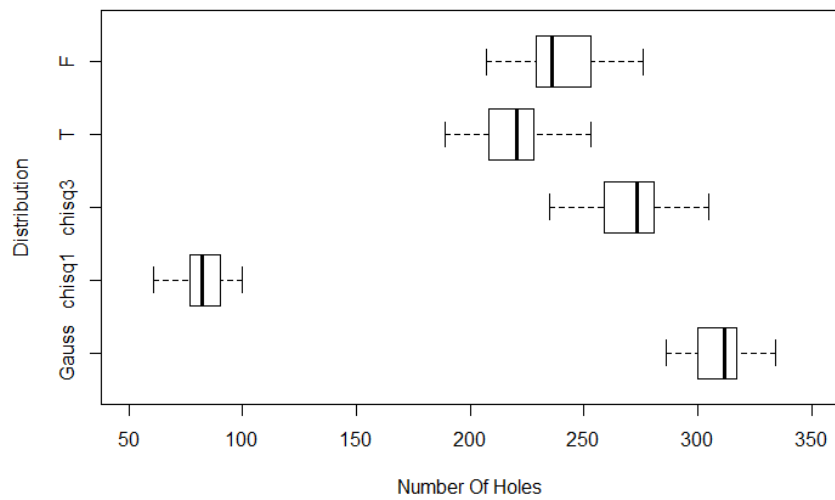
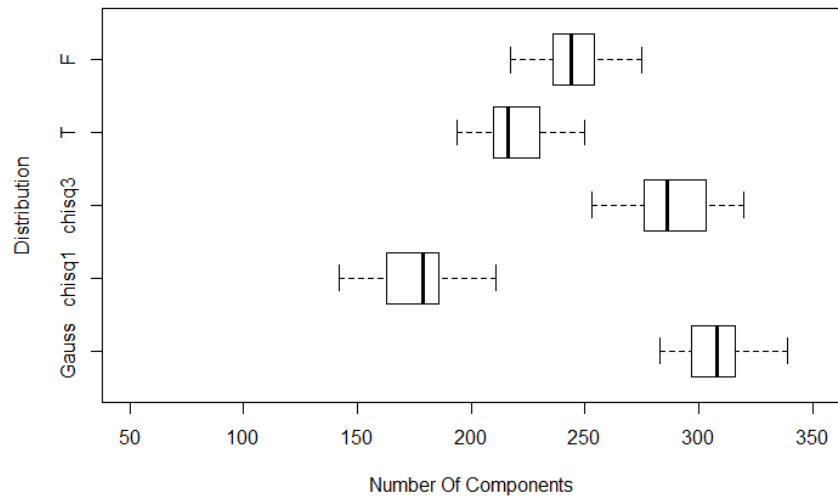


Figure 91: Boxplots of the number of components (top) and holes (bottom) from 50 simulations of each of five distributions on 65 by 65 lattices, all with target correlation  $(\nu, \eta) = (1, 5)$ .

From Tables 38 and 39 it can be seen that the simulations on the Gaussian

field match the expected number, however Figures 88 - 91 show that very little distinction can be made between the Gaussian and other distributions based on number of components and holes.

Figures 92 and 93 explore this further by fixing the target correlation parameter  $\nu = 0.5$  and  $\nu = 1$  respectively and investigating the difference in components and holes of the different distributions as  $\eta$  changes.

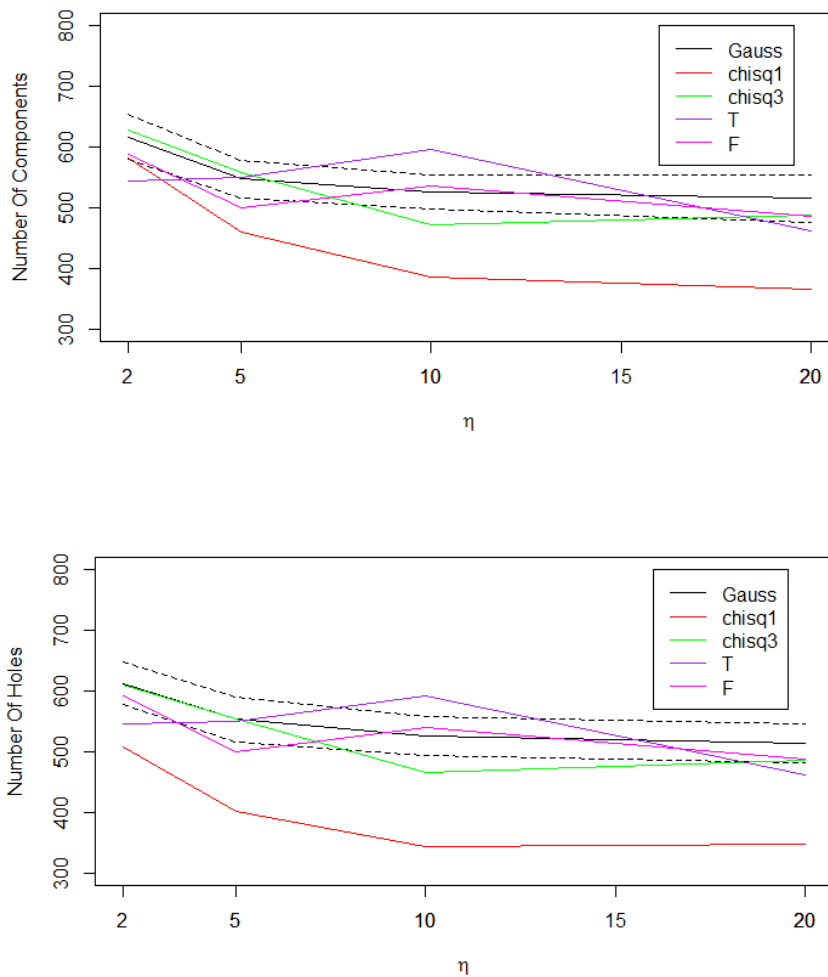


Figure 92: The mean number of components and holes (and two standard deviation interval for Gauss - black dashed), obtained from 50 simulations on a 65 by 65 lattice. The Matérn parameter  $\nu = 0.5$  for the Gaussian field in all cases and  $\eta$  varies. Parameters for other fields are chosen such that all fields have correlation structures that are indistinguishable.

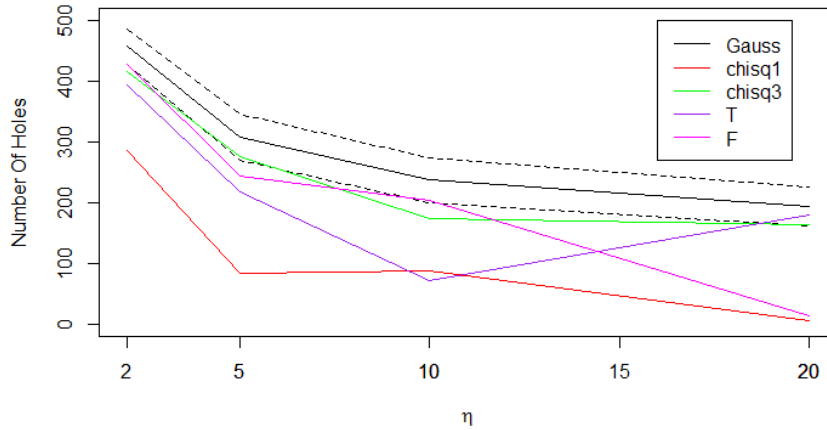
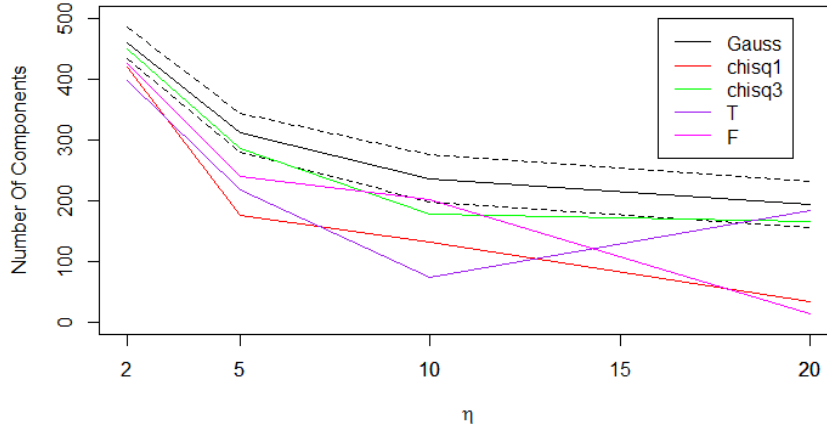


Figure 93: The mean number of components and holes (and two standard deviation interval for Gauss - black dashed), obtained from 50 simulations on a 65 by 65 lattice. The Matérn parameter  $\nu = 1$  for the Gaussian field in all cases and  $\eta$  varies. Parameters for other fields are chosen such that all fields have correlation structures that are indistinguishable.

Figures 92 and 93 again show the asymmetry in components and holes on the  $\chi_1^2$  field in all cases. The figures also show more overlap between the results on the Gaussian and other fields, where at least one distribution in

all plots overlap with the Gaussian, and there are several cases where this happens even as  $\eta$  increases, which is in contrast to the results on the larger lattice. When  $\eta = 2$ , which is similar to that on the brain slice  $z = 42$ , the results for components on the  $\chi_3^2$  and  $F_{3,3}$  fields are within two standard deviations of the results for the Gaussian (when  $\nu = 0.5$ ), and the  $\chi_1^2$  results lie just outside the interval. This was not true for the larger lattice.

### 7.2.4 Simulations On Regular Three Dimensional Lattices

We will now use simulations to determine if distinctions can be made between the distributions on a regular three dimensional lattice of dimension  $60^3 = 216000$  points, similar to that of the full brain. As described in Section 7.1, there is no theoretical result for the expected number of holes in three dimensions, and so we will consider only components and voids which are described by the minima and maxima of the field.

Tables 40 and 41 show the mean and standard deviation of the number of components and voids from 50 simulations for each of the five distributions and each of the four correlations considered. Also given in the table is the expected number of components/voids for the Gaussian field and approximations of the standard deviation using the theoretical result.

Distribution	$(\nu, \eta) = (0.5, 2)$		$(\nu, \eta) = (0.5, 20)$	
	Components	Voids	Components	Voids
Gauss Theory	19622 (116)	19622 (116)	15280 (110)	15280 (110)
Gauss	19619 (113)	19608 (105)	12205 (6170)	11902 (6196)
$\chi_1^2$	19471 (221)	14484 (179)	3855 (1644)	1686 (2029)
$\chi_3^2$	20759 (164)	19360 (144)	6597 (5068)	7315 (5449)
T	16731 (102)	16698 (111)	9127 (4892)	10307 (5741)
F	18556 (144)	18581 (121)	6832 (5504)	6753 (5506)

Table 40: The mean and standard deviation of the number of components and voids from 50 simulations of each of five different 60 by 60 by 60 lattice fields all standard normalised, with Gaussian Matérn parameters specified.

Distribution	$(\nu, \eta) = (0.5, 5)$		$(\nu, \eta) = (1, 5)$	
	Components	Voids	Components	Voids
Gauss Theory	16837 (112)	16837 (112)	7201 (82)	7201 (82)
Gauss	16844 (126)	16824 (127)	7224 (109)	7201 (127)
$\chi_1^2$	14012 (444)	10319 (262)	5656 (400)	870 (56)
$\chi_3^2$	17430 (315)	16846 (236)	6678 (142)	5941 (166)
T	16915 (148)	16926 (137)	4486 (110)	4444 (103)
F	14737 (172)	14698 (172)	5071 (134)	5098 (124)

Table 41: The mean and standard deviation of the number of components and voids from 50 simulations of each of five different 60 by 60 by 60 lattice fields all standard normalised, with Gaussian Matérn parameters specified.

Tables 40 and 41 show that for the simulations with Matérn parameters  $(\nu, \eta) = (0.5, 2)$  or  $(\nu, \eta) = (1, 5)$  distinction can be made between all distributions, either due to significantly different number of components and voids, or due to the asymmetry in the number of components and voids for the  $\chi^2$  distributions. When  $(\nu, \eta) = (0.5, 5)$  there is some overlap between the Gauss and  $T$  distributions.

From the tables it can also be seen that the expected number of components and voids for the Gaussian agree with the simulations, except for the case where  $(\nu, \eta) = (0.5, 20)$ . The simulated figures in the table were obtained using the *RandomFields* package in *R* to simulate data on the lattice. To check the results, an alternative package in *R* was used to simulate the data - *gstat*. Using this package the mean number of components and voids for the case where  $(\nu, \eta) = (0.5, 20)$  is 15366 and 15387 with standard deviations of 177 and 163 respectively (obtained from 50 simulations). These figures are within two standard deviation of the expected value seen in Table 40. This indicates that there is something wrong with the *RandomFields* package in *R*, when  $\eta$  gets large with respect to the size of the lattice. This will be investigated further in Section 7.2.5.

Figures 94 - 91 summarise the above results using boxplots. Note that the boxplot results for the case where  $(\nu, \eta) = (0.5, 20)$  were obtained using the *gstat* package.



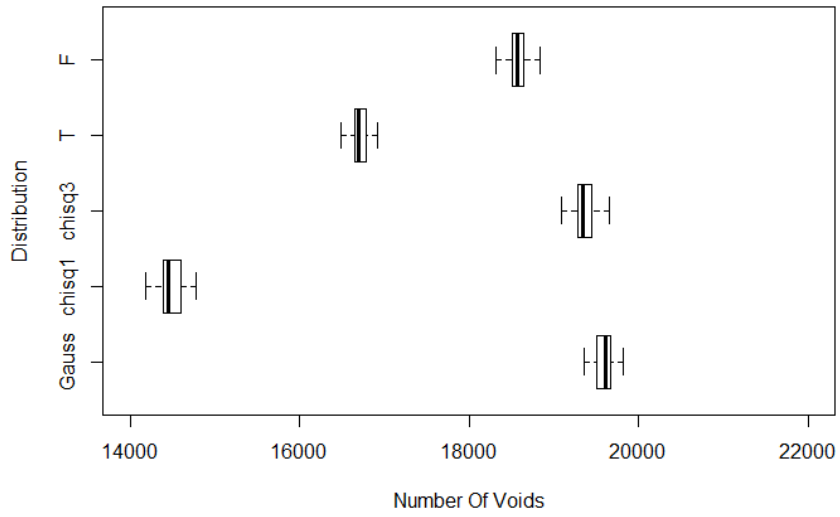
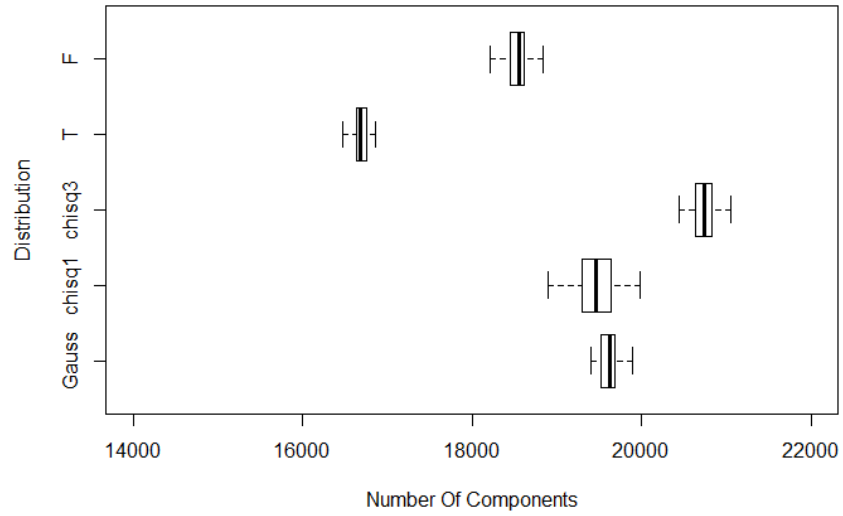


Figure 94: Boxplots of the number of components (top) and voids (bottom) from 50 simulations of each of five distributions on  $60^3$  lattices, all with target correlation  $(\nu, \eta) = (0.5, 2)$ .

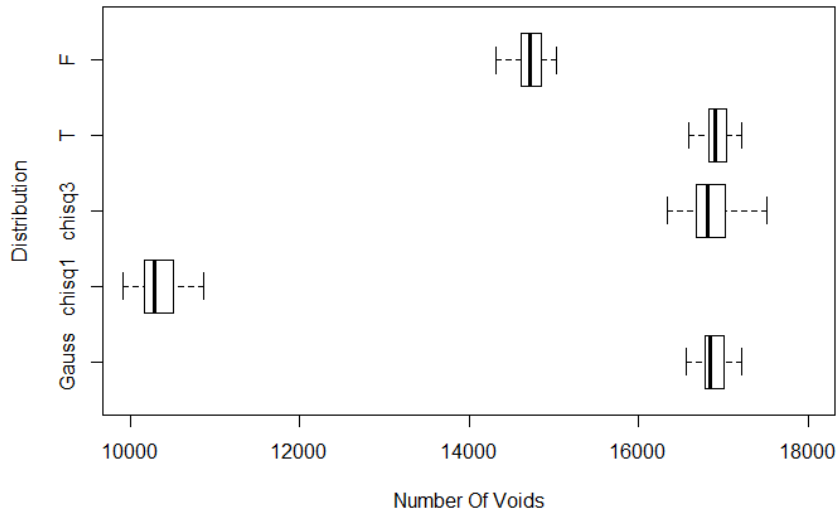
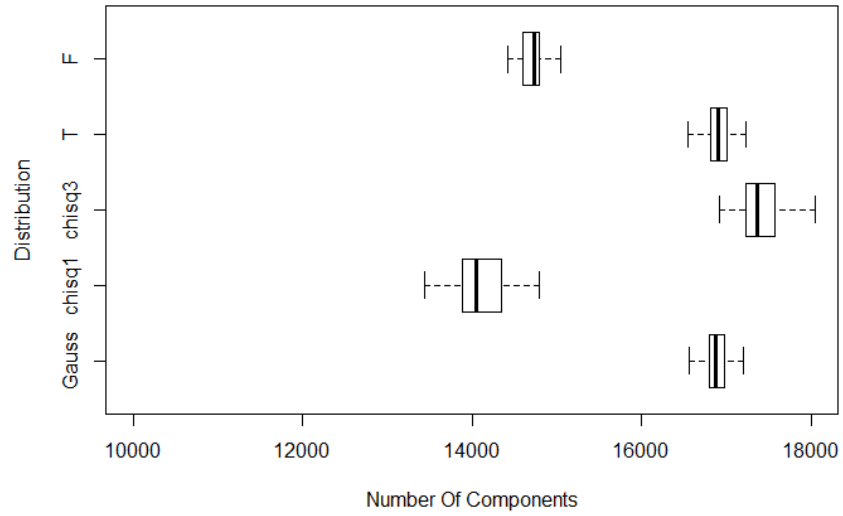


Figure 95: Boxplots of the number of components (top) and voids (bottom) from 50 simulations of each of five distributions on  $60^3$  lattices, all with target correlation  $(\nu, \eta) = (0.5, 5)$ .

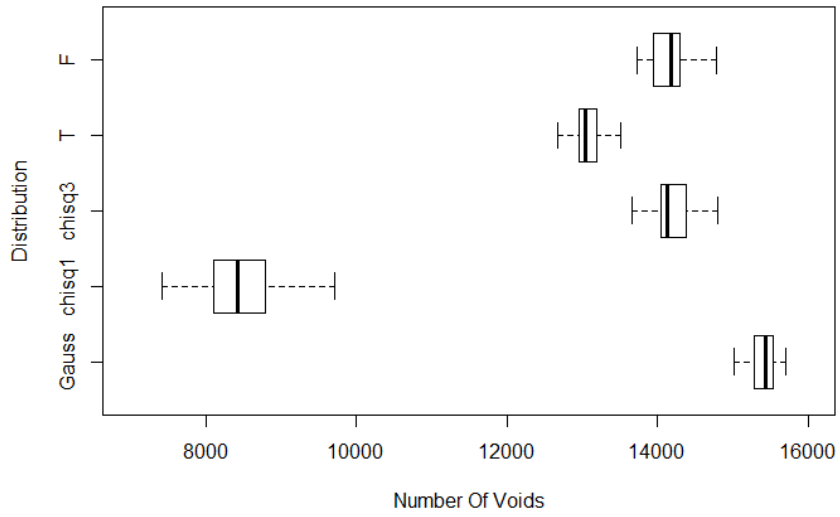
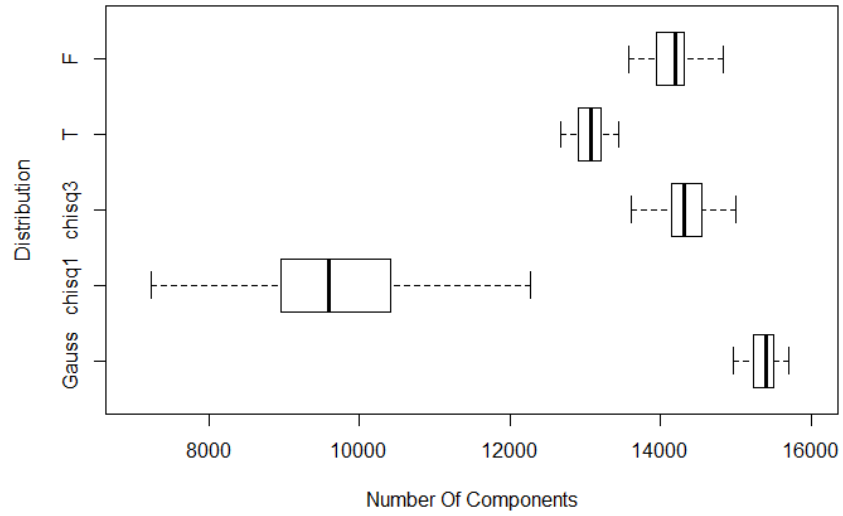


Figure 96: Boxplots of the number of components (top) and voids (bottom) from 50 simulations of each of five distributions on  $60^3$  lattices, all with target correlation  $(\nu, \eta) = (0.5, 20)$ .

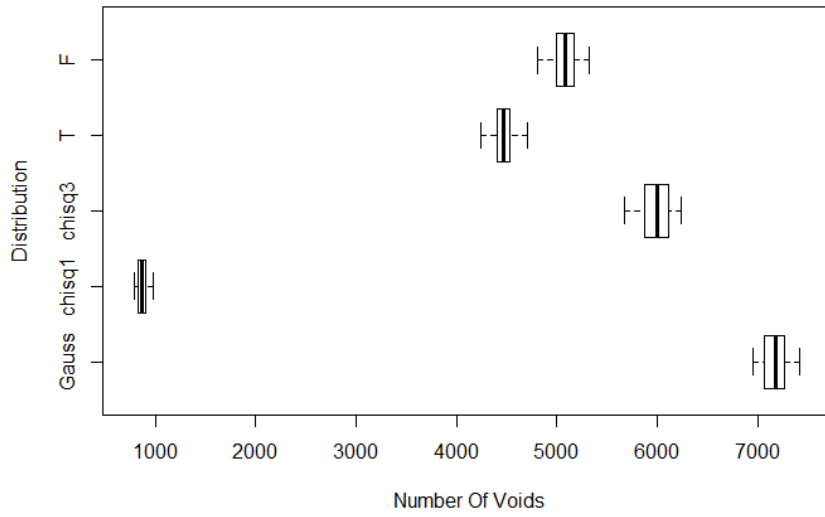
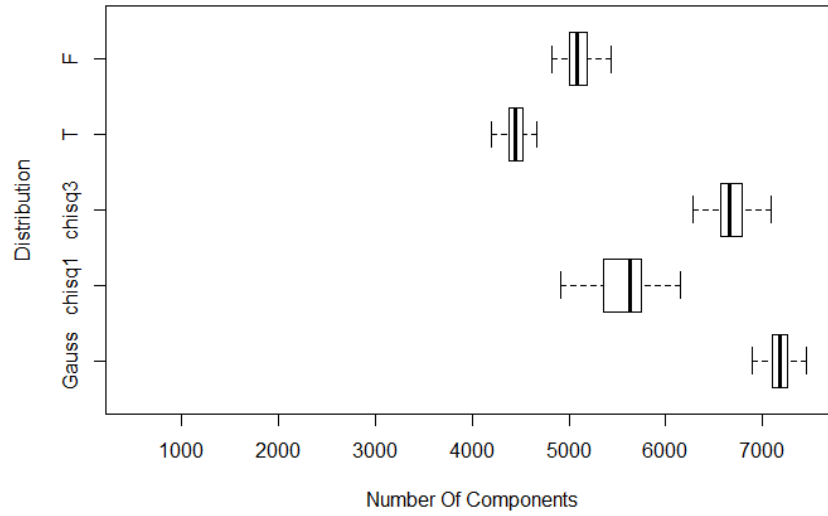


Figure 97: Boxplots of the number of components (top) and voids (bottom) from 50 simulations of each of five distributions on  $60^3$  lattices, all with target correlation  $(\nu, \eta) = (1, 5)$ .

In this section it has been shown that on a three dimensional lattice of similar size to the full brain, distinctions can be made between all distributions either due to significantly different number of components and voids, or due to the asymmetry in the number of components and voids for the  $\chi^2$  distributions. However when  $(\nu, \eta) = (0.5, 5)$  there is some overlap between the Gauss and  $T$  distributions. It has also been found that the *RandomFields* package in *R* produces some inconsistent results when  $(\nu, \eta) = (0.5, 20)$  and this will be further explored in the next section.

### 7.2.5 Investigating Random Fields Package In R

In Section 7.2.4 it was found that the *RandomFields* package in *R*, used for simulating spatial data, did not generate Gaussian fields on a three dimensional square lattice with sides of length 60 units, with the expected number of components and voids when the correlation parameters  $(\nu, \eta) = (0.5, 20)$ . In this section we will investigate this further.

Figure 98 shows the simulated number of components and voids obtained using both the *RandomFields* package, and the *gstat* package against the expected number for varying values of  $\eta$ .

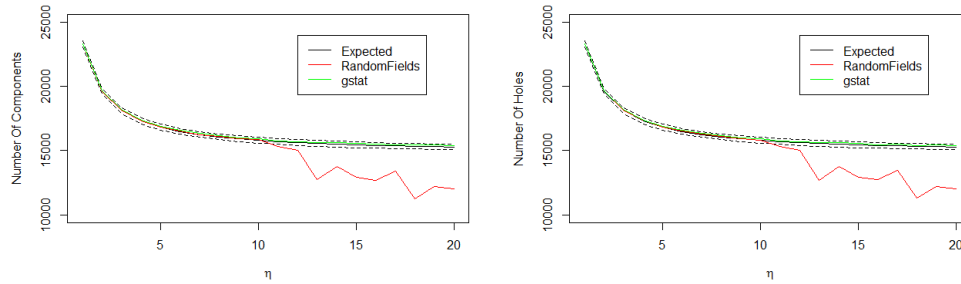


Figure 98: The expected number of components (left) and voids (right) (black solid) plus/minus two standard deviations (black dashed) on a  $65^3$  Gaussian lattice with Matérn correlation parameters  $\nu = 0.5$  and varying  $\eta$ . Also shown is the mean number obtained from 50 simulations using the *RandomFields* package (red) and *gstat* package (green) in *R*.

From Figure 98 it can be seen that both the *R* packages agree with the expected number of components and voids for  $\eta \leq 10$ . As  $\eta$  increases

beyond ten, the number of components and voids obtained by using the *RandomFields* package decreases below two standard deviations of the expected number, however the *gstat* package produces results consistent with the expected number.

Figure 99 shows normal q-q plots of four random realisations on the  $60^3$  lattice when  $(\nu, \eta) = (0.5, 20)$  to check that the *RandomFields* package is generating Gaussian data as expected.

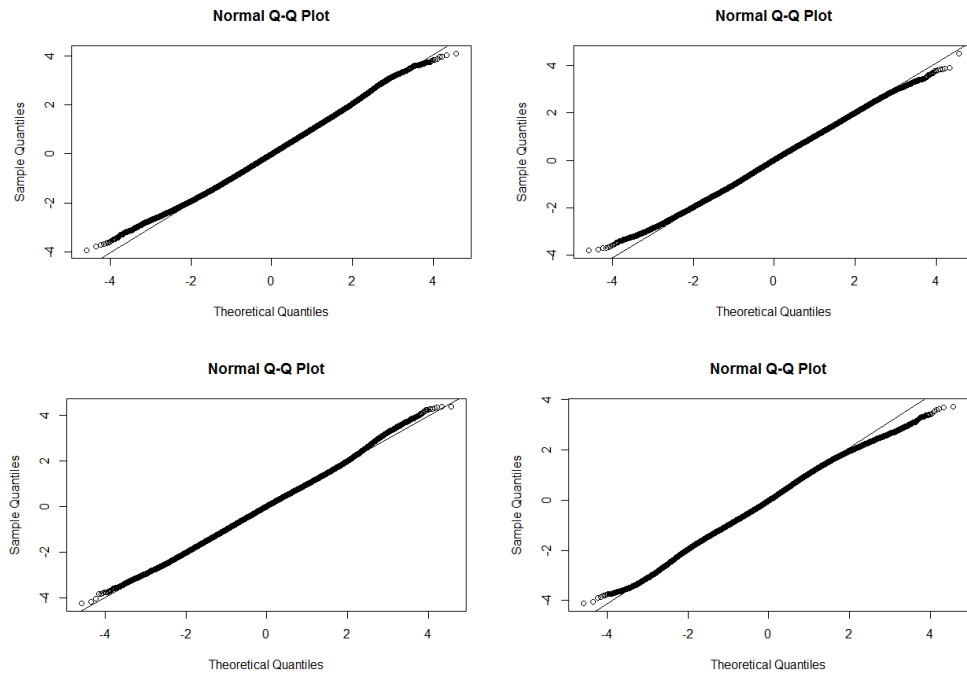


Figure 99: Q-Q plots of four simulations on the  $60^3$  lattice from Gaussian distributions with  $(\nu, \eta) = (0.5, 20)$  using the Random Fields package.

From Figure 99 it can be seen that the data being produced looks Gaussian, however the number of components and voids are not consistent with the expected number (15280) or the number we find using the alternative *gstat* package in *R*. For these four realisations the number of components top left, top right, bottom left and bottom right were 13374, 18976, 26245 and 9794 respectively, with similar numbers for voids. Hence although the data look Gaussian, there is high variability in the number of components and holes as was seen the simulations in Section 7.2.4.

Figure 100 shows the value of  $\eta$  where the number of components on a Gaussian lattice generated using the *RandomFields* package first disagrees with the expected number for different values of  $d$ , the side length of the cube.

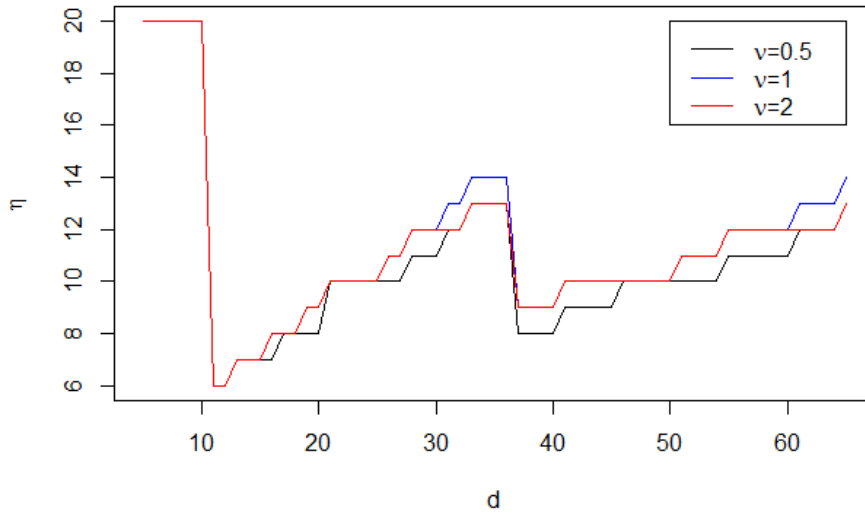


Figure 100: The first value of  $\eta$  when the number of components on a simulated Gaussian  $d^3$  lattice disagrees with the expected number, when using the Random Fields package. Note only values of  $\eta = 1 - 20$  tested.

From Figure 100 it can be seen that for a three dimensional lattice of side length up to ten, the expected number of components on the Gaussian fields generated using the *RandomFields* package agrees with the expected number for all values of  $\eta$  from 1 – 20. However when  $d = 11$ , the number of simulated components diverges from the expected number as soon as  $\eta = 6$ . The maximum value of  $\eta$  for which the *RandomFields* package works as expected increases as  $d$  increases until  $d = 37$  where we see a decrease again. The figure also shows three lines for three different values of  $\nu$  which control the local correlation. It can be seen that the results for all three values of  $\nu$  are very similar, suggesting that it is  $\eta$  which is causing the problem, the long distance correlation. However there does not appear to be any systematic

pattern, ie the maximum value of  $\eta$  is not always increasing with  $d$ , as can be seen by the fact that there is no problem with small cubes, and with the dip at  $d = 37$ .

Similar results were not found in two dimensions. Simulations were carried out for  $\nu = 0.5, 1, 2$  and  $\eta = 1 - 20$  for two dimensional fields with similar numbers of points to the three dimension cases above. For example, in the three dimension simulations, the *RandomFields* package did not compute the correct number of components when  $d = 60$  (216000 points) and  $\eta = 20$ , however for the two dimensional case where  $d = 465$  (216225) points, there were no issues when  $\eta = 20$ .

The source code for all *R* packages are openly available, however no obvious error could be found following inspection of the code.

### 7.2.6 Simulations On Two Dimensional Brain Slices

In this section, we aim to simulate data from the various distributions as in Section 7.2.3 and compare the number of components and holes, however we will now simulate on irregular lattices of the same shape as some brain slices. We also use this section to ensure the theoretical and simulated number of components and holes for a Gaussian field are the same on the irregular lattice. We will focus on three brain slices in this section, those with  $z$  coordinates  $z = 42$ ,  $z = 60$  and  $z = 70$  (see Figure 101). Brain slice  $z = 42$  is approximately midway through the brain, and the other two slices were chosen at random. These choices are arbitrary as there are too many possible two dimensional slices to consider them all.



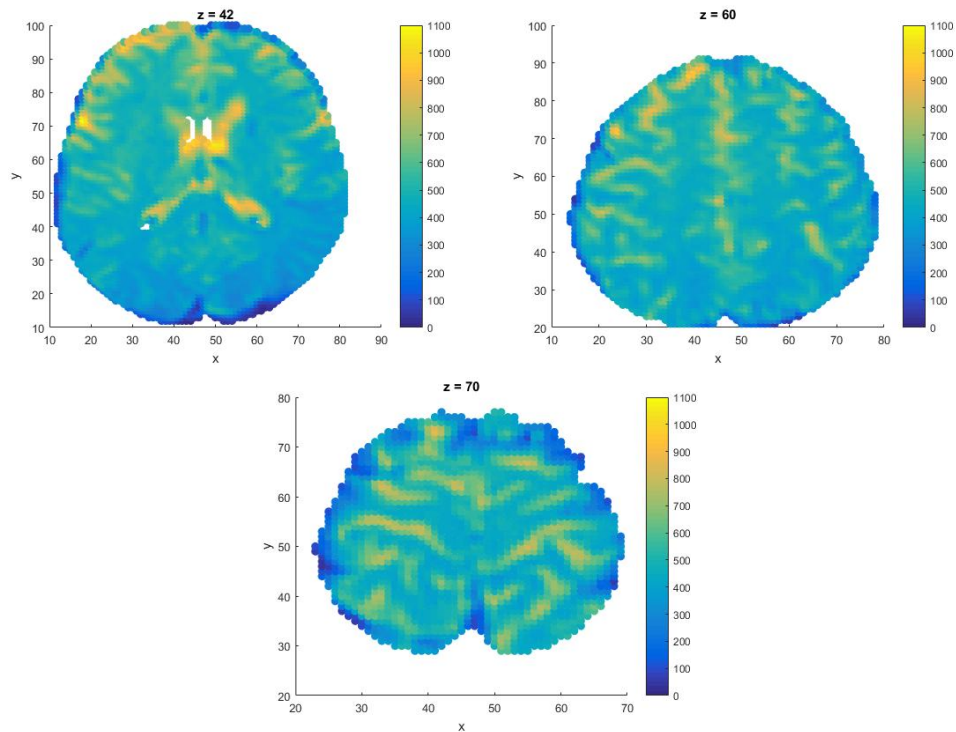


Figure 101: Examples of the fMRI measurements at time  $t=1$  for voxels on three different brain slices.

The theoretical result in Section 7.1 states that the probability of a point being a component depends only on local correlations. Hence to adjust for the irregular shape of the brain, all that is required is to calculate the number of points with the different neighbour configurations (Figure 77) and the lag 1,  $\sqrt{2}$  and 2 correlations. Table 42 shows the counts of the points with different numbers of neighbours for the brain slices seen in Figure 101. As there are two, two-neighbour configurations,  $2_1$  denotes a point where there is a lag of 2 between the neighbours, and  $2_2$  denotes a point where there is a lag of  $\sqrt{2}$  between the neighbours.

Number of Neighbours	0	1	2 <sub>1</sub>	2 <sub>2</sub>	3	4
Count $z = 42$	0	11	0	279	387	3042
Count $z = 60$	0	0	0	137	218	2652
Count $z = 70$	0	3	0	69	115	1446

Table 42: The counts of the number of points with different numbers of neighbours on three different brain slices. 2<sub>1</sub> and 2<sub>2</sub> refer to the two different two neighbour configurations seen in Figure 77.

Tables 43-45 show the results of simulating on the brain slices  $z = 42$ ,  $z = 60$  and  $z = 70$  respectively using the various distributions, with Gaussian Matérn parameters  $(\nu, \eta) = (0.5, 2)$ . In later sections, it will be shown that on the brain slices, the empirical correlation length is close to that of the exponential with correlation length of two, thus only these results are presented here. In each table, the theoretical number is also given in the caption. Figures 102-104 show boxplots of these results.

Distribution	Components	Holes
Gauss Theory	595 (17)	595 (17)
Gauss	599 (15)	602 (15)
$\chi_1^2$	573 (18)	515 (17)
$\chi_3^2$	613 (15)	600 (14)
$T_3$	540 (14)	541 (16)
$F_{3,3}$	578 (17)	579 (17)

Table 43: The mean (and standard deviation) of the number of components and holes from 50 simulations of each of five different fields on brain slice  $z = 42$ , all standard normalised, with target correlation  $(\nu, \eta) = (0.5, 2)$ .

Distribution	Components	Holes
Gauss Theory	463 (15)	463 (15)
Gauss	461 (12)	460 (15)
$\chi_1^2$	435 (14)	390 (15)
$\chi_3^2$	476 (16)	459 (15)
$T_3$	416 (13)	415 (13)
$F_{3,3}$	444 (13)	443 (15)

Table 44: The mean (and standard deviation) of the number of components and holes from 50 simulations of each of five different fields on brain slice  $z = 60$ , all standard normalised, with target correlation  $(\nu, \eta) = (0.5, 2)$ .

Distribution	Components	Holes
Gauss Theory	250 (11)	250 (11)
Gauss	250 (10)	251 (9)
$\chi_1^2$	240 (12)	212 (11)
$\chi_3^2$	256 (12)	250 (12)
$T_3$	224 (10)	226 (8)
$F_{3,3}$	238 (10)	239 (11)

Table 45: The mean (and standard deviation) of the number of components and holes from 50 simulations of each of five different fields on brain slice  $z = 70$ , all standard normalised, with target correlation  $(\nu, \eta) = (0.5, 2)$ .

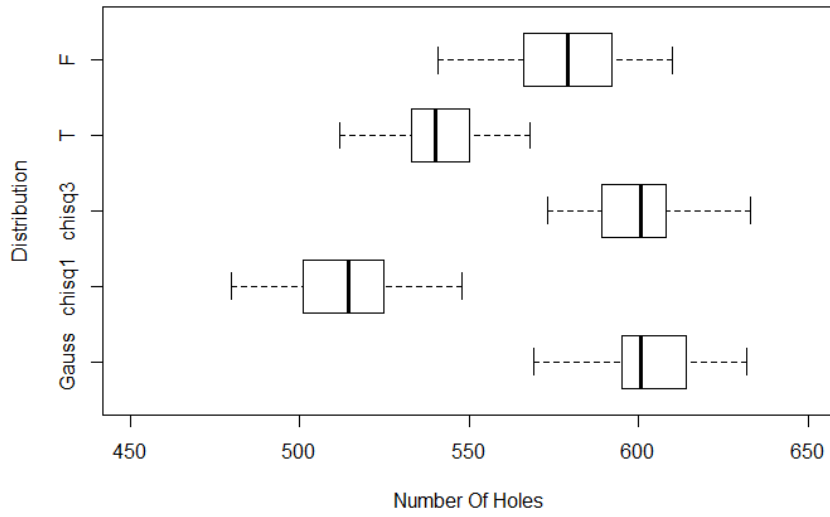
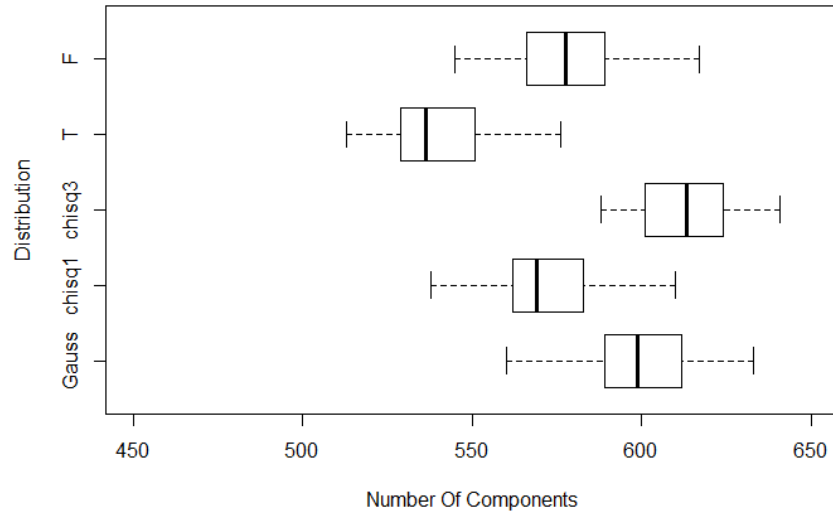


Figure 102: Boxplots of the number of components (top) and holes (bottom) from 50 simulations of each of five distributions on brain slice  $z = 42$ , all with target correlation  $(\nu, \eta) = (0.5, 2)$ .

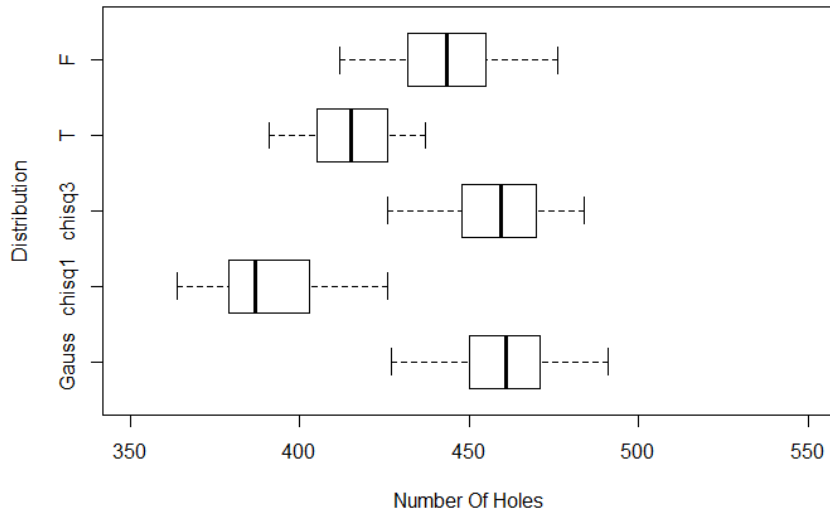
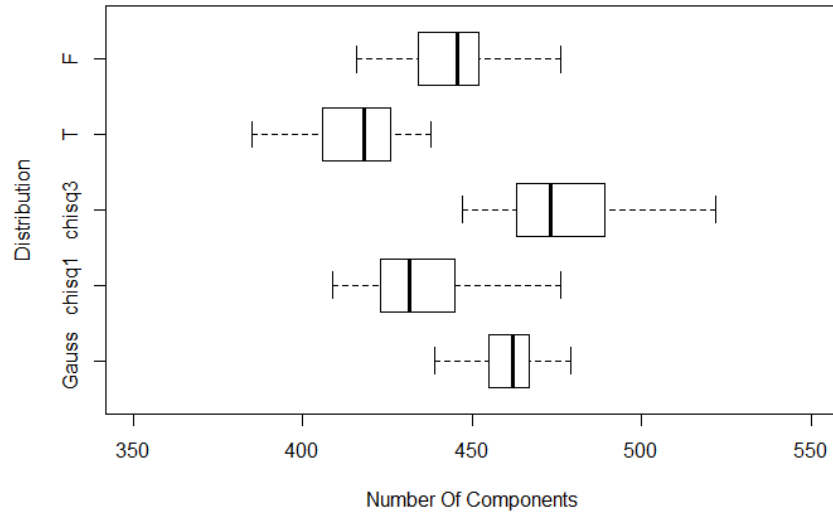


Figure 103: Boxplots of the number of components (top) and holes (bottom) from 50 simulations of each of five distributions on brain slice  $z = 60$ , all with target correlation  $(\nu, \eta) = (0.5, 2)$ .

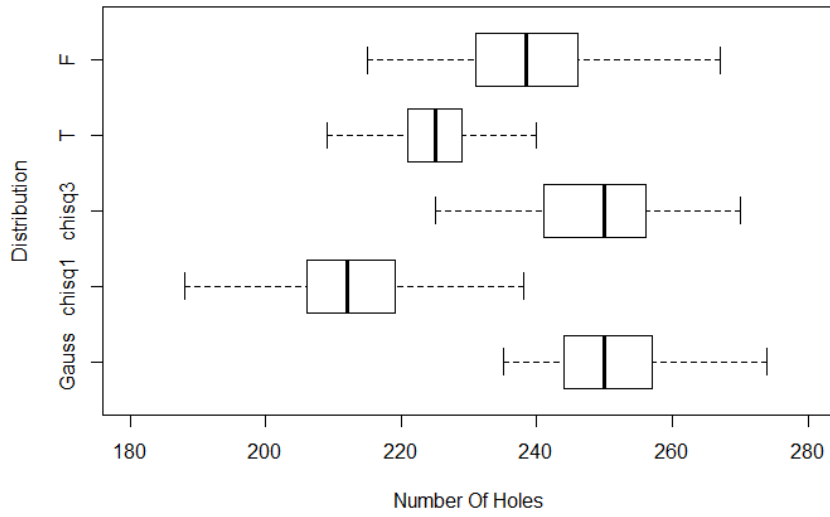
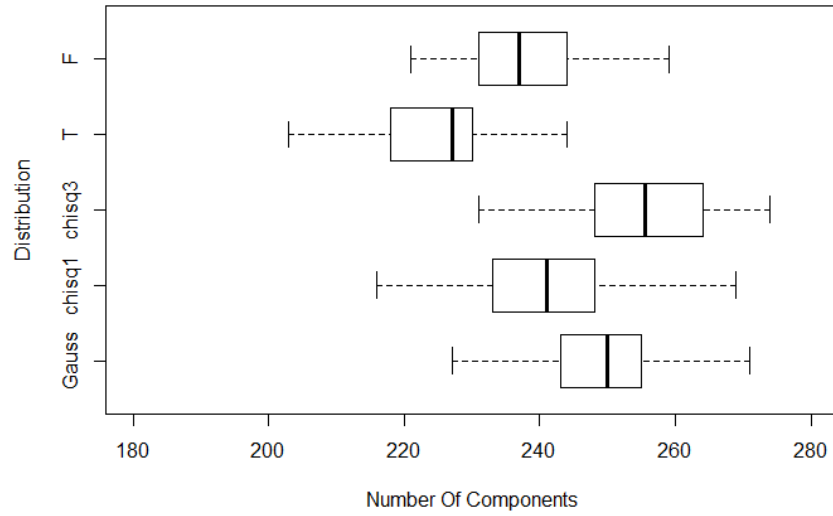


Figure 104: Boxplots of the number of components (top) and holes (bottom) from 50 simulations of each of five distributions on brain slice  $z = 70$ , all with target correlation  $(\nu, \eta) = (0.5, 2)$ .

As in Section 7.2.3, Figures 102-104 shows little distinction between the underlying distributions in terms of number of components and holes. In particular the Gauss,  $T$  and  $F$  distributions overlap in all plots. Also note that the number of components and holes on the  $\chi_1^2$  field are asymmetric as previously seen.

The simulations in Section 7.2.3 and in this section have shown that on smaller fields, we may not be able to distinguish between all the distributions based on number of components and holes. However we have demonstrated that simulating Gaussian data on brain slices with irregular lattice structures give results which are consistent with the expected theoretical number of components and holes.

### 7.2.7 Simulations On Three Dimensional Full Brain

In this section we extend the work in Section 7.2.6 to three dimensional fields. As before, to calculate the expected number of components and voids for the Gaussian field, the local correlations and number of neighbours for each point type is required.

In the three dimensional case, each point can have up to six neighbours. For the isotropic case there are two, two, three and four neighbour configurations. The two two-neighbour configurations are the same as in the two dimensional case in Figure 77. Figure 105 shows the two, three and four neighbour configurations.

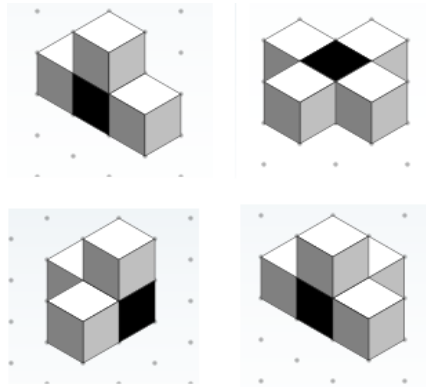


Figure 105: The three dimensional three and four isotropic neighbour configurations.

Table 46 shows the counts of the number of points on the full brain with different neighbour configurations.

Neighbours	0	1	2 <sub>1</sub>	2 <sub>2</sub>	3 <sub>1</sub>	3 <sub>2</sub>	4 <sub>1</sub>	4 <sub>2</sub>	5	6
Count	0	0	0	477	104	6221	6	9053	15578	120756

Table 46: The counts of the number of points with different numbers of neighbours on the full brain. 2<sub>1</sub> and 2<sub>2</sub> refer to the two different two neighbour configurations seen in Figure 77 and 3<sub>1</sub>, 3<sub>2</sub>, 4<sub>1</sub> and 4<sub>2</sub> refer to the two types of three and four neighbour configurations seen in Figure 105.

Table 47 show the simulated number of components and voids on the full brain assuming  $(\nu, \eta) = (0.5, 2)$ .

Distribution	Components	Voids
Gauss Theory	15254 (103)	15254 (103)
Gauss	15264 (90)	15274 (96)
$\chi_1^2$	15020 (155)	11740 (132)
$\chi_3^2$	15991 (118)	15076 (99)
T	13217 (76)	13219 (75)
F	14508 (107)	14522 (110)

Table 47: The mean (and standard deviation) of the number of components and voids from 50 simulations of each of five different fields all standard normalised, with correlation parameters  $(\nu, \eta) = (0.5, 2)$  (Full Brain).

As in Section 7.2.6, Table 47 shows that there is asymmetry in the number of components and voids on the  $\chi_1^2$  field, however it can also be seen that there is now some asymmetry on the  $\chi_3^2$  field also. On the much larger field, it can now be seen that on the full brain distinctions can be made between the Gaussian and the other distributions, either due to different numbers, or the asymmetry of the of components and voids.

### 7.3 Application

The aim of this section is to compare the expected number of components/holes as given in Section 7.1 with the observed number of components and holes on various brain slices and the full brain.



The theoretical result in Section 7.1 states that the probability of a point being a component depends only on local correlations. Hence to adjust for the irregular shape of the brain, all that is required is to calculate the number of points with the different neighbour configurations (Figure 77) and the lag 1,  $\sqrt{2}$  and 2 correlations. In this section, the expected number will be calculated using estimates of the local correlations found by a fitted Matérn, and this will be compared to the theoretical result found when using the empirical local isotropic correlations. A small number of simulations will also be included in this section to show that expected number of components on a Gaussian field with the fitted Matérn correlation obtained from the brain data, is consistent with the simulated Gauss data with the same local correlations. If the actual data is Gaussian and isotropic, then the expected, simulated and actual number of components should be similar.

### 7.3.1 Number Of Components And Holes On Brain Slices

To obtain the expected number of components for the Gaussian field, we will first estimate the local correlation by fitting a Matérn model using least squares to the empirical isotropic correlations. This is time dependent, and this procedure will need to be repeated when considering different time points. Figure 106 shows the fitted Matérn against the actual field correlations for brain slice  $z = 42$  at  $t = 1$ .

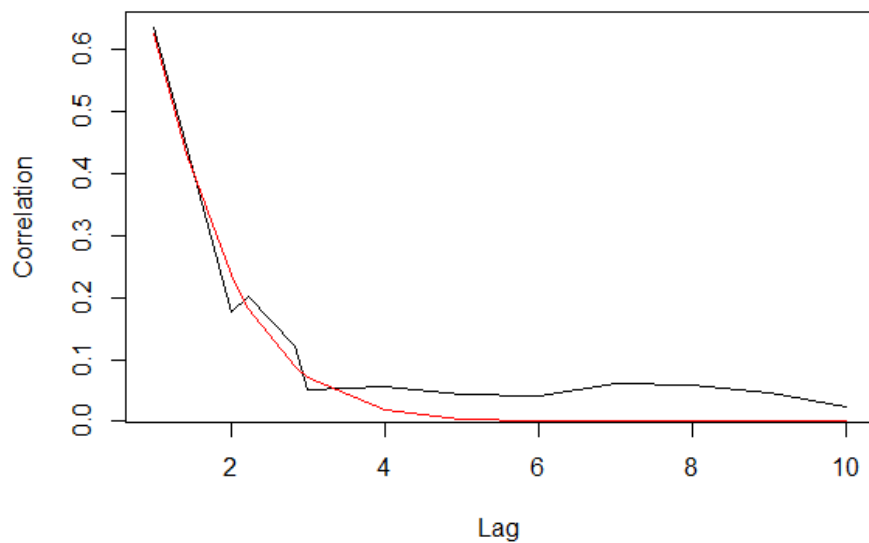


Figure 106: The correlation against lag of the brain slice  $z = 42$  at  $t = 1$  (black), and the fitted Matérn (red).

Note that in Figure 106 the local lags that we are interested in  $(1, \sqrt{2}, 2)$  are fitted well using the Matérn correlation function. Using this Matérn we can now obtain the expected number of components for the Gaussian field. Table 49 shows the results, along with the count of the observed number of components and holes on the brain slice  $z = 42$  at different time points. Also included is the result of 50 simulations to ensure the expected matches simulated using the fitted Matérn correlation.

Time	$E[N]$	$SD[N]$	Sim Min	Sim Max	Obs Min	Obs Max
1	394.1	10.5	401.3 (14.5)	397.4 (13.8)	382	404
11	439.3	11.3	447.1 (12.4)	446.7 (13.2)	436	426
12	418.1	11.4	424.8 (13.8)	424.0 (13.1)	407	387
13	416.9	11.9	423.2 (11.0)	419.8 (14.4)	395	396
14	390.1	11.3	395.7 (11.2)	395.3 (15.2)	405	358

Table 48: The theoretical expected number of components (min) and holes (max) on a Gaussian field (using a fitted Matérn), the mean (and standard deviation) number from 50 simulations, and the observed number for various time points on brain slice  $z = 42$ .

An alternative to fitting a Matérn to the empirical isotropic correlations, is to use the actual empirical correlations in the calculation for the expected number of components. Table 49 shows the results in this case.

Time	$E[N]$	$SD[N]$	Obs Min	Obs Max
1	415.6	15.5	382	404
11	459.1	16.5	436	426
12	417.4	15.5	407	387
13	419.5	15.6	395	396
14	390.1	14.8	405	358

Table 49: The theoretical expected number of components (min) and holes (max) on a Gaussian field (using actual field correlations) and the observed number for various time points on brain slice  $z = 42$ .

Table 48 shows that the calculations for the expected number of components on the irregular field are very similar to the simulated number, when fitting a Matérn to the empirical slice correlations, and also to the actual number on the brain slice. Table 49 shows results for the expected number of components when the empirical correlations are used. From this it can be seen that although there are some differences, due to the fitted Matérn not being exactly the same as the empirical correlation, the expected number is similar to that obtained when using the Matérn fit.

Figures 107 and 108 show the results for all time points.

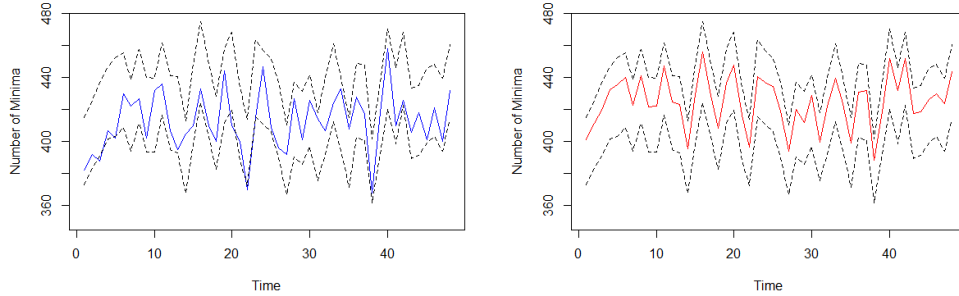


Figure 107: The actual number of components on slice  $z = 42$  (blue), simulated number (red), and expected number plus/minus two standard deviations (black dashed).

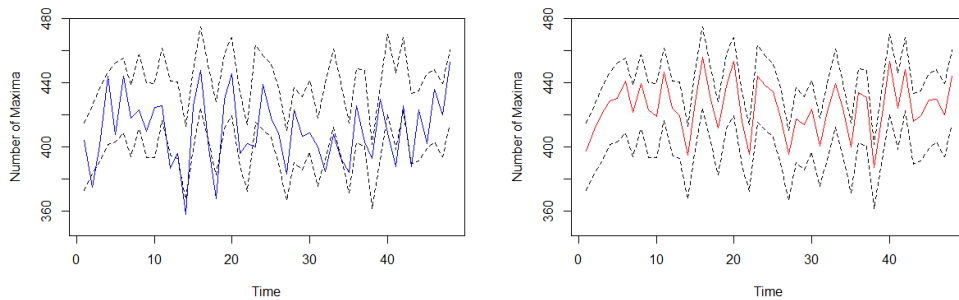


Figure 108: The actual number of holes on slice  $z = 42$  (blue), simulated number (red), and expected number plus/minus two standard deviations (black dashed).

Figures 107 and 108 show that the simulated Gaussian data agree with the theoretical calculation for the irregular two dimensional field shape on the brain slice. It can also be seen that the actual number observed on the brain slice falls within two standard deviations of the expected number at all time points. This indicates that the model residuals at each time point are consistent with a Gaussian Random Field. Although results here are presented only for the brain slice  $z = 42$ , similar results were found for other brain slices.

As noted earlier, an assumption for the expected number of components and holes on the Gaussian field is that the correlations are isotropic. Next, we will investigate this assumption. It will now be allowed that the local correlations depend on direction as well as distance. Considering only lags 1,  $\sqrt{2}$ , 2, six empirical correlations on the brain slices will be calculated. Note here that there is now also an increase in the number of point types  $P_i$ , as orientation as well as number of neighbours needs to be considered. Table 50 shows the values of the local lag correlations for slice  $z = 42$  and  $t = 1$  for both the isotropic and anisotropic cases. In the table the lag subscript refers to orientation of the neighbours. For example,  $2_E$  refers to the lag two correlations in the east/west directions, while  $2_N$  refers to the lag two correlations in the north/south direction.

Lag	Isotropic Correlation	Anisotropic Correlation
$1_E$	0.649	0.635
$1_N$	0.649	0.664
$\sqrt{2}_{SE}$	0.452	0.444
$\sqrt{2}_{NE}$	0.452	0.459
$2_E$	0.212	0.177
$2_N$	0.212	0.250

Table 50: The local correlations on the brain slice  $z = 42$  at  $t = 1$  under the assumptions of isotropy and anisotropy.

Table 51 shows the average correlations over time.

Lag	Isotropic Correlation	Anisotropic Correlation
$1_E$	0.640 (0.04)	0.646 (0.03)
$1_N$	0.640 (0.04)	0.634 (0.05)
$\sqrt{2}_{SE}$	0.446 (0.05)	0.457 (0.06)
$\sqrt{2}_{NE}$	0.446 (0.05)	0.435 (0.05)
$2_E$	0.223 (0.06)	0.213 (0.05)
$2_N$	0.223 (0.06)	0.233 (0.08)

Table 51: The mean (and standard deviation) of local correlations on the brain slice  $z = 42$  over time under the assumption of isotropy and anisotropy.

Table 51 shows that the anisotropic correlations are very similar to the

isotropic correlations, meaning that the field is approximately isotropic, at least locally, and assuming stationarity.

### 7.3.2 Number Of Components And Voids On The Full Brain

In this section, the theoretical number of components and voids for three dimensional parts of the brain will be compared with the observed number in the model residuals. Initially a region of interest will be tested, and then the full brain. Figures 109 and 110 shows the number of components and voids in region seven, the simulated number and the expected interval found by fitting the Matérn to the empirical correlations, which are assumed to be isotropic.

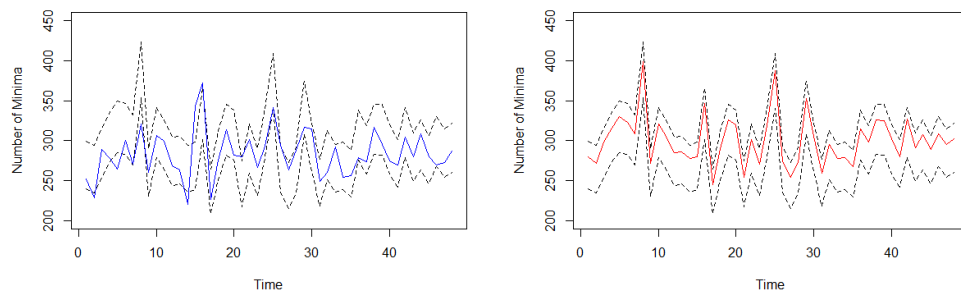


Figure 109: The actual number of components in region seven (blue), simulated number (red), and expected number plus/minus two standard deviations (black dashed).

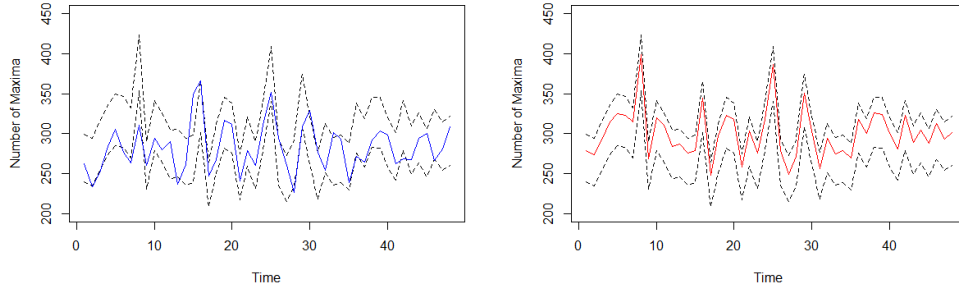


Figure 110: The actual number of voids in region seven (blue), simulated number (red), and expected number plus/minus two standard deviations (black dashed).

Like with the brain slice, Figures 109 and 110 show that the observed number of extrema in region seven is consistent with an isotropic Gaussian field. Similar results were obtained for other regions of interest. Table 52 shows some results for the full brain.

Time	$E[N]$	$SD[N]$	Sim Min	Sim Max	Obs Min	Obs Max
1	9940.8	88.8	9933.7 (78.2)	9959.4 (79.8)	6634	6702
11	9983.2	87.6	10001.8 (67.6)	10002.9 (65.4)	8347	8003
12	9809.7	88.0	9841.4 (61.1)	9827.9 (73.0)	7391	7437
13	8004.3	81.0	7956.44 (236.7)	7963.2 (216.1)	6701	6314
14	8395.2	83.1	8407.7 (83.1)	8405.3 (88.2)	6699	6691

Table 52: The theoretical expected number of components (min) and voids (max) on a Gaussian field (using a fitted Matérn), the mean (and standard deviation) number from 50 simulations, and the observed number for various time points on the full brain.

From Table 52 it can be seen that the simulated agree with the theoretical number of extrema on the full brain when fitting an Matérn to the empirical isotropic correlations, however these figures do not match with the observed number found on the field. Table 53 shows the expected number of extrema using the empirical correlations found in the full brain, assuming stationarity and isotropy.

Time	$E[N]$	$SD[N]$	Observed Components	Observed Voids
1	6357.0	67.8	6634	6702
11	7789.5	70.4	8347	8003
12	7028.0	69.3	7391	7437
13	4773.5	62.2	6701	6314
14	6557.5	68.3	6699	6691

Table 53: The theoretical expected number of components and voids on a Gaussian field (using empirical correlations) and the observed number for various time points on the full brain.

From Table 53 it can be seen that the expected number of extrema has reduced substantially in comparison to the expected number found using a fitted Matérn to the correlations. Note that at  $t = 13$  the expected number has fallen far more than the others, and this is due to an increase in lag one correlation at this time point (see Figure 111 below). The expected numbers now fall slightly under the number observed on the brain. Figure 111 compares the fitted correlations using the Matérn and the empirical correlations.



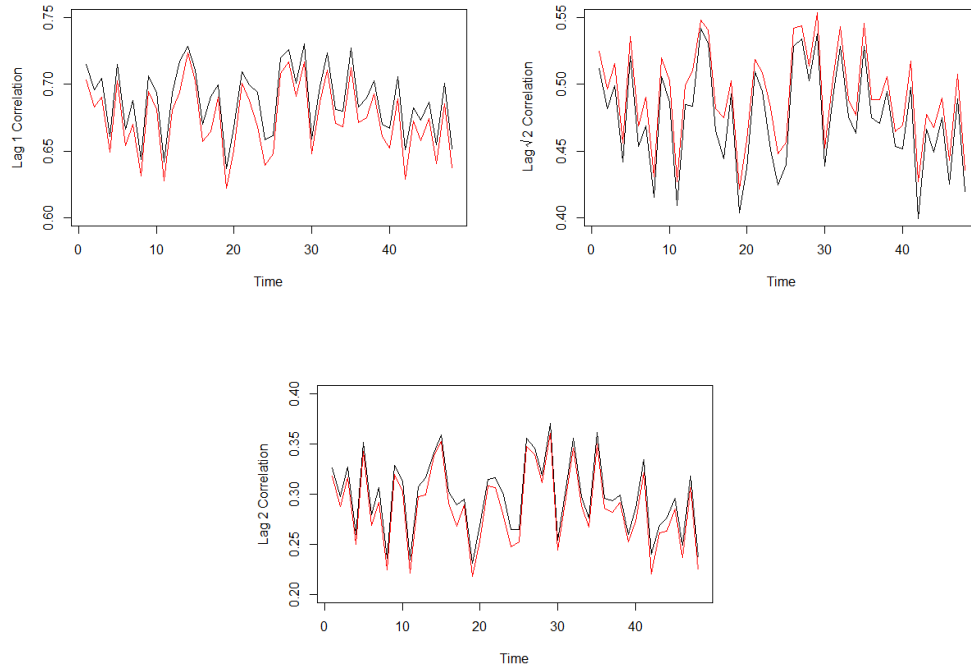


Figure 111: The empirical lag 1,  $\sqrt{2}$ , 2 correlations (black) and the fitted Matérn correlations (red) for the full brain.

From Figure 111 it can be seen that the Matérn fitted correlations are very similar to the empirical ones. Therefore to explore why the Matérn fit fails, we will now explore the sensitivity of the theoretical result to small changes in local correlation. Figure 112 shows the effect of reducing the correlation at local lags on the expected number of maxima on the brain slice.

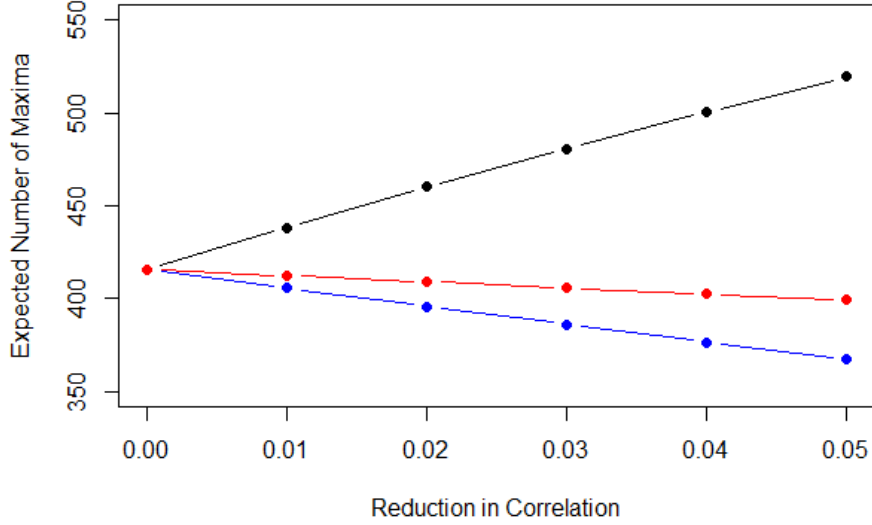


Figure 112: The expected number of maxima on the brain slice  $z = 42$  at  $t = 1$  when reducing each of the lags independently (lags 1,  $\sqrt{2}$ , 2 - black, blue, red).

From Figure 112 it can be seen that if the lag one correlation reduces, the number of maxima increases, however if the lag  $\sqrt{2}$  and 2 correlation decreases, so does the number of maxima. Note also that the reduction in local correlations in Figure 112 are very small and can change the number of maxima significantly. Returning to Figure 111 it can be seen that the fitted Matérn lag one correlations are always slightly lower than the empirical ones, and the  $\sqrt{2}$  fitted correlations are always higher than the empirical ones. This, combined with the much larger number of points on the full brain, explains why the expected number of extrema on the full brain is much higher when using the fitted values.

To explain why this happens, consider the probability of a point in the field being a maxima as in Equation (67). The probability is calculated by evaluating the zero mean,  $k$ -dimensional cumulative normal distribution at  $0_k$ , where the variance is given by

$$V(r_1, r_{\sqrt{2}}, r_2) = R + 1_k 1_k^T - r 1_k^T - 1_k r^T \quad (72)$$

where  $r_1, r_{\sqrt{2}}, r_2$  are the lag 1,  $\sqrt{2}$  and 2 correlations respectively. Recall that  $R$  is a  $k$  by  $k$  matrix of the correlations between the neighbours of the point of interest, and  $r$  is the  $k$ -vector of correlations between the point of interest and its immediate neighbours. Therefore,  $R = R(r_{\sqrt{2}}, r_2)$  depends only on correlations of lag  $\sqrt{2}$  and 2, and  $r = r(r_1)$  depends only on lag 1 correlations. Hence from (72) it can be seen that for fixed lags  $\sqrt{2}$  and 2, and so fixed  $R$ , a increase in lag 1 correlation (increasing  $r$ ) will reduce all elements of  $V$ , in particular the off diagonal elements. These represent the correlation between the variables of the multivariate normal distribution and so by considering the contours of a multivariate normal, it can be deduced that a reduction in these correlations will decrease the probability of obtaining a maximum. Thus if lag 1 correlation increases, the probability of obtaining a maximum decreases, as seen in Figure 112. To highlight this using an example, consider the case of a point  $z_1$  with two neighbours,  $z_2$  and  $z_3$ , each of distance one from  $z_1$ . In this case, the probability of  $z_1$  being a maximum is determined using the cumulative bivariate normal distribution as given in Equation (67) with  $k = 2$ , evaluated at zero. Suppose  $R$  is fixed such that there is independence between the neighbours of  $z_1$ , and let us consider three cases for different lag 1 correlations. If  $r_1 = 0$ ,  $r_1 = 0.5$  and  $r_1 = 0.7$  then we have

$$V(0, 0, 0) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, V(0.5, 0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, V(0.7, 0, 0) = \begin{pmatrix} 0.6 & -0.4 \\ -0.4 & 0.6 \end{pmatrix}.$$

Here it can be calculated that the probability of  $z_1$  being a maximum is 0.33, 0.25 and 0.13 respectively. Hence increasing lag 1 correlation reduces the probability of a maximum. This can also be seen by observing the contours of the bivariate normal distribution in these cases. Figure 113 shows the three cases, and note that the probability of a maximum is determined by the area where  $z_2$  and  $z_3$  are less than zero. Therefore we can see that increasing lag 1 correlation reduces the probability of obtaining a maximum.

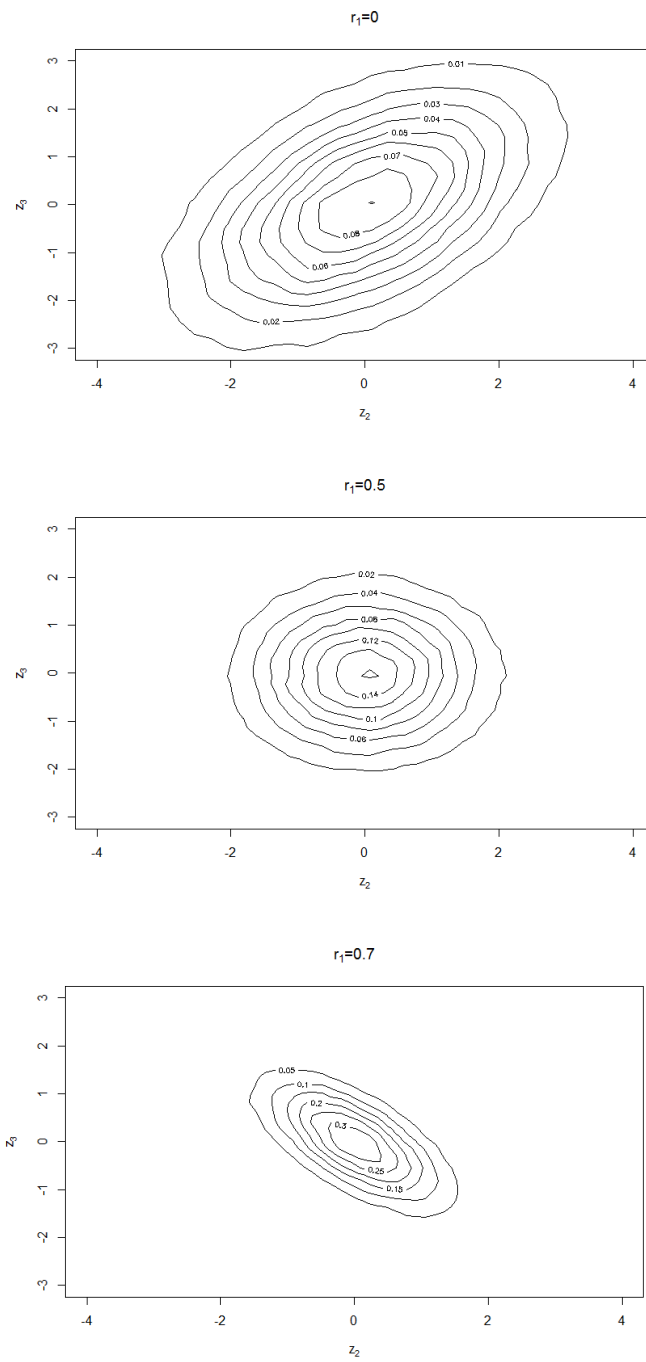


Figure 113: Contours of the bivariate normal distribution for fixed  $R$  and varying lag 1 correlations.

In a similar way, if lag 1 correlation is fixed, and either lag  $\sqrt{2}$  or lag 2 correlation increases, then some of the off diagonal elements of  $V$  increase, and so the probability of obtaining a maximum increases, also seen in Figure 112.

## 7.4 Conclusion

In this chapter it has been shown that the number of components and holes in the model residuals on various brain slices and the full brain are consistent with that of a isotropic Gaussian field. However, simulations on lattices have shown that distinctions between fields which are marginally normal with the same correlation may not be possible for smaller fields using components and holes alone. Also of interest is that the expected number of components and holes on the Gaussian field is extremely sensitive to local correlation, with only small changes in correlation causing large differences in number of components and holes on the field. In the next chapter, we will investigate if the spatial distribution of the components and holes on brain slices and the full brain are consistent with an isotropic Gaussian field.

## 8 Spatial TDA

In Chapter 7, we investigated the number of components and holes (or voids) on the brain to determine if the data is consistent with a Gaussian field. This extended work in [51] to irregular and three dimensional fields. In this chapter we will focus on the point patterns of the topological features of a field. The aim is to introduce a new method to use these point patterns to compare fields and to determine if a field is Gaussian, and apply the theory to the brain data.

In this chapter we initially test for spatial homogeneity in the model residuals, in particular the focus will be on testing if certain topological features are distributed randomly across the brain. Following this, methods and simulations will be used to show how to compare two point patterns, in particular to determine if a Gaussian point pattern can be distinguished from other fields as described in Section 7.2.1. The section will end with application to brain data, where we will test if the point pattern of topological features on the brain are consistent with a Gaussian field.

### 8.1 Introduction

Suppose we observe some measurements in space, we may ask if the observations are randomly located or if there is any clustering present. One simple method is quadrant count analysis [12] [36] [22], which splits the regions where observations may be made into blocks, or quadrants, and compares the number of observations in a block with the expected number under randomness. This method usually involves computing a single statistic, derived from the average number of counts in the blocks for the point pattern, which can be used to test for departure from spatial randomness.

Alternatives to quadrant count analysis which are commonly used are Ripley's K-function [65], Besag's L-function [38] and the pairwise correlation function (PCF), sometimes also known as the radial distribution function, described in [55]. In contrast to quadrant count analysis, these functions analyse the point pattern at different scales, and so can give more information. In this section, we will use both quadrant count analysis and the alternatives to investigate spatial point patterns, both to simulated data, and then to applications of topological features of the model residuals.

Another interesting question is how to compare two or more point patterns. In this section we will use the cross K-function to determine if there

is attraction or repulsion between two point patterns, which will allow us to determine, for example, if the positions of components influence the positions of holes or if the position of components at a particular time point influences the position at the next time point. Another method that will be explored will be to use metrics to determine the distance between two point patterns. This will allow us to investigate if the distances between the components, for example, from pairs of Gaussian fields, differ from pairs of fields from other distributions, in the case where the target correlation is the same. This will be applied to the brain data by constructing a test to determine if the components and holes on the brain slice are consistent with a Gaussian field. Quadrant count analysis will also be used to compare point patterns by counting the number of points for each pattern in each quadrant and performing an association test.

## 8.2 Complete Spatial Randomness - Quadrant Count Analysis

Quadrant count analysis is concerned with testing if observations are randomly distributed in space. For a Poisson distribution with mean  $\lambda$ , the probability of obtaining sample value  $n$  is

$$Pr(N = n) = \frac{\lambda^n}{n!} e^{-\lambda}$$

where  $E[N] = Var(N) = \lambda$ . Suppose the space where observations may be made is split up into a set of quadrants of equal size,  $V$ . Then the expected number of observed points in each quadrant, assuming a homogeneous Poisson process is  $\lambda V$ . If there are  $1, \dots, n$  quadrants, and  $N_1, N_2, \dots, N_n$  observations are made in each, with mean  $\bar{N}$ , then the ratio of the variance and mean of the counts is given by

$$IOD = \frac{1}{(n-1)\bar{N}} \sum_{i=1}^n (N_i - \bar{N})^2 \quad (73)$$

which is known as the index of dispersion. If the point pattern is consistent with a homogeneous Poisson process, then the  $IOD$  will be close to one. Note that this is very similar to the Pearson goodness of fit test

$$X^2 = \frac{\sum_{i=1}^n (N_i - \bar{N})^2}{\bar{N}} \quad (74)$$

which can be tested against the  $\chi_{n-1}^2$  distribution. In [12] it is noted that if the test statistic is large this implies high variability suggesting clustering. Similarly if the test statistic is low this implies lower than expected variability under randomness, suggesting regularity.

One important question when using quadrant count analysis is the choice of quadrant shape and number of quadrants. In [12] it is noted that the shape of the quadrants are not important, though the size should be such that the mean observed point count is at least one, and more than six quadrants should be used.

### 8.3 Complete Spatial Randomness - Ripley's K-function

Quadrant count analysis derives a single statistic from the point pattern to test for complete spatial randomness (CSR). An alternative is Ripley's K-function which tests for CSR at different spatial scales. In this section the background to the method will be given.

Let  $A$  be the area or volume of a space in two or three dimensions where a quantity of interest may be observed. Let  $n$  be the number of observed values within  $A$ . Then the expected number of observed values (under the assumption of spatial homogeneity) per unit area/volume (intensity of points) is given by

$$\lambda = \frac{n}{A}. \quad (75)$$

The intuition behind the K-function is to test for each individual observed value if it has the expected number of neighbours within a particular distance under the assumption that the points are randomly scattered. To do this, a circle (two dimensions) or sphere (three dimensions) of radius  $r$ , centered at the observed value is drawn, and neighbours within the circle/sphere are counted. Let  $K(r)$  denote the area of a circle (or volume of a sphere) of radius  $r$ , then given the expected number of points per unit area/volume is given by (75), the expected number of points,  $E(r)$ , within the circle/sphere is then

$$\lambda K(r) = E(r), \quad (76)$$

which implies that

$$K(r) = \frac{E(r)}{\lambda} = \pi r^2 \quad (77)$$



in two dimensions, and

$$K(r) = \frac{E(r)}{\lambda} = \frac{4}{3}\pi r^3 \quad (78)$$

in three dimensions.

To obtain an estimate for  $K(r)$  from observed data,  $E$  is replaced by counting the number of points within circles/spheres of radius  $r$  for all observations in  $A$  and this is averaged, to give the estimated K-function

$$\hat{K}(r) = \frac{1}{n} \frac{1}{\lambda} \sum_{i \neq j} I(d_{ij} < r) \quad (79)$$

where  $d_{ij}$  is the distance between observations  $i$  and  $j$  and  $I$  is the indicator function. This estimate should be approximately equal to  $\pi r^2$  (or  $\frac{4}{3}\pi r^3$  in three dimensions) if the data is randomly scattered. If there is clustering, more points will be observed than expected and so  $\hat{K}(r) > \pi r^2$ , and  $\hat{K}(r) < \pi r^2$  if the observed points are more regular than random data.

The K-function defined above for random data varies depending on the dimension of the data being considered. Also, as the distance,  $r$ , being considered increases, it is known [55] that fluctuations in the K-function increases also. Therefore often the L-function defined by

$$L(r) = \sqrt{\frac{d K(r)}{\pi}} = r \quad (80)$$

is used. The L-function is approximately  $r$  for all dimensions  $d$ , and the transform stabilises the large distance fluctuations in the K-function.

The K-function and L-functions are very similar in the information that they give about the point pattern process. They are cumulative in the sense that all points up to and including the desired  $r$  are considered. Depending on application, this may be a disadvantage. For example, suppose we are interested in the distribution of points at distance  $r_1 > r$ . If  $K(r) > \pi r^2$ , indicating clustering at the smaller distance, this information will impact on  $K(r_1)$  even if the point pattern is random at distance  $r_1$ . The pairwise correlation function (PCF) is an alternative to the K-function and L-function, which looks at the point pattern process in intervals. In two dimensions, let  $g(r)$  be the region between circles of radius  $r$  and  $r+dr$ . Given point intensity  $\lambda$  as above, then

$$\lambda g(r) = E(r) \quad (81)$$

where  $E(r)$  is now the expected number of points within the region  $g(r)$  if the points follow a Poisson process. Therefore

$$g(r) = \frac{E(r)}{\lambda} \approx A(g(r)) \quad (82)$$

where  $A(g(r))$  is area of the region  $g(r)$ .

To obtain an estimate from observed data, as before the expected value  $E(r)$  is replaced by averaged counts of points

$$\hat{g}(r) = \frac{1}{n} \frac{1}{\lambda} \sum_{i \neq j} I(r < d_{ij} < (r + dr)). \quad (83)$$

To stabilise the result at large distances the result is usually divided by  $A(g(r))$  [55], [6] and [10] so that  $\hat{g}(r) = 1$  for random data,  $\hat{g}(r) > 1$  for clustered data and  $\hat{g}(r) < 1$  for regular data.

### 8.3.1 Edge Effects

If two-dimensional observed data are randomly scattered within the area then  $\hat{K}(r) = \pi r^2$  in two dimensions. However this is only true if the circles about all points lie entirely within  $A$ . If this is not the case, then this is an incorrect theoretical value for  $K(r)$ , as if all circles do not fully lie within  $A$  then under spatial homogeneity we would expect  $K(r) < \pi r^2$ . In subsequent sections, the impact of such points will be shown to be significant when making inferences, therefore an adjustment needs to be made to (79) to account for the edge effect. From Equation (76) we have

$$\lambda K(r) = E(r)$$

where  $K(r)$  is the area of a circle of radius  $r$ . Let  $\alpha_i$  denote the proportion of the circle centered on point  $i$ , which lies within the area  $A$ , then

$$\lambda \alpha_i K_i(r) = E_i(r) \quad (84)$$

is the expected number of points within radius  $r$  of point  $i$  which are within the area  $A$ , and so

$$K_i(r) = \frac{E_i(r)}{\lambda \alpha_i} = \pi r^2. \quad (85)$$

As before, to estimate the K-function, the expected number of points  $E_i(r)$  is replaced with the observed count of points within radius  $r$  of point  $i$ , and an average over all points  $i$  is taken (Equation (86)).

$$\hat{K}(r) = \frac{1}{n} \frac{1}{\lambda} \sum_{i \neq j} w_i I(d_{ij} < r) \quad (86)$$

where  $w_i = 1/\alpha_i$  is the inverse of the proportion of the circle about point  $i$  which lies within  $A$ .

In order to calculate the weights  $w_i$ , we need to be able to calculate the proportion of the circle centered at point  $i$  which is within the area  $A$ . Fortunately the brain data are observations on a lattice, therefore we can overcome this problem of edge effects by making some small adjustments to the theory in this section. See Section 8.3.5 for more details.

### 8.3.2 Confidence Envelopes

As noted in Section 8.3, if the observed point pattern is homogeneous then the expected K-function in two dimensions is  $K(r) = \pi r^2$ . To assess if the observed K-function is close to enough to  $\pi r^2$  to conclude that the point pattern is homogeneous a confidence envelope is used. The idea is to simulate lots of homogeneous data on the area of interest, and using the minimum and maximum values of the K-function on the simulations, produce a lower and upper limit curve for the theoretical value. Then if the observed K-function lies within this envelope it can be concluded that the point pattern is consistent with homogeneity. The same idea can also be applied to the L-function and PCF.

### 8.3.3 Example: Homogeneous Data On A Square Region

In this section, two-dimensional simulations will be used to highlight how the K-function is typically implemented and the importance of edge effects.

The first simulation will be of 400 random points on a 65 by 65 square grid. This was chosen as it has approximately the same number of points as the largest brain slice, which has approximately 400 components. In this case  $A = 65^2 = 4225$  units squared and  $n = 400$ , thus the point intensity

$$\lambda = \frac{n}{A} = \frac{400}{4225} = 0.09$$

points per unit area.

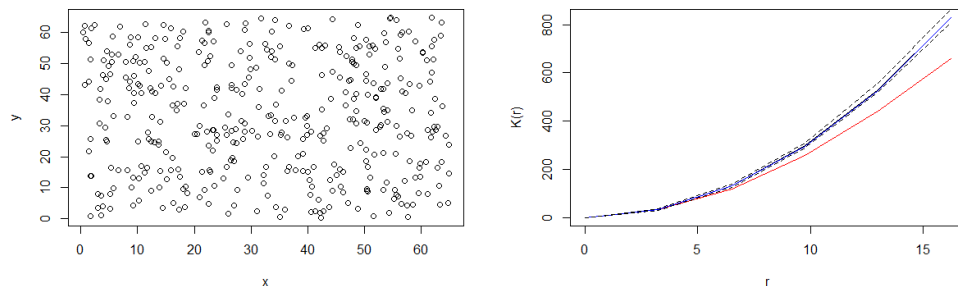


Figure 114: A random set of 400 points on a 65 by 65 grid (left) and the K-function under assumption of complete spatial randomness (black), with the observed K-function with (blue) and without (red) edge correction. Also included is the confidence envelope obtained from 50 simulations (black dashed).

From Figure 114 it can be seen that the observed K-function (red) is not consistent with the random point pattern, especially as  $r$  increases. This is because in this case no edge correction has been made. However, the blue curve shows that when edge correction is accounted for, the observed K-function does lie within the confidence interval for a random point pattern. In subsequent sections, all results given will use edge correction.

### 8.3.4 Example: Clustered And Regular Data On A Square Region

Figure 115 shows the K-function when applied to clustered and regular data. In this simulation, two clusters were simulated on the two-dimensional 65 by 65 grid, again using 400 observations, and the regular data were simulated so that the 400 observations were evenly spaced across the region.

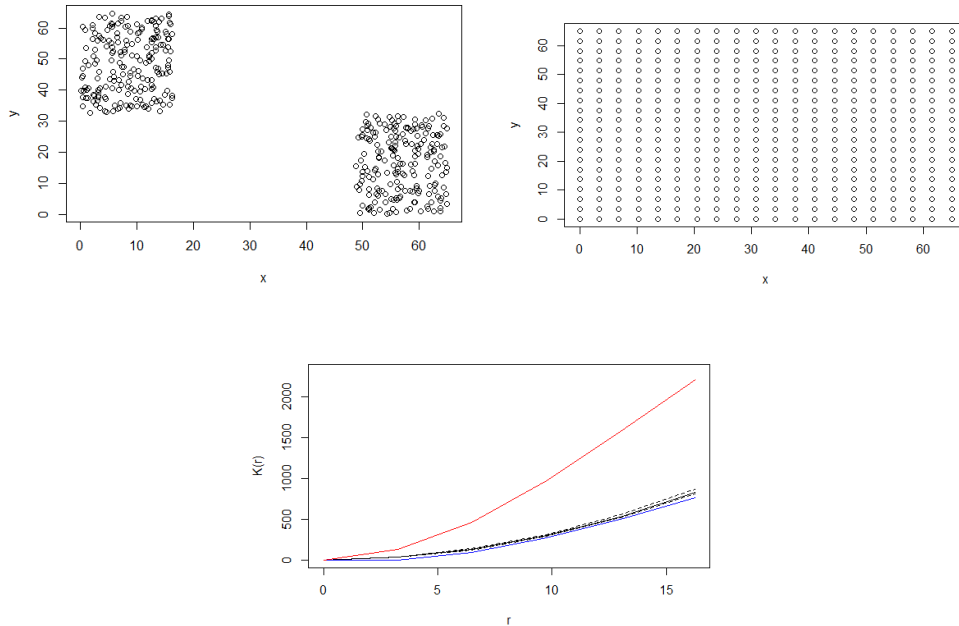


Figure 115: Sets of 400 clustered (top left) and evenly spaced points (top right) on a 65 by 65 grid and the K-function under assumption of complete spatial randomness (black), with confidence envelopes (black-dashed) and the observed K-function with edge correction for clustered (red) and regular data (blue).

Figure 115 shows that the observed K-function (red) for the clustered data is much higher than expected for homogeneous data, thus indicating clustering. Also observe that if data are evenly spaced (regular), the observed K-function falls only very slightly below the homogeneous confidence envelope. This suggests that the K-function may not work well when testing for regularity.

### 8.3.5 Testing For Homogeneity On A Lattice

In this section, we will make some adjustments to the K-function and PCF to test for spatial homogeneity on any lattice or set of coordinates, like the brain data. First, let us redefine

$$\lambda = \frac{n}{A} \quad (87)$$

where  $A$  is the total number of lattice points where an observation may be made, and  $n$  is the number of lattice points where an observation is made. Then  $\lambda$  is the marginal probability that any one of the  $A$  lattice points is an observed point.

Let  $\mathcal{P}_A$  denote the full set of lattice points where an observation may be made,  $\mathcal{O}_n \in \mathcal{P}_A$  be the  $n$  points on the lattice where observations are made and  $d_{ij}$  be the distance between point  $i$  and point  $j$  on the lattice. Then the expected number of points within radius  $r$  (under spatial randomness), from any point  $i$  on the lattice is given by

$$E_i(r) = \lambda \sum_{j \in \mathcal{P}_A, j \neq i} I(d_{ij} < r). \quad (88)$$

Summing over all points on the lattice, and averaging over the  $A$  total lattice points gives

$$E(r) = \frac{\lambda}{A} \sum_{(i,j) \in \mathcal{P}_A, j \neq i} I(d_{ij} < r). \quad (89)$$

This can then be compared with the average observed number of points,  $\hat{E}(r)$ , within radius  $r$  for all points in  $\mathcal{O}_n$ , given by

$$\hat{E}(r) = \frac{1}{n} \sum_{j \neq i, (i,j) \in \mathcal{O}_n} I(d_{ij} < r). \quad (90)$$

This method has an advantage over the K-function when dealing with a lattice, as edge effects are automatically accounted for. Note that  $E(r)$  now depends on  $\mathcal{P}_A$ , in contrast to the Ripley K-function which always has expected value  $\pi r^2$  in two dimensions. This is not a concern as  $\mathcal{P}_A$  is fixed for every given application.

Similar adjustments can be made to the PCF. Let  $E_g(r)$  denote the expected number of points that satisfy  $r < d_{ij} < r + dr$  under complete spatial randomness. Then

$$E_g(r) = \frac{\lambda}{A} \sum_{(i,j) \in \mathcal{P}_A, j \neq i} I(r < d_{ij} < r + dr). \quad (91)$$

Given an observed point pattern, this can be estimated as before using

$$\hat{E}_g(r) = \frac{1}{n} \sum_{j \neq i, (i,j) \in \mathcal{O}_n} I(r < d_{ij} < r + dr). \quad (92)$$

Analogous to the PCF (see Section 8.3), we will divide  $\hat{E}_g(r)$  by its expected value under randomness  $E_g(r)$  such that  $\hat{E}_g(r) = 1$  for homogeneous data,  $\hat{E}_g(r) > 1$  for clustered data, and  $\hat{E}_g(r) < 1$  for regular data.

## 8.4 Comparing Spatial Point Patterns

So far, this chapter has been concerned with testing if a single point pattern is completely random, clustered or dispersed. Another question of interest is the relationship between multiple point patterns. One method would be to calculate the distance between the two sets of points using a mapping which minimises the distance. However, if we would like to understand the relationship between point patterns for various scales, the cross K-function can be used [65], [3], [39].

### 8.4.1 Minimum Distance Between Point Patterns

A starting point for measuring the similarity between two point patterns is to calculate the distance between the two patterns. In this section, some commonly used distance metrics, as introduced in [17] and discussed in [67] will be described. Let the position of points in each point pattern  $X$  and  $Y$ , be respectively denoted  $x_i$  and  $y_j$ , where  $i = 1, \dots, n_1$  and  $j = 1, \dots, n_2$  with  $n_2 \geq n_1$ . Define  $d_0^*(x_i, y_j)$  to be the Euclidean distance between points  $x_i$  and  $y_j$ , and let  $d_0(x_i, y_j) \leq 1$  to be the min-max normalised distance given by

$$d_0(x_i, y_j) = \frac{d_0^*(x_i, y_j) - \min(d_0^*(x_i, y_j))}{\max(d_0^*(x_i, y_j)) - \min(d_0^*(x_i, y_j))}.$$

Then

$$d_1(X, Y) = \min_{\pi \in \Pi_{(n_1, n_2)}} \frac{1}{n_1} \sum_{i=1}^{n_1} d_0(x_i, y_{\pi_i}) \quad (93)$$

is the minimum average distance between the two point patterns, considering all possible mappings of the smaller point pattern  $X$  on to the larger point pattern  $Y$ . Here  $\Pi_{(n_1, n_2)}$  is the set of all possible combinations of  $n_1$  elements

of  $Y$ , and  $y_{\pi_i}$  is the  $i$ -th point of  $Y$  for combination  $\pi$ . In the general problem of comparing pairs of point patterns, it is important that the Euclidean distance is normalised, so that the results are based on the similarity of the distribution of points rather than the physical coordinates. For example, consider three points patterns  $A$ ,  $B$  and  $C$  where  $A$  and  $B$  have very similar point patterns but have very different coordinates, but  $A$  and  $C$  are close in space but have very different point patterns. Without normalising the distances, we would conclude that  $A$  is more similar to  $C$  than  $B$ .

The distance metric  $d_1(X, Y)$  could be extended to allow for a  $p$ -th order average

$$d_2(X, Y) = \min_{\pi \in \Pi_{(n_1, n_2)}} \frac{1}{n_1} \left( \sum_{i=1}^{n_1} d_0(x_i, y_{\pi_i})^p \right)^{1/p}. \quad (94)$$

A disadvantage of the  $d_1$  and  $d_2$  metrics are that if  $n_1 \neq n_2$  then  $d_1(X, Y)$  will be smaller than a similar point pattern with  $n_1 = n_2$ , as it is more likely to find a permutation which maps  $X$  on to  $Y$  with smaller distance if  $Y$  has many more points than  $X$ , as not all points in  $Y$  are required to map to a point in  $X$ . A common method [17] to overcome differences in number of points is to use the metric

$$d_3(X, Y) = \begin{cases} d_2(X, Y), & \text{if } n_1 = n_2 \\ 1, & \text{if } n_1 \neq n_2. \end{cases} \quad (95)$$

This metric has the opposite problem to  $d_2$ . Instead of giving small distances when the number of points in the two patterns differ,  $d_3$  assigns the maximum distance. This is disadvantageous as one could imagine two almost identical point patterns, differing by a very small number of points, giving a result which would indicate the maximum possible distance between the two patterns. In [67], a solution to the problem of different numbers of points in the patterns is proposed by using the metric

$$d_4(X, Y) = \min_{\pi \in \Pi_{(n_1, n_2)}} \frac{1}{n_2} \left( \sum_{i=1}^{n_2} d_0(x_i, y_{\pi_i})^p + c(n_2 - n_1) \right)^{1/p} \quad (96)$$

for some choice of  $p \geq 1$  and  $0 \leq c \leq 1$ . The effect of adding  $(n_2 - n_1)$  inside the sum of Equation (96) is to add the difference in number of points to the smaller point pattern at the maximum allowed distance of one (for  $c = 1$ ). In



Section 8.6.1 we will use simulation studies to show that the metric  $d_4(X, Y)$  is a better metric than the first three when the number of points differ in the point patterns, and the choice for values of  $p$  and  $c$  will be explored.

Finally, we will consider the well known discrete linear transport problem [62] which in general aims to minimise the costs associated with transporting supply between locations. In this context the distances between topological features of two point patterns (components or holes) are the costs which we aim to minimise, and we assume that each point pattern has total mass (supply)  $n_1$  to be spread evenly across the points. The *spatspat* package in *R* implements an algorithm to find the weights  $w_{ij}$  which minimises the weighted distance,  $d_5(X, Y)$ , between the two sets of points as described in (97).

$$d_5(X, Y) = \min \sum_i \sum_j w_{ij} d_0(x_i, y_j) \quad (97)$$

subject to the constraints

$$\begin{aligned} \sum_j w_{ij} &= 1, \forall i \in [1, n_1] \\ \sum_i w_{ij} &= \frac{n_1}{n_2}, \forall j \in [1, n_2] \\ w_{ij} &> 0 \end{aligned}$$

In Section 8.6.1 these various distance metrics will be explored using simulations.

#### 8.4.2 Quadrant Count Analysis

As well as testing if a single pattern is homogeneous (see Section 8.2 ) quadrant count analysis can be used to compare two point patterns. Suppose we observe two point patterns on the same space. As in Section 8.2, suppose the space is split into a set of  $n$  quadrants, not now necessarily equal in size. To compare the two point patterns, one could count the number of observations of each pattern in each quadrant, and perform a  $\chi^2$  association test, to determine if there is an association between the quadrants and the type of point.

If there are  $n$  quadrants, we can count the number of points of pattern X and pattern Y in each quadrant to obtain the table

Quadrant	1	2	3	...	$n$	Total
Pattern X	$O_{1,1}$	$O_{1,2}$	$O_{1,3}$	...	$O_{1,n}$	$r_1$
Pattern Y	$O_{2,1}$	$O_{2,2}$	$O_{2,3}$	...	$O_{2,n}$	$r_2$
Total	$c_1$	$c_2$	$c_3$	...	$c_n$	$N$

Table 54: The observed number  $O_{i,j}$  of points of pattern  $i$  in quadrant  $j$ .

Then the  $\chi^2$  test statistic is given by

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^n \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

where  $O_{i,j}$  is the number of observed points of pattern  $i$  in quadrant  $j$ , and  $E_{i,j}$  is the expected frequency given by

$$E_{i,j} = \frac{r_i c_j}{N}$$

where  $r_i$  is the total number of observed points in pattern  $i$  and  $c_j$  is the total number of observed samples of any point type in quadrant  $j$ . The test statistic can be tested against the  $\chi^2$  distribution with  $(n - 1)$  degrees of freedom.

### 8.4.3 Cross K-function

The cross K-function is very similar in definition to the standard K-function described in Section 8.3. As we are working on lattice data, all definitions given here will be adjusted to account for the lattice configurations. Given two spatial point patterns, both observable on the lattice points  $\mathcal{P}_A$ , with  $n_1$  and  $n_2$  points observed from point pattern X and Y respectively, we can define

$$\lambda_1 = \frac{n_1}{A} \tag{98}$$

$$\lambda_2 = \frac{n_2}{A} \tag{99}$$

as the marginal probabilities that any one of the  $A$  lattice points are from point pattern one or two. Let  $E_1(r)$  denote the expected number of points of type one within distance  $r$  of any given point on the lattice  $A$ , then as in Equation (89) we have

$$E_1(r) = \frac{\lambda_1}{A} \sum_{(i,j) \in \mathcal{P}_A, j \neq i} I(d_{ij} < r) \quad (100)$$

with a similar expression for  $E_2(r)$ . As before let  $\mathcal{P}_A$  denote the full set of lattice points where an observation may be made,  $\mathcal{O}_{n_1} \in \mathcal{P}_A$  be the  $n_1$  points on the lattice where observations of point pattern one are made and  $\mathcal{O}_{n_2} \in \mathcal{P}_A$  be the  $n_2$  points on the lattice where observations of point pattern two are made. Then

$$\hat{E}_{12}(r) = \frac{1}{n_1} \sum_{i \in \mathcal{O}_{n_1}, j \in \mathcal{O}_{n_2}} I(d_{ij} < r) \quad (101)$$

is the average observed number of points of type two within distance  $r$  of the type one points on the lattice  $A$ . We can then compare  $\hat{E}_{12}(r)$  with  $E_2(r)$  to determine if more or less points are observed than expected under randomness.

#### 8.4.4 Confidence Envelopes

When investigating a single point pattern, the goal was to test if the pattern was random. Now we would like to test for similarity between point patterns. In particular we would like to know if the point patterns are independent, or if there is attraction or repulsion between them. One such method is a random labelling test. Full details can be found in [3] and [47], and here brief details will be given.

Given observed locations  $(s_1, \dots, s_n)$  with labels  $(m_1, \dots, m_n)$  with  $n_1$  points from process one, and  $n_2$  points from process two, randomly permute the labels  $\tau = 1 \dots N$  times. For each permutation compute the observed cross K-function  $\hat{E}_{12}^\tau(r)$ . If the observed point processes are spatially indistinguishable, then the observed cross K-function  $\hat{E}_{12}(r)$  should be a typical sample from  $\hat{E}_{12}^\tau(r)$ . Using the permuted samples, we can construct  $p$ -values to test this.

Let  $M_+(r)$  and  $M_-(r)$  denote the number of permutations where  $\hat{E}_{12}^\tau(r) \geq \hat{E}_{12}(r)$  and  $\hat{E}_{12}^\tau(r) < \hat{E}_{12}(r)$  respectively. Then define attraction and repulsion  $p$ -values as

$$\hat{p}_{\text{attraction}}(r) = \frac{M_+(r) + 1}{N + 1} \quad (102)$$

$$\hat{p}_{\text{repulsion}}(r) = \frac{M_-(r) + 1}{N + 1}. \quad (103)$$

The attraction  $p$ -value is the proportion of permutations where the permuted K-function is at least as large as the observed. Hence if  $M_+(r)$  is small, this means that the observed cross K-function is larger than expected under the random permutations, implying an attraction between the point processes. Similarly if  $M_-(r)$  is small, this implies that the original point patterns have smaller cross K-function than expected under random permutations, implying a repulsion between points in the point processes.

Note that the method described here, presented in [3] and other works, is based on permuting the point labels for the two spatial patterns only, and it ignores the overall space where observations may be made. The argument for this is that it is advantageous when the space where observations may be made is irregular, maybe with gaps or holes. For example, analysing the relationship between positions of two different species of plants in a park, where there is a lake in the middle where plants cannot grow. However, consider the case where there are clustered data, where each cluster is a mix of the two point patterns. Randomly permuting the labels of the points and calculating the cross K-function will very likely give a similar result to that of the observed, ie if we allow only the observed points to be permuted, there is a high chance that each cluster will contain a mix of the two point types. Therefore in this case, any confidence interval for spatially indistinguishable point patterns will very likely include the observed cross K-function for the clustered data. This may be interpreted as independence between the points, even if clear attraction can be seen.

To overcome this, note that the theoretical work in [3] does not limit the labels of points to be restricted to two groups. Therefore, in this work we will allow a third group which are points where observations are not made but are in the overall space where points could be made. This overcomes the problem of the given example. Simulated examples of this will be given in Section 8.6.4.

## 8.5 Simulations I - Complete Spatial Randomness

### 8.5.1 Quadrant Count Analysis On A Square Lattice

In this section we will use quadrant count analysis on simulated data sets to determine if departures from randomness can be detected using this method. For all two dimensional simulations on a lattice, the dimension of the lattice will be 65 by 65. The number of points inside the lattice where we assume observations are made will be chosen to be 400. Hence for these simulations

$$A = 65^2 = 4225$$

lattice points, and so the probability that any one of the 4225 points is in the observed set is

$$\lambda = \frac{n}{A} = \frac{400}{4225} = 0.09$$

As noted in Section 8.2, there is no general rule for the size of the quadrants, the number chosen is usually application-dependent. However it is asserted in [12] that the expected mean count of points in each quadrant should exceed one. In this situation, this means that the number of lattice points per quadrant should exceed  $1/0.09 \approx 12$  lattice points.

Figure 116 shows a single simulation of the random points on the lattice, and the  $p$ -value obtained from the  $\chi^2$  goodness of fit test for a number of different quadrant sizes. Non-overlapping quadrants were chosen, and in each case the quadrant sizes were chosen such that all quadrants were of equal size, and each lattice point was contained in exactly one quadrant. The figure shows the 0.05 critical value for the test (black dashed). Also included is the 0.95 critical value that in [12] is used as a threshold to indicate regular points.

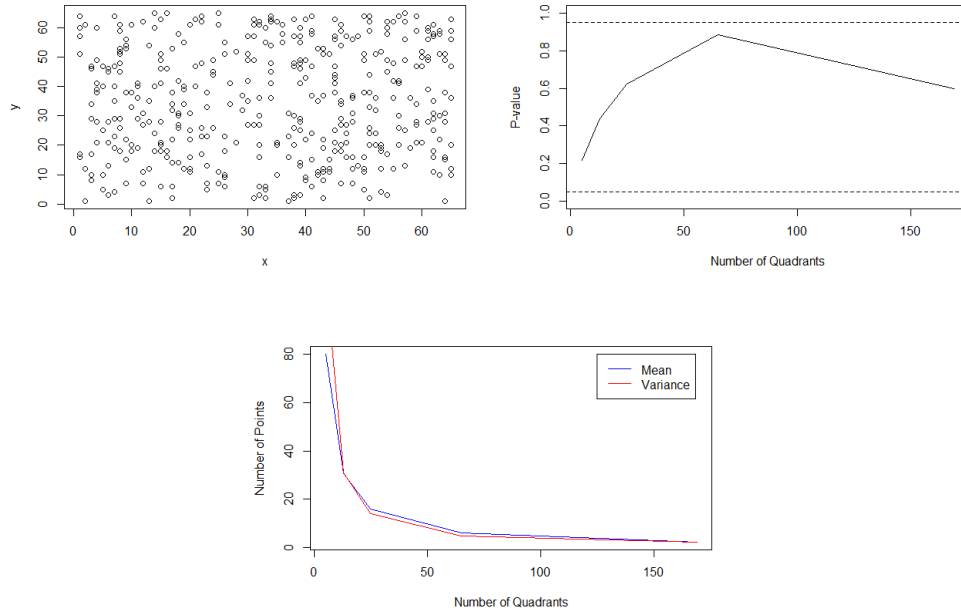


Figure 116: A random set of 400 points on a 65 by 65 lattice (top left), the  $\chi^2$   $p$ -value for a number of different number of quadrants, with critical values (black dashed) (top right), and the mean and variance of points in the quadrants (bottom centre).

Figure 116 shows that for a random set of points on the lattice, the  $\chi^2$   $p$ -value is greater than 0.05 and less than 0.95 for all quadrant sizes which means there is no evidence in any case for departure from randomness, as expected. The figure also shows that in all cases, the mean and variance of points in the quadrants are approximately equal, as expected from a random Poisson process.

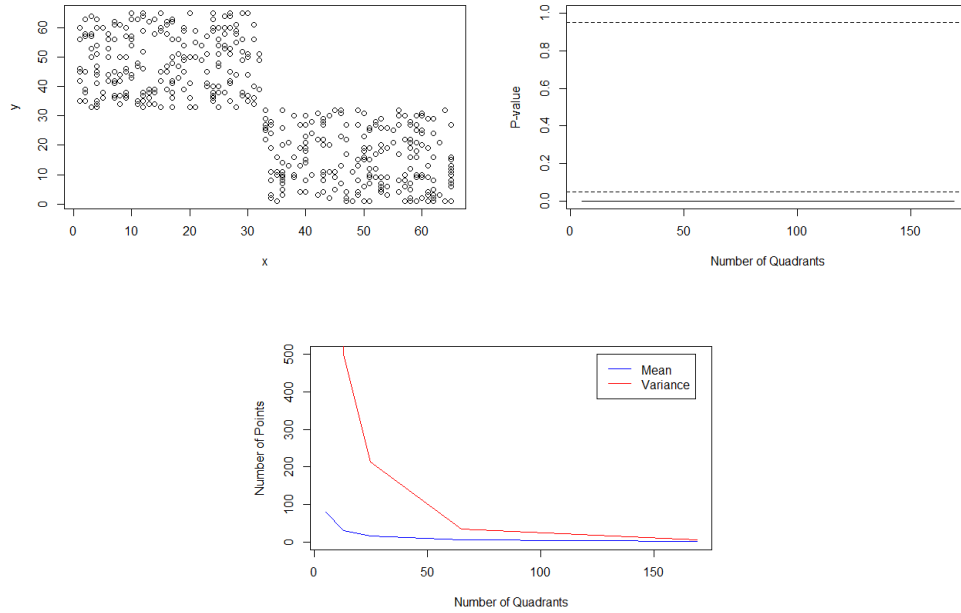


Figure 117: A clustered set of 400 points on a 65 by 65 lattice (top left), the  $\chi^2$   $p$ -value for a number of different number of quadrants, with critical values (black dashed) (top right), and the mean and variance of points in the quadrants (bottom centre).

Figure 117 shows results for the case of extreme clustering, and from this it can be seen that the  $\chi^2$   $p$ -value is very small in all cases, indicating that the counts are much higher than expected. Also note that the mean and variance differ significantly, which shows that the point pattern is not consistent with the homogeneous Poisson process. Here the variance is higher than the mean as expected for clustered data.

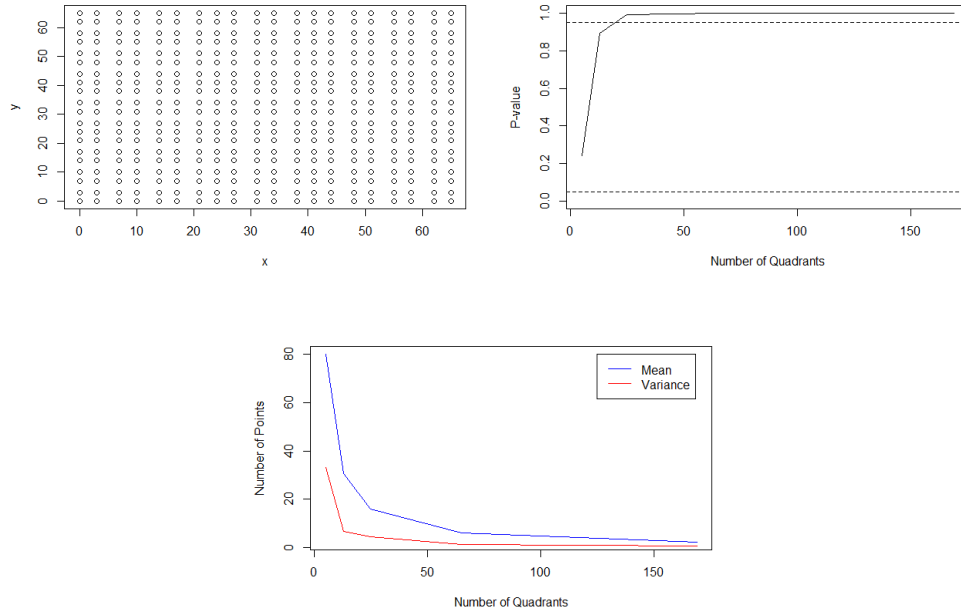


Figure 118: A regular set of 400 points on a 65 by 65 lattice (top left), the  $\chi^2$   $p$ -value for a number of different number of quadrants, with critical values (black dashed) (top right), and the mean and variance of points in the quadrants (bottom centre).

Finally Figure 118 shows the results for regular data. From this it can be seen that for the lowest number of quadrants, there is no evidence from departure from randomness. However as the size of the quadrants increase, the  $p$ -value is larger than 0.95. The mean/variance plot shows that the variance is lower than the mean, which means we can conclude that the point pattern is regular, as expected.

### 8.5.2 K-function On A Square Lattice

In this section we will compute the adjusted K-function and pairwise correlation function (PCF) for the random set of points on the 65 by 65 lattice as seen in Figure 116. Figure 119 shows  $E(r)$  and the confidence envelope obtained from 50 simulations of random data on the lattice.



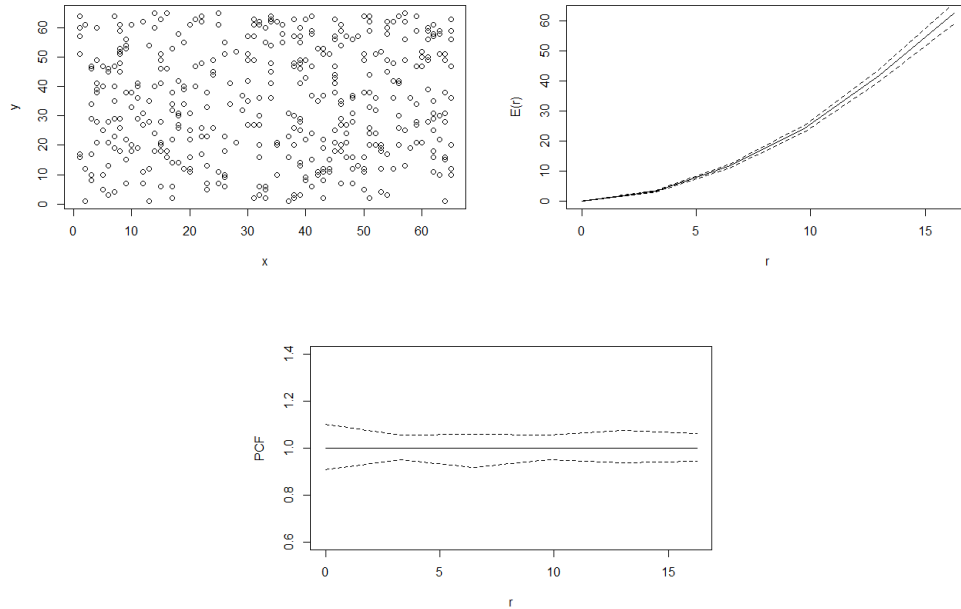


Figure 119: A random set of 400 points on a 65 by 65 lattice (top left),  $E(r)$  under assumption of complete spatial randomness (black), with the confidence envelope obtained from 50 simulations (black dashed) (top right) and  $E_g(r)$  with its associated confidence envelopes (bottom centre).

Figure 120 shows the K-function when applied to clustered and regular data, as seen in Figures 117 and 118. From this it can be seen that  $\hat{E}(r) > E(r)$  for the clustered data, and  $\hat{E}(r) < E(r)$  for the regular data as expected. The simulation results in this section show that we can assess spatial homogeneity on a lattice, and detect when there is clustering or regularity in the data.

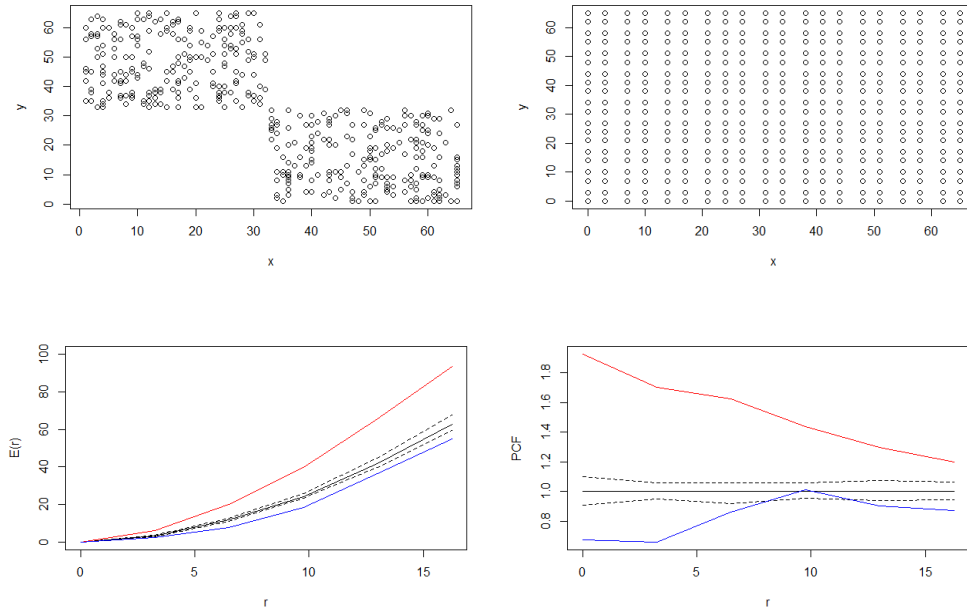


Figure 120: A set of 400 clustered points (top left) and regular points (top right) on a 65 by 65 lattice,  $E(r)$  (bottom left) and  $E_g(r)$  (bottom right) under assumption of complete spatial randomness (black), with confidence envelopes (black-dashed) and estimated functions for the clustered (red) and dispersed data (blue).

### 8.5.3 Comparing The K-function Of Different Distributions

In this section we will now return to point patterns generated using topological data analysis. As in Section 7.2.1, various random fields will be generated on the lattice such that they all are marginally normalised with the same correlation structure. In Chapter 7 it was found that under certain circumstances, differences could be seen in the number of components and holes between the distributions, when considering the lattice data. The goal in this section is to investigate if differences can be seen in how these components and holes are distributed across the lattice, by calculating and comparing the K-function for the various point patterns.

The method here is to simulate 50 data sets from all distributions, using the correlation parameters given in Section 7.2.1, then to compute the

K-function and radial distribution function to determine if there is any differences between the distributions.

Figures 121 and 122 show examples of the locations of the components and holes from two simulations of the various distributions on the lattice when  $(\nu, \eta) = (1, 5)$ . These correspond to the heatmaps seen in Section 7.2.2, Figure 80. Appendix 10.5 shows the locations of the components and holes for the other sets of parameters and distributions.

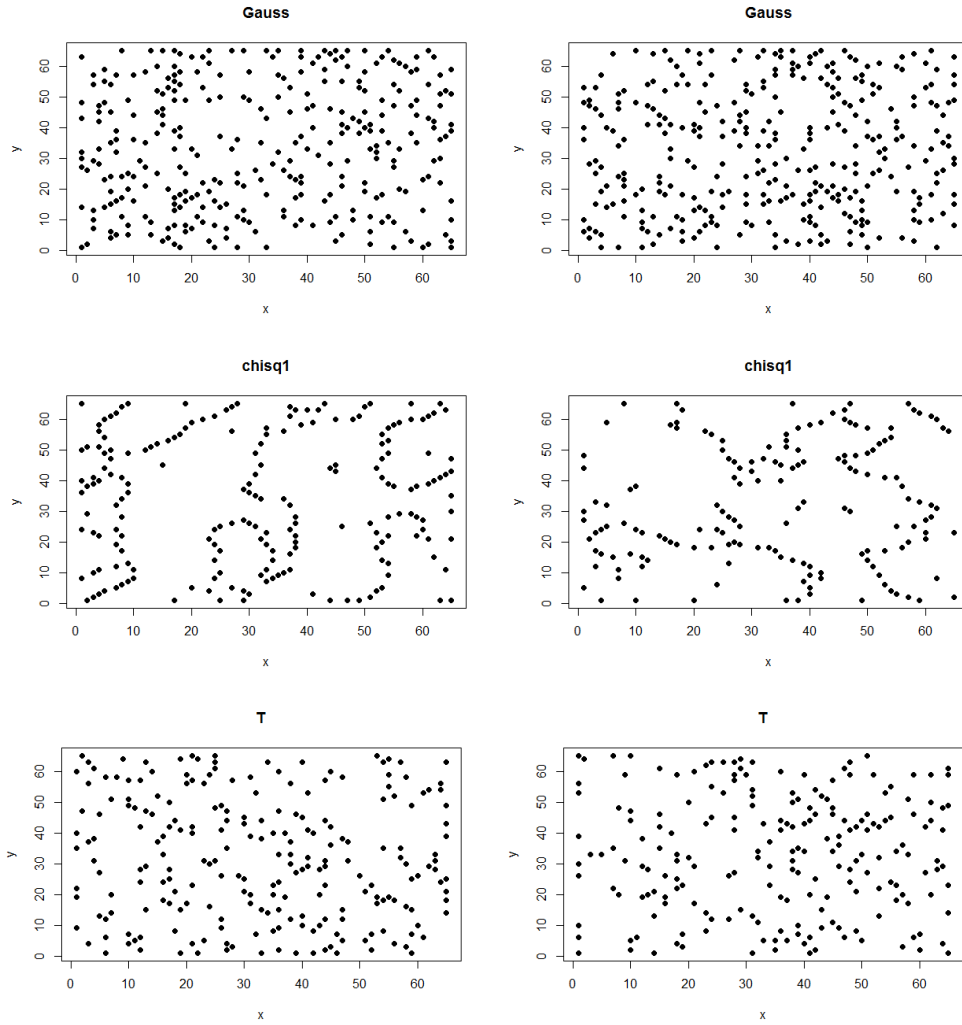


Figure 121: The position of the components from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (1, 5)$ .

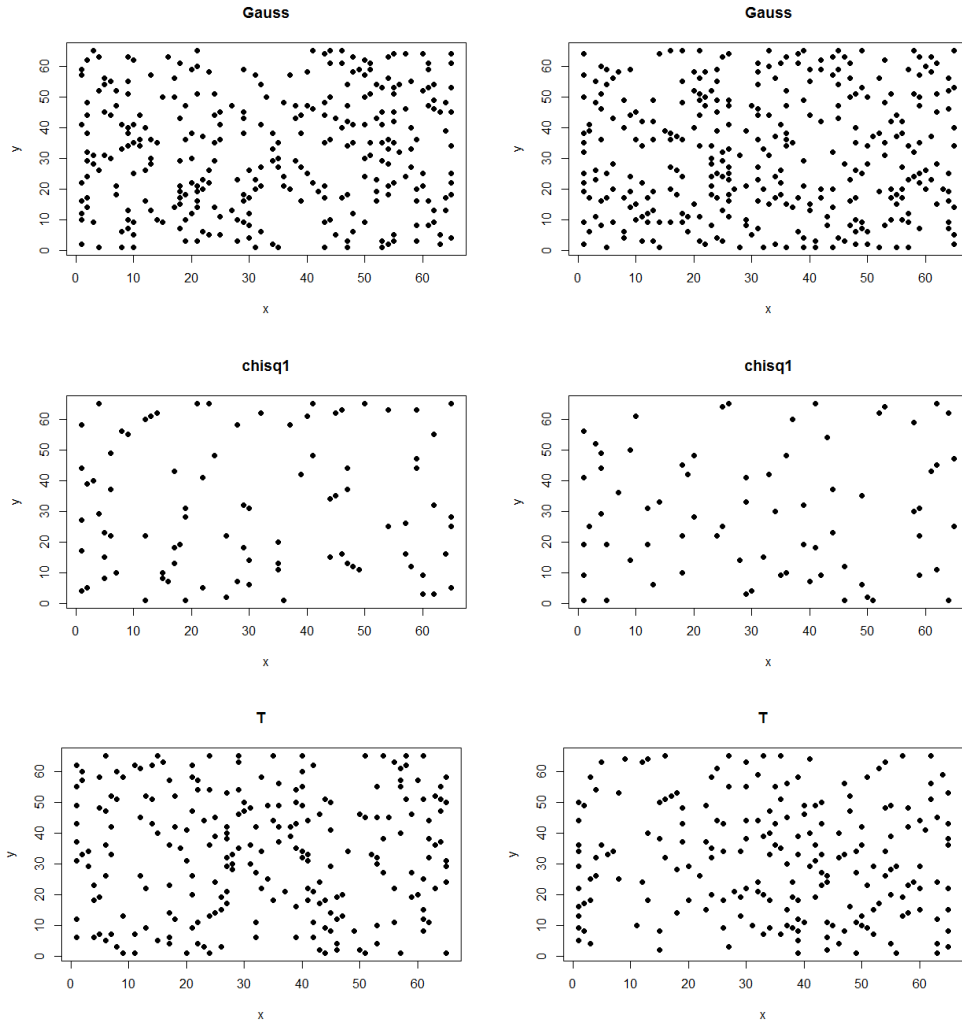


Figure 122: The position of the holes from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (1, 5)$ .

Figures 123 and 124 shows the mean K-function from all the distributions and parameters considered in Section 7.2.1 from 50 simulations. The plot shows (black dashed) the mean plus/minus two standard deviations for the Gaussian. Figure 125 shows the results for smaller values of  $r$ .

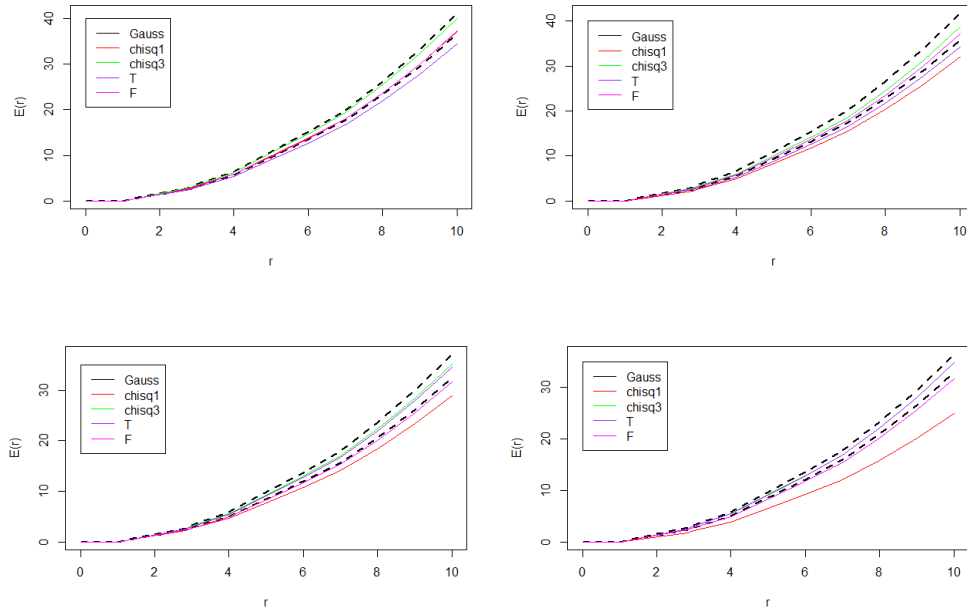


Figure 123: The mean K-functions for each distribution (black dashed is Gaussian interval) from the point pattern of components (left) and holes (right) on a 65 by 65 lattice from 50 simulations. All distributions marginally normal with similar correlation. Top row shows results for Gaussian Matérn  $(\nu, \eta) = (0.5, 2)$  and bottom row  $(\nu, \eta) = (0.5, 5)$ .

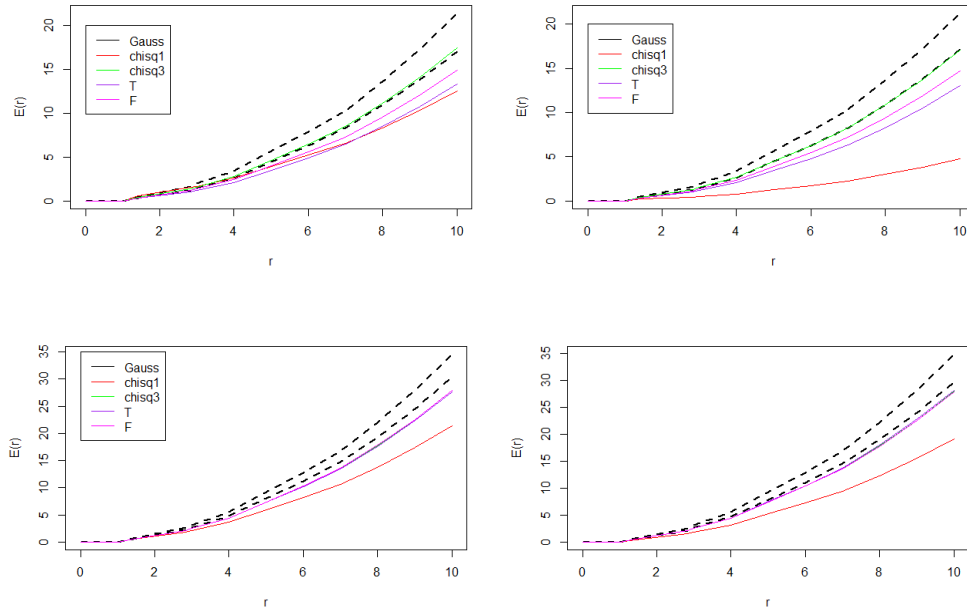


Figure 124: The mean K-functions for each distribution (black dashed is Gaussian interval) from the point pattern of components (left) and holes (right) on a 65 by 65 lattice from 50 simulations. All distributions marginally normal with similar correlation. Top row shows results for Gaussian Matérn  $(\nu, \eta) = (1, 5)$  and the bottom row  $(\nu, \eta) = (0.5, 20)$ .

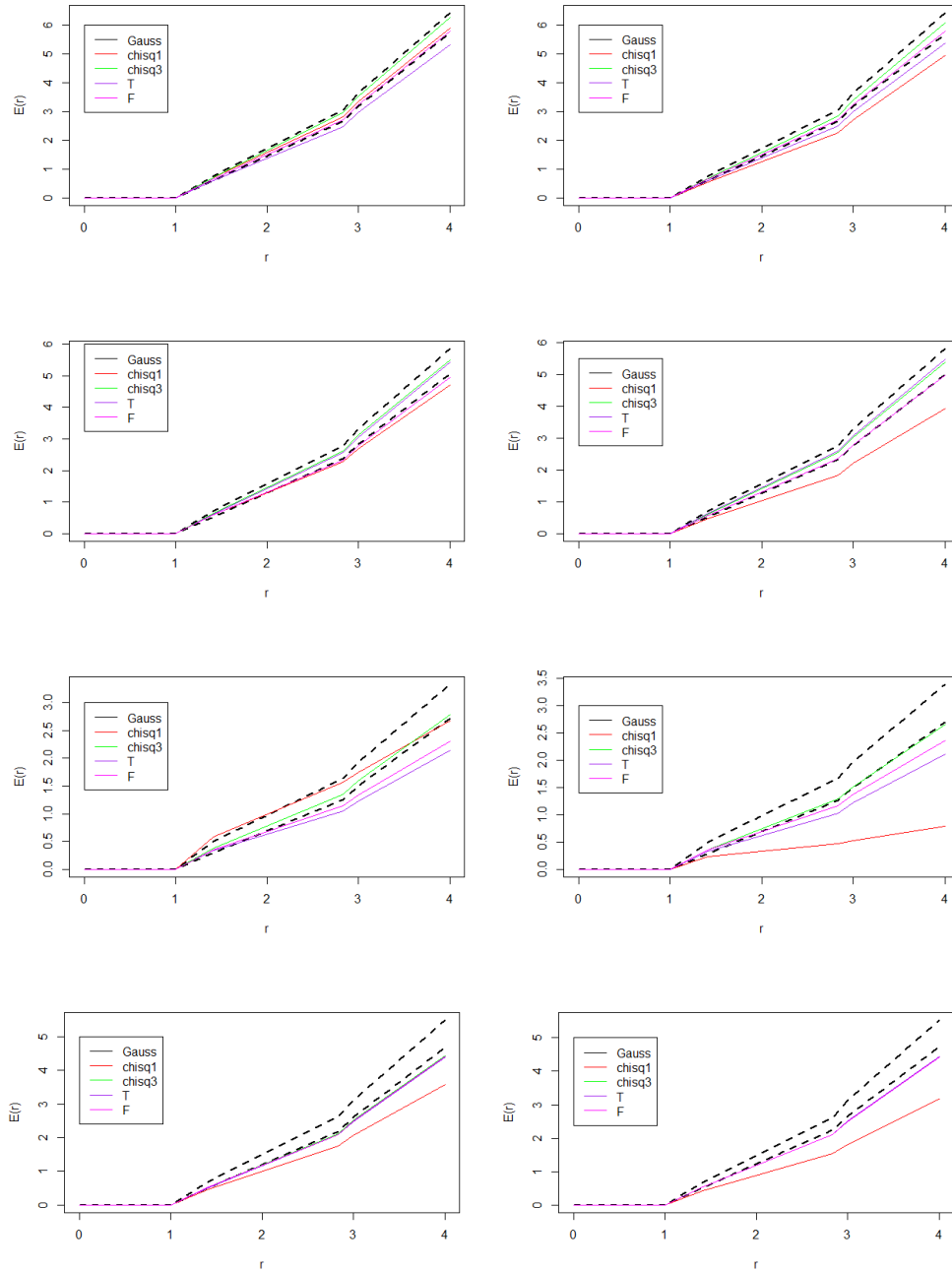


Figure 125: As in Figures 123 and 124 for small values of  $r$ . Top to bottom: Gaussian Matérn  $(\nu, \eta) = (0.5, 2)$ ,  $(\nu, \eta) = (0.5, 5)$ ,  $(\nu, \eta) = (1, 5)$  and  $(\nu, \eta) = (0.5, 20)$



From Figures 123 and 125 it can be seen that there is a lot of overlap between the various random fields. The results from the  $T$  distribution lie outside two standard deviations of the Gaussian in almost all cases. Also, the asymmetry of the  $\chi_1^2$  distribution can be seen as before in almost all cases. Therefore, a distinction can be made between the Gaussian,  $\chi_1^2$  and  $T$  distributions. However in almost all cases, the  $\chi_3^2$  and  $F$  results for both components and holes lie inside the Gaussian interval.

Table 55 and 56 shows the numerical results for the case where  $(\nu, \eta) = (0.5, 2)$ .

	Gauss	$\chi_1^2$	$\chi_3^2$	$T_3$	$F_{3,3}$
$r$					
1	0	0	0	0	0
$\sqrt{2}$	0.68(0.04)	0.71 (0.05)	0.72 (0.04)	0.59 (0.05)	0.65 (0.05)
2	2.85(0.08)	2.82 (0.12)	2.97 (0.10)	2.48 (0.09)	2.72 (0.09)
$\sqrt{8}$	3.40(0.10)	3.34 (0.13)	3.53 (0.11)	2.97 (0.11)	3.24 (0.11)
3	6.05(0.17)	5.89 (0.25)	6.30 (0.17)	5.32 (0.15)	5.80 (0.19)
4	10.22(0.29)	9.80 (0.40)	10.59 (0.29)	9.00 (0.24)	9.76 (0.28)
5	14.29(0.38)	13.70 (0.58)	14.78 (0.39)	12.62 (0.34)	13.68 (0.44)

Table 55: The average observed number of components within distance  $r$  on a 65 by 65 lattice with each given root distribution, all marginally normal with similar correlation (here Gaussian Matérn  $(\nu, \eta) = (0.5, 2)$ ). The mean (standard deviation) from 50 simulations of each distribution is given.

	Gauss	$\chi_1^2$	$\chi_3^2$	$T_3$	$F_{3,3}$
$r$					
1	0	0	0	0	0
$\sqrt{2}$	0.68(0.06)	0.55 (0.05)	0.68 (0.04)	0.59 (0.04)	0.65 (0.04)
2	2.85(0.11)	2.28 (0.11)	2.84 (0.11)	2.45 (0.11)	2.72 (0.10)
$\sqrt{8}$	3.40(0.13)	2.75 (0.13)	3.39 (0.12)	2.93 (0.12)	3.26 (0.11)
3	6.09(0.20)	4.99 (0.22)	6.06 (0.17)	5.27 (0.17)	5.81 (0.19)
4	10.27(0.32)	8.52 (0.37)	10.22 (0.30)	8.94 (0.30)	9.78 (0.27)
5	14.38(0.42)	11.95 (0.52)	14.32 (0.42)	12.52 (0.39)	13.71 (0.38)

Table 56: The average observed number of holes within distance  $r$  on a 65 by 65 lattice with each given root distribution, all marginally normal with similar correlation (here Gaussian Matérn  $(\nu, \eta) = (0.5, 2)$ ). The mean (standard deviation) from 50 simulations of each distribution is given.

Table 55 shows that the average number of components within radius  $r$  is very similar for all distributions. In almost all cases, the results lie within two standard deviations of each other. However note that the results from the  $T$  distribution are all slightly lower than the rest. Table 56 shows similar results for the holes. Note that also in this case the  $\chi_1^2$  results are also lower than the rest, again highlighting the asymmetry in the components and holes on the  $\chi_1^2$  field. There are no observed components within distance  $r = 1$  for any distribution, as it is not possible that two neighbours on a lattice are both components. Figure 126 shows the corresponding  $PCF$  results.

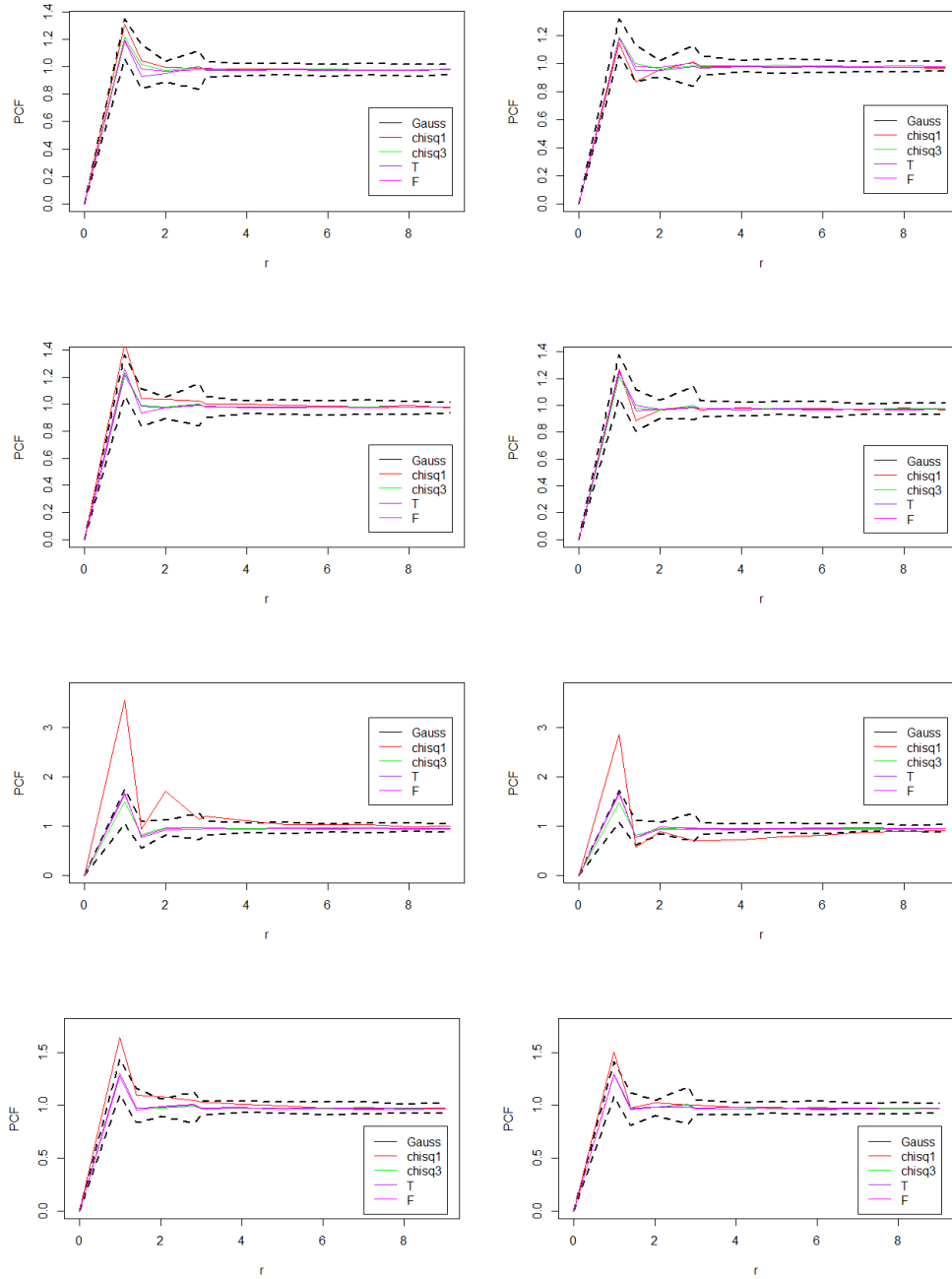


Figure 126: As in Figure 125 but for the pcf.

Again, Figure 126 shows that there is little difference between distributions. There are some exceptions though. For all correlation lengths, in particular the higher ones, there is evidence again of the asymmetry in the  $\chi_1^2$  results. This is most notable in the cases where the Gaussian Matérn correlation parameters are  $(\nu, \eta) = (1, 5)$  and  $(\nu, \eta) = (0.5, 20)$ . In both these cases, there is high local correlation, and in particular for  $(\nu, \eta) = (1, 5)$  there is a very large difference between the number of components and holes on the  $\chi_1^2$  field and the other distributions, in the intervals  $1 < r \leq \sqrt{2}$  and  $2 < r \leq \sqrt{8}$ .

## 8.6 Simulations II - Comparing Point Patterns

### 8.6.1 Minimum Distance Between Point Patterns - Cumulative Distribution Functions

In this section we will use simulations to compare two spatial point patterns using the distance metrics introduced in Section 8.4.1. As in Section 7.2.1 data will be simulated from various random fields, all marginally  $N(0, 1)$  and with the same target correlation structure. From these simulated fields, the point patterns of the position of components and/or holes will be compared for the various distributions. The goal is to determine if differences between the fields can be seen by the distance between the positions of the components or holes.

Firstly, we will simulate 100 pairs of random realisations of each of the five different field types described in Section 7.2.1. For each pair we will calculate the distance between the components on the fields, to obtain five vectors of length 100, one for each distribution. We will plot the cumulative density function of each to see if there are any differences in the distance between components on the various fields. This will be repeated for the holes, and also for the four sets of parameters. Figures 127 to 130 show the results when the metric in Equation (97) is used.

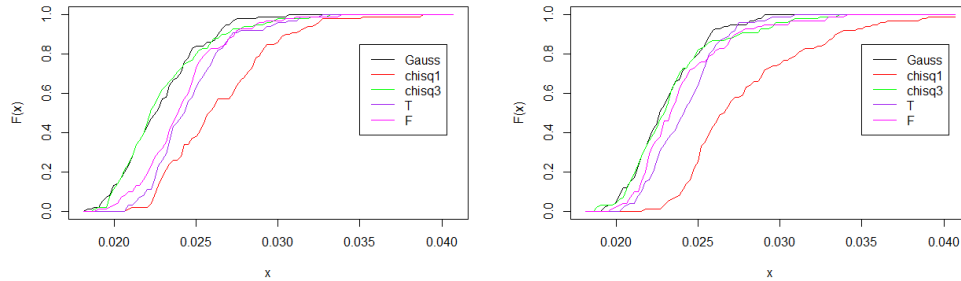


Figure 127: The cumulative distribution function of the minimum distances between components (left) and holes (right) of 100 simulated pairs of each field on a 65 by 65 lattice and correlation parameters  $(\nu, \eta) = (0.5, 2)$ .

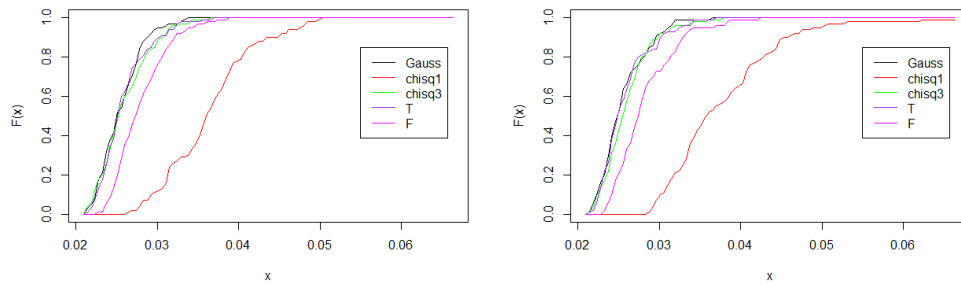


Figure 128: The cumulative distribution function of the minimum distances between components (left) and holes (right) of 100 simulated pairs of each field on a 65 by 65 lattice and correlation parameters  $(\nu, \eta) = (0.5, 5)$ .

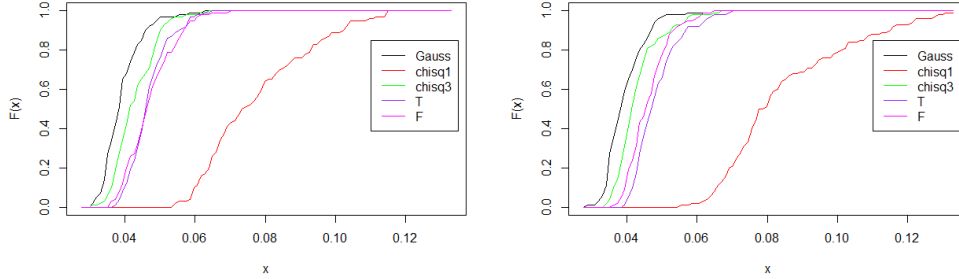


Figure 129: The cumulative distribution function of the minimum distances between components (left) and holes (right) of 100 simulated pairs of each field on a 65 by 65 lattice and correlation parameters  $(\nu, \eta) = (1, 5)$ .

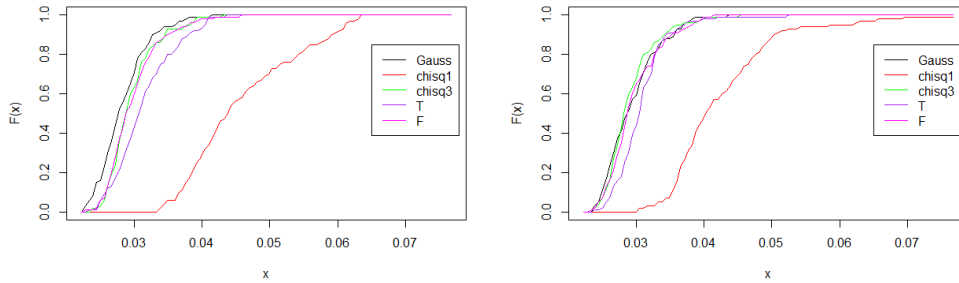


Figure 130: The cumulative distribution function of the minimum distances between components (left) and holes (right) of 100 simulated pairs of each field on a 65 by 65 lattice and correlation parameters  $(\nu, \eta) = (0.5, 20)$ .

Figures 127 - 130 show that the distance between the components (or holes) from simulations on pairs of fields from the same distributions vary depending on what that distribution is, even if they have the same correlation structure. In particular the  $\chi_1^2$  field has much larger distances between components and holes than the others, and this difference is more noticeable as the local correlations on the field increases. This pattern is similar for all distance metrics described in Section 8.4.1 (results not shown). This is particularly notable when using the metric in Equation (96), which corrects for differing numbers of points in each pattern. The similarity between the

results whilst using the different metrics, and choices for  $p$  and  $c$  (where applicable) is due to pairs of the same fields having an approximately equal number of components or holes in each case.

Next, we will repeat the above but for each simulation look at the distance between the components of different fields. For example, simulate 100 pairs of Gaussian and  $\chi_1^2$  fields, each time computing the minimum distance between the components. This will allow us again to plot the cdfs and see if there are any notable difference between the distance between the components of two Gaussian fields, and a Gaussian and some other distribution. Note that in this example, the number of components (or holes) will vary between the point patterns we are comparing, as the two point patterns will come from different distributions. Therefore we will present results using two of the distance metrics to highlight the importance of correcting for the different number of points. Firstly we use the metric (94) which does not correct for the differering number of points, and then metric (97) which does. Figure 131 shows the results using the distance metric (94).

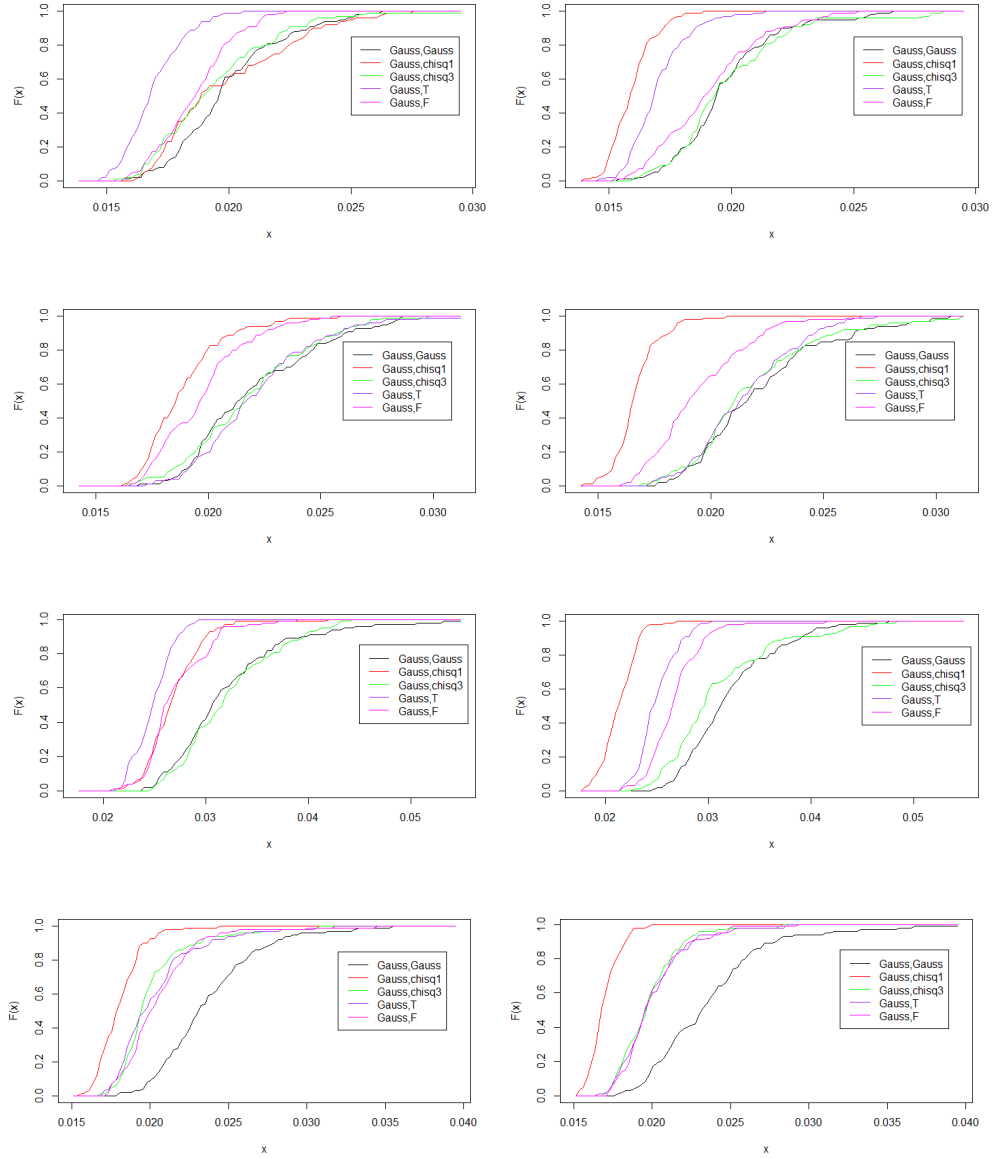


Figure 131: The cumulative distribution function of the minimum distances (metric (94)) between components (left) and holes (right) of 100 simulated pairs of each field on a 65 by 65 lattice and correlation parameters (top to bottom)  $(\nu, \eta) = (0.5, 2), (0.5, 5), (1, 5), (0.5, 20)$ .



Figure 131 shows that the distance between components on pairs of Gaussian fields and a Gaussian and some other fields is different. This difference is more notable again as correlation length increases. An interesting thing to note is that the distance between the components on pairs of Gaussian fields is typically bigger than the distance between the components of a Gaussian and some other type of field. This is due to the difference in the number of components on fields with different underlying distributions. Therefore when using the metric (94) which does not account for the differences, it becomes easier to find a mapping between the two sets of points which can give a smaller distance. Hence we obtain a counter-intuitive result. Figure 132 shows the results obtained by using the distance metric in Equation (97) which corrects for differing number of points in each pattern.

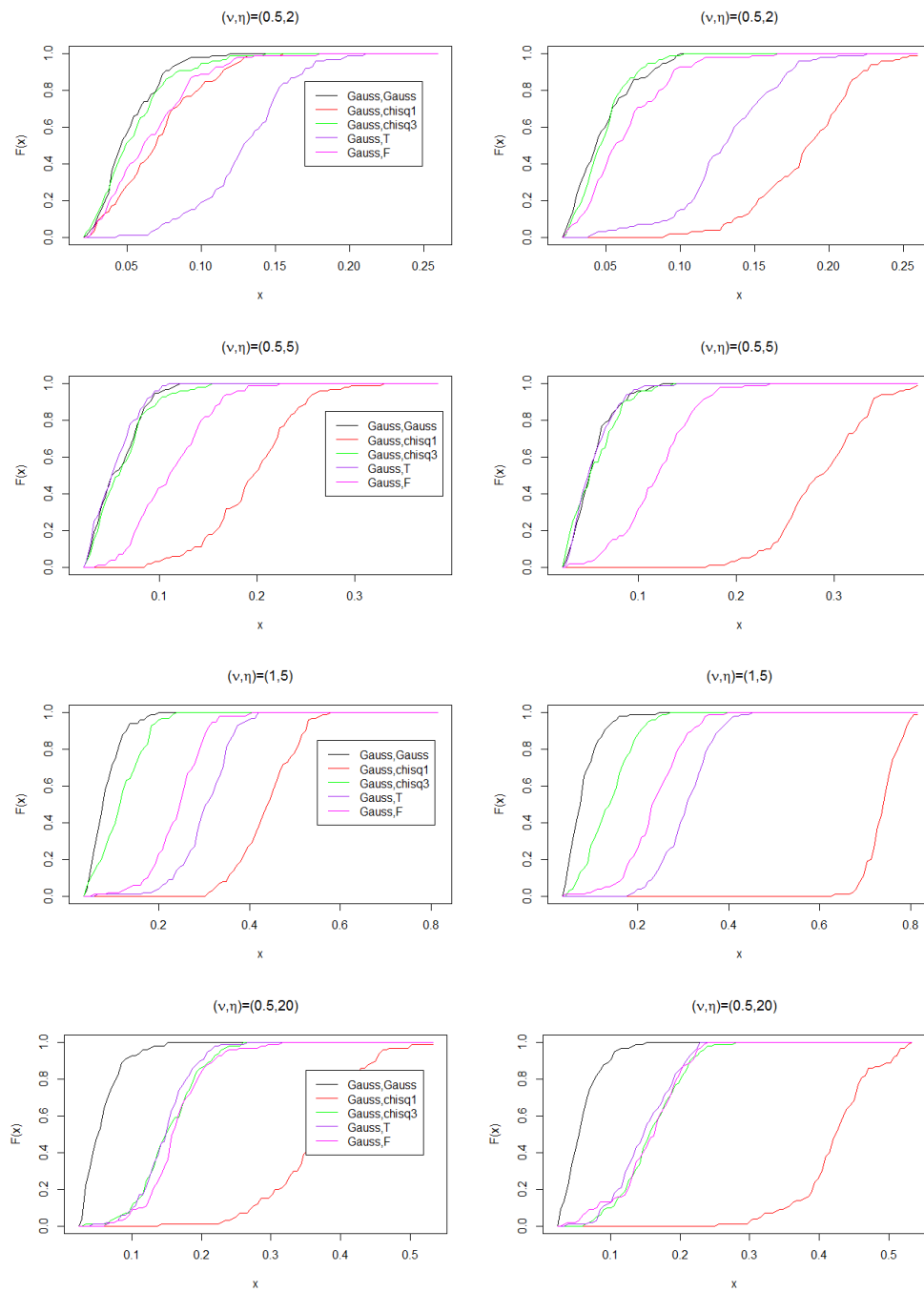


Figure 132: As in Figure 131 but using the distance metric (97).

Figure 132 shows that we now obtain the results we would expect, the distance between components on pairs of Gaussian fields are smaller than the distance between components on the Gaussian and components on one of the other fields. Note that similar results were found using metric (96) and for various sets of  $(p, c)$ . All combinations of  $p = (1, 2, 3)$  and  $c = (0, 0.25, 0.5, 0.75, 1)$  were investigated, with little difference in conclusions.

Next, we will use the same method to compare the distance between components for pairs of Gaussian fields, for the four different sets of parameter choices.

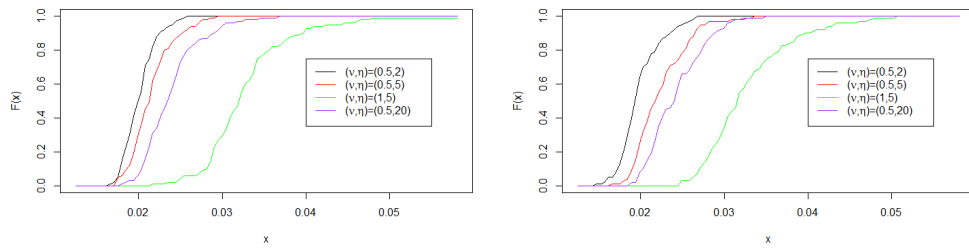


Figure 133: The cumulative distribution function of the minimum distances between components (left) and holes (right) of 100 simulated pairs of Gaussian fields on a 65 by 65 lattice and with different correlation parameters.

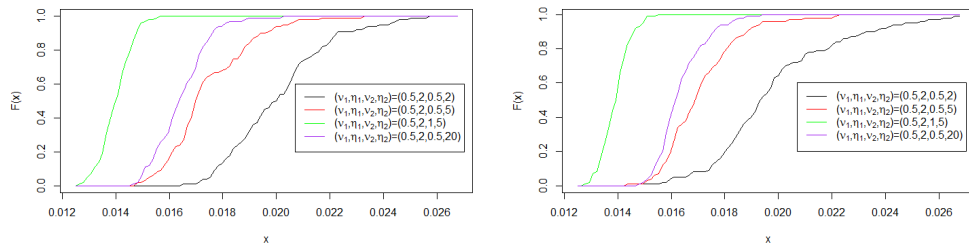


Figure 134: The cumulative distribution function of the minimum distances between components (left) and holes (right) of 100 simulated pairs of Gaussian fields on a 65 by 65 lattice and with different correlation parameters.

Figures 133 and 134 show that as local correlation on the field increases, then so does the distance between components (or holes) on pairs of simulated

Gaussian fields. However, the distance between components (or holes) on pairs of two Gaussian fields with different sets of correlation parameters is less than if the parameters on both fields were the same (for the reference case where  $(\nu, \eta) = (0.5, 2)$ ). Again this can be attributed to the large difference in number of points between fields with different correlation parameters. Figure 135 shows results when the distance metric is corrected for the differing number of points.

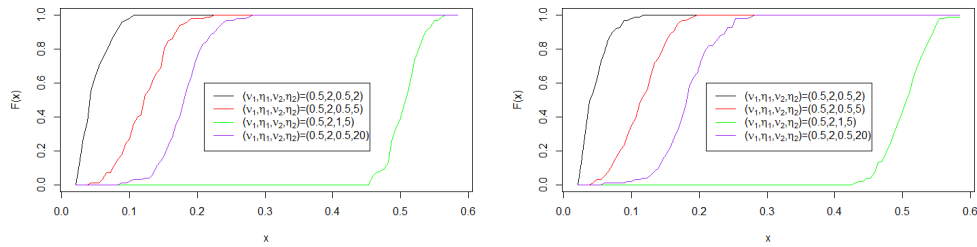


Figure 135: The cumulative distribution function of the minimum distances between components (left) and holes (right) of 100 simulated pairs of Gaussian fields on a 65 by 65 lattice and with different correlation parameters. Here the distance metric is Equation (96) which corrects for different numbers of points. Results show case with  $p = 1$ ,  $c = 1$ .

Figure 135 again shows that the use of the distance metric in Equation (96) which corrects for the different numbers of points in the patterns has given results which are as expected. That is, the distance between components (or holes) on two Gaussian fields with parameters  $(\nu, \eta) = (0.5, 2)$  are closer to each other than the distance between components (or holes) where we have one Gaussian field with parameters  $(\nu, \eta) = (0.5, 2)$  and one with some set of correlation parameters.

Finally, we will look at the difference between the distance between components and holes on the five different fields.

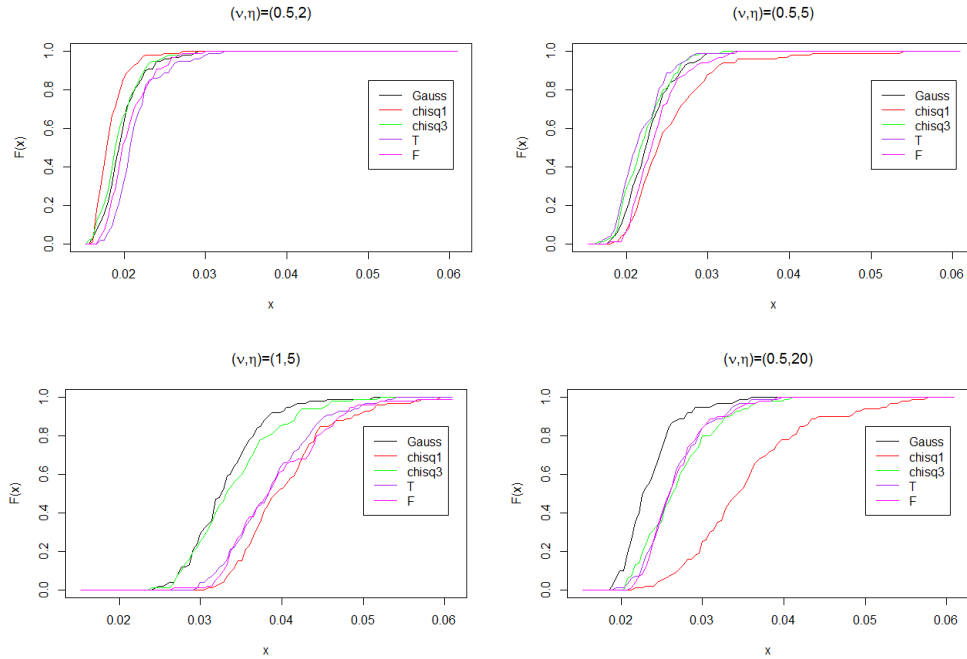


Figure 136: The cumulative distribution function of the minimum distances between components and holes of 100 simulated pairs of fields on a 65 by 65 lattice and with different correlation parameters.

Figure 136 shows that the difference between components and holes on pairs of fields is notable, especially when the local correlation increases, where the Gaussian has the smallest distance between components and holes, and as before the  $\chi_1^2$  field has large difference between the components and holes.

In this section we have seen that the distance between components (or holes) on pairs of different fields, all marginally normalised and with the same correlation structure, appear to distinguish between the distributions. Furthermore, the distance between components on pairs of Gaussian fields is closer than the distance between the components of a Gaussian and some other field. It has also been demonstrated that the distance metrics given in Equation (96) and (97) are very important when the number of points in the two point patterns being investigated are different, so that any differences are not based different number of points, but genuine differences in the spatial distribution of the points.

Although the cumulative distribution functions appear to show differ-

ences, it is difficult to assess the significance of these differences. One method is to repeat the simulations many times to obtain many cdf's from which we could determine the significance of the overlap between the distributions. However this is very computationally expensive. In the next section we will use an alternative method to assess the similarity and differences between the distances of all the comparisons made in this section.

### 8.6.2 Minimum Distance Between Point Patterns - Probability Functions

In this section, we will further investigate the similarity between point patterns of different distributions. In contrast to the work in Section 8.6.1, we will look at the pdfs of the distances between point patterns as opposed to the cdfs.

The first questions we will consider as before is to investigate if distances between the point pattern of the positions of the components (or holes) of two Gaussian fields are similar or different to a pair of  $\chi_1^2$ ,  $\chi_3^2$ ,  $T$  or  $F$  fields, when all the fields are simulated such that the target correlation on each are approximately equal.

The method will be to simulate  $j = 1 - 1000$  pairs of Gaussian fields, and for each pair compute a distance,  $D_j$ , between the point patterns of the components (or holes) using the distance metric given in Equation (97). Then for each of the other distributions,  $i = 1 - 100$  pairs of fields will be simulated, and 100 distances,  $d_i$  between the point patterns computed. We then define the  $p$ -value,  $p_i$ , for distance  $d_i$  of the alternative distribution as

$$p_i = \frac{\sum_{j=1}^{1000} I(D_j > d_i)}{1000} \quad (104)$$

where  $I(D_j > d_i) = 1$  if  $D_j > d_i$  and zero otherwise. If the distance  $d_i$  from the non-Gaussian distribution is typical of a distance between a pair of Gaussian distributions, then we expect  $p_i$  to be larger than some critical value (here 0.05). Alternatively if  $p_i < 0.05$  we can conclude that the distance is significantly different from what is expected from a pair of Gaussian fields.

The  $p$ -value  $p_i$  will be computed for  $i = 1 - 100$  to obtain a set of  $p$ -values. Then the proportion of the  $p$ -values which are less than 0.05 is the estimated power of the test, ie the probability of detecting a non-Gaussian pair of distributions when the distance is from a pair of non-Gaussian distributions.

$$Power = \frac{\sum_{i=1}^{100} I(p_i < 0.05)}{100}. \quad (105)$$

Parameters	$\chi_1^2$	$\chi_3^2$	$T_3$	$F_{3,3}$
(0.5, 2)	0.13/0.16	0.05/0.03	0.05/0.05	0.09/0.09
(0.5, 5)	0.29/0.31	0.08/0.06	0.04/0.06	0.13/0.07
(1, 5)	0.48/0.58	0.06/0.11	0.14/0.13	0.07/0.12
(0.5, 20)	0.53/0.39	0.10/0.05	0.08/0.14	0.10/0.09

Table 57: The power obtained from comparing the components/holes of 100 pairs of simulated fields from different distributions on 65 by 65 lattices to 1000 Gaussian distances.

Table 57 shows that as local correlation on the fields increases, the power also increases. This is in agreement with results found in Section 8.6.1. However, in all cases the power is low, which suggests that the differences between the distributions are not consistently detected. To confirm these figures, Figure 137 shows histograms of the 1000 Gaussian distances, and 100  $\chi_1^2$  distances when  $(\nu, \eta) = (0.5, 2)$ .

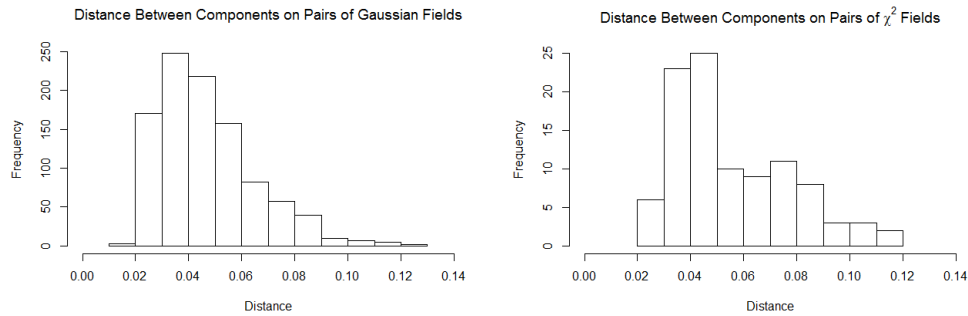


Figure 137: Histograms of the distances between components on pairs of Gaussian fields (left) and  $\chi_1^2$  fields (right), when  $(\nu, \eta) = (0.5, 2)$

From Figure 137 it can be seen that the two histograms are quite similar. In fact, 95% of the Gaussian distances are less than 0.08, and from the histogram on the right it can be seen that very few of the  $\chi_1^2$  distances are

larger than 0.08 (13%), and so in most cases we would conclude that the  $\chi_1^2$  distances are consistent with the Gaussian distribution of distances, thus the power is small. Histograms for the other distributions were also consistent with the results in Table 57.

Table 58 shows the power when comparing distances between components (or holes) from pairs of Gaussian fields, to the distances between a Gaussian field and another distribution (see Figure 132).

Parameters	$\chi_1^2$	$\chi_3^2$	$T_3$	$F_{3,3}$
(0.5, 2)	0.34/0.98	0.09/0.04	0.95/0.90	0.20/0.28
(0.5, 5)	0.99/1.00	0.08/0.10	0.05/0.09	0.56/0.57
(1, 5)	1.00/1.00	0.21/0.44	1.00/1.00	0.96/0.93
(0.5, 20)	0.99/1.00	0.18/0.16	0.61/0.63	0.26/0.17

Table 58: The power obtained from comparing the components/holes of 100 pairs of simulated fields, one a Gaussian field, the other a field from a different distribution, on 65 by 65 lattices to 1000 distances obtained from pairs of Gaussian fields.

Table 58 shows that in almost all cases the distance between components or holes on two Gaussian fields can be distinguished from the distance between components or holes on a Gaussian and a  $\chi_1^2$  field. It can also be seen that distinctions can be made for some correlation parameters between the Gauss-Gauss and Gauss- $T$  or Gauss- $F$  fields.

Table 59 compares the point patterns of the components and holes on pairs of Gaussian fields with correlation parameters  $(\nu, \eta) = (0.5, 2)$  and pairs of Gaussian fields with the other three sets of correlation parameters.

Parameters	Power
(0.5, 5)	0.08/0.05
(1, 5)	0.40/0.33
(0.5, 20)	0.20/0.16

Table 59: The power obtained from comparing the components/holes of 100 pairs of Gaussian simulated fields with different correlation parameters on 65 by 65 lattices to 1000 distances obtained from pairs of Gaussian fields with parameters  $(\nu, \eta) = (0.5, 2)$ .



Table 59 shows that there is little evidence that we can consistently distinguish between pairs of Gaussian fields based on the locations of the components and holes.

### 8.6.3 Quadrant Count Analysis

In this section we will implement the theory in Section 8.4.2 and use a  $\chi^2$  association test based on quadrant counts to compare fields.

Table 60 shows the proportion of  $p$ -values with value less than or equal to 0.05 obtained from 100 simulations of pairs of the various fields, where in all cases the correlation parameters are Matérn  $(\nu, \eta) = (1, 5)$ . Hence the table shows the probability of rejecting the null hypothesis that the two point patterns are from the same distribution. Therefore, we would expect the numbers on the diagonal to be small, and for a high enough number of simulations to be approximately 0.05, and the off diagonal elements are the power of the test.

	Gauss	$\chi_1^2$	$\chi_3^2$	$T_3$	$F_{3,3}$
Gauss	0.00/0.00				
$\chi_1^2$	0.46/0.01	0.78/0.04			
$\chi_3^2$	0.02/0.00	0.45/0.02	0.00/0.02		
$T_3$	0.02/0.00	0.44/0.02	0.01/0.00	0.01/0.01	
$F_{3,3}$	0.01/0.00	0.40/0.02	0.02/0.00	0.00/0.02	0.02/0.00

Table 60: The proportion of  $p$ -values (obtained from an association test using 25 quadrants on a 65 by 65 lattice) less than or equal to 0.05 obtained from 100 simulations of pairs of the various fields. Results obtained using the point patterns of components/holes of pairs of fields with Matérn  $(\nu, \eta) = (1, 5)$  correlation.

Table 60 shows a lot of very small values. For example, using quadrant counts to compare the position of components (or holes) of two Gaussian fields did not give any  $p$ -values less than 0.05 from the 100 simulations. From Figures 121 and 122 it can be seen that the distribution of components and holes are similar from different random realisations of the Gaussian field, therefore the quadrant analysis always gives a very high  $p$ -value ( $> 0.9$ ) when making the comparisons.

Table 60 also shows interesting results for the cases where one of the distributions is  $\chi_1^2$  and the point pattern is that of the components. Here it can be seen that the probability of detecting that the two distributions are different is almost 0.5 in all cases. What is more interesting is that the probability of rejecting that both distributions are  $\chi_1^2$  is high, here 0.78. These results are not true when considering the holes. This result suggests that although the number of components on the  $\chi_1^2$  field remain fairly constant across simulations (see results in Section 7.2.3), they vary in location and are not evenly distributed across the quadrants, thus some clustering takes place. This is not true or noticeable if the local correlations are smaller, as similar results were not found in these cases. Results were obtained for the three other sets of correlation parameters, though the results here are the most extreme, and the other cases had all values of power less than 0.05 for every comparison.

Figure 138 shows some heatmaps of the  $\chi_1^2$  fields for different correlation parameters. From this it is noticeable that as local correlation increases, patterns emerge on the field which explains the association test results. For example, for  $(\nu, \eta) = (0.5, 2)$ , the components are fairly evenly distributed across the grid, however when  $(\nu, \eta) = (1, 5)$ , there are definite patterns, or clustering of components on the grid. Figure 139 shows four examples of the  $\chi_1^2$  fields when  $(\nu, \eta) = (1, 5)$ . From the figure it can be seen that the components (dark red) are not consistent across different simulations, which explains the results in Table 60, ie the position of the components are different across different simulations, but the qualitative patterns are the same. This is indicative of an inhomogeneous process such as the Cox process [35] [58] where the point intensity is itself a random process which can change over time, or in this case space.

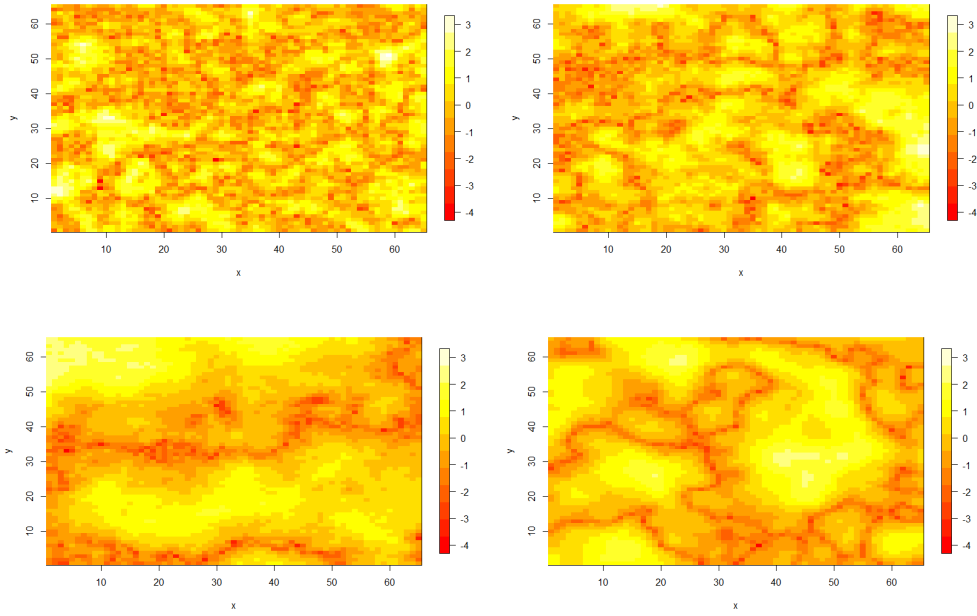


Figure 138: Heatmap images of simulated data from the  $\chi_1^2$  distribution with correlation parameters  $(\nu, \eta) = (0.5, 2)$  (top left),  $(\nu, \eta) = (0.5, 5)$  (top right),  $(\nu, \eta) = (0.5, 20)$  (bottom left) and  $(\nu, \eta) = (1, 5)$  (bottom right).

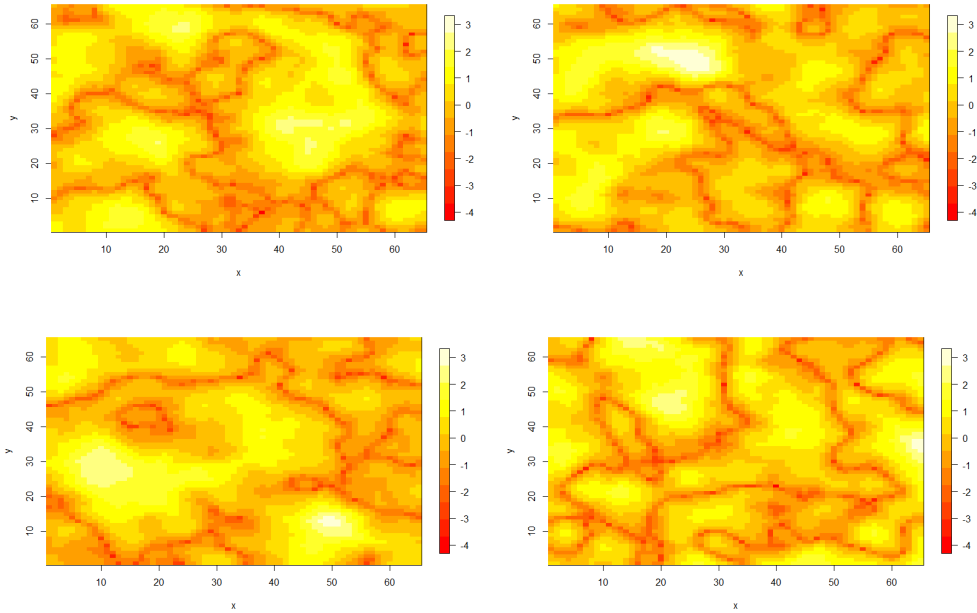


Figure 139: Examples of heatmap images of simulated data from the  $\chi_1^2$  distribution with correlation parameters  $(\nu, \eta) = (1, 5)$ .

Comparisons were also made between the point patterns of the components or holes of pairs of Gaussian fields with different correlation parameters. For all possible combinations of the four sets of parameters, 100 pairs of Gaussian fields were simulated and the power calculated. In all cases the power was less than 0.05 suggesting that no distinction can be made between Gaussian fields with different correlation parameters.

$(\nu, \eta) =$	(0.5, 2)	(0.5, 5)	(1, 5)	(0.5, 20)
Gauss	0.00	0.00	0.01	0.00
$\chi_1^2$	0.00	0.00	0.26	0.10
$\chi_3^2$	0.00	0.00	0.00	0.00
$T_3$	0.00	0.00	0.01	0.00
$F_{3,3}$	0.00	0.00	0.00	0.01

Table 61: The proportion of  $p$ -values (obtained from an association test using 25 quadrants on a 65 by 65 lattice) less than or equal to 0.05 obtained from 100 simulations of pairs of the various fields. Results obtained by comparing the point patterns of components and holes of pairs of fields.

Table 61 shows a comparison between components and holes on the various fields, for the various correlation parameters. From the table, it can be seen that there is very little distinction between the point patterns of the components and holes on the various fields. Again refer to Figures 121 and 122 to see the similarity between the point patterns of the components (or holes) from different random realisations of the various fields. Quadrant count analysis, using 25 quadrants on the field of size 65 by 65 can not detect differences between the point patterns.

The simulations in this section show that using an association test on quadrant counts, is not an effective method to detect differences amongst distributions which have the same correlation structure. However, like the other methods, it can detect the unusual distribution of components on the  $\chi_1^2$  field.

#### 8.6.4 Cross K-function

An alternative method of comparing two point patterns, this time at different scales, is the cross K-function as described in Section 8.4. Again simulations will be used on a two dimensional lattice to show how the cross K-function can detect attraction, repulsion or independence between two point patterns.

Let us consider two spatial point patterns, both on a lattice of size 65 by 65 (4225 points). Let the size of the first point pattern be  $n_1 = 400$  and the second pattern  $n_2 = 400$  (similar to the number of components and holes on the brain slice of similar size). Therefore here  $\lambda_1 = \lambda_2 = 400/4225 = 0.09$ . The following examples will show how the estimated cross K-function

(101) compares with the expected (100), which gives a sense of attraction or repulsion between the two point patterns. Also shown will be a plot with a confidence band to assess the significance of any departure from independence as per the details in Section 8.4.4.

The first example, seen in Figure 140, shows an extreme example of attraction between two sets of points.

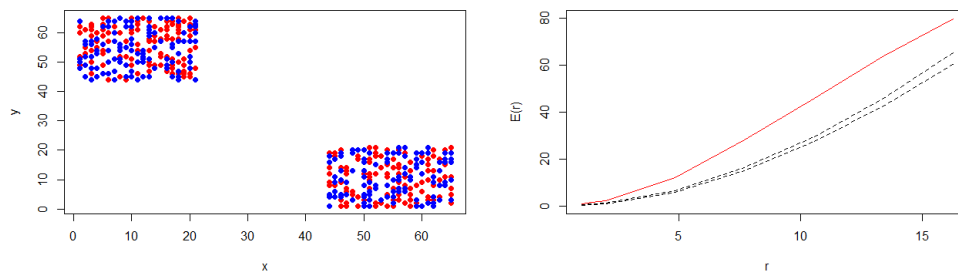


Figure 140: Two set of attracted points (left), observed cross k-function (right - red) and expected under randomness (right-black). Also shown is the confidence envelope under independence (black dashed).

Figure 140 shows two sets of points, red and blue, which are positioned (attracted) near each other in the space. The cross K-function shows that there are more red points close to blue points than expected under spatial randomness, ie if the red points were randomly distributed across the full lattice. The confidence envelope is the result of  $N = 99$  permutations of both sets of points across the space, and so as the observed cross K-function falls outside and above this envelope, this shows that the attraction is significant. Therefore this example shows that if the two point patterns are positioned close to each other, the cross K-function can detect the attraction.

Figure 141 shows the results for two sets of points which are simulated such that they are repulsed, and Figure 142 shows the results for two sets of independent points.

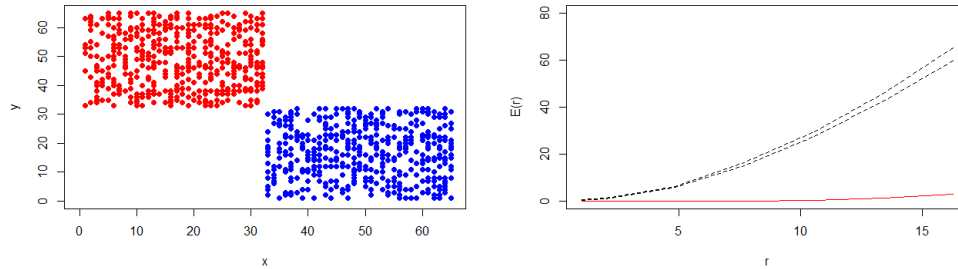


Figure 141: Two set of repulsed points (left), observed cross K-function (right - red) and expected under randomness (right-black). Also shown is the confidence envelope under independence (black dashed).

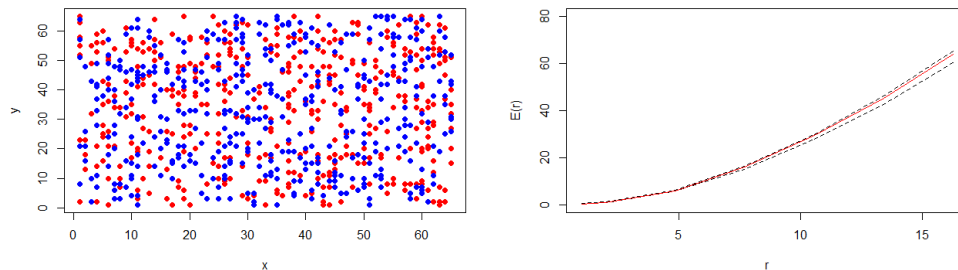


Figure 142: Two set of random points (left), observed cross K-function (right - red) and expected under randomness (right-black). Also shown is the confidence envelope under independence (black dashed).

Figure 141 shows that when the two sets of points are repulsed, the observed cross K-function is well below that which is expected under spatial randomness, and Figure 142 shows that the observed cross K-function is within the confidence band for spatial randomness, when the two point patterns are randomly simulated.

Figures 143 and 144 show some comparisons of observed cross K-functions for the components and holes of pairs of fields from different distributions as described in Section 7.2.1. In particular an interval obtained from the minimum and maximum cross K-function of 100 pairs of Gaussian fields is

shown as well as the cross  $K$ -function from the other distributions. From the plots it can be seen that for some correlation parameters, in particular for cases where the local correlation increases, differences can be seen between the distributions.



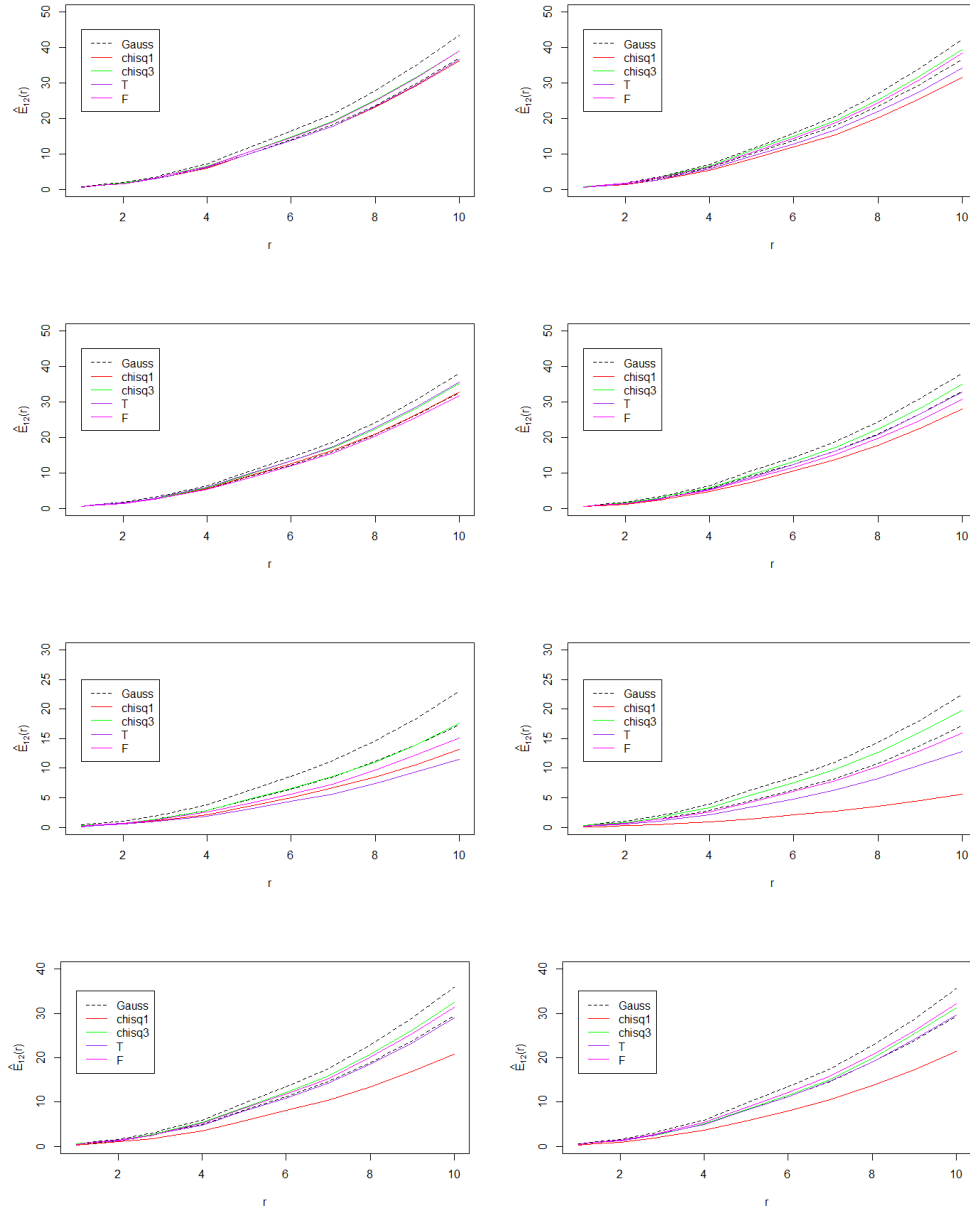


Figure 143: The cross K-function from the point pattern of the components (left) and holes (right) of pairs of each of the distributions (black dashed is Gaussian interval from 100 simulations) on a 65 by 65 lattice. All distributions are marginally normal with similar correlation. Top to bottom: Gaussian Matérn  $(\nu, \eta) = (0.5, 2)$ ,  $(\nu, \eta) = (0.5, 5)$ ,  $(\nu, \eta) = (1, 5)$ ,  $(\nu, \eta) = (0.5, 20)$ .

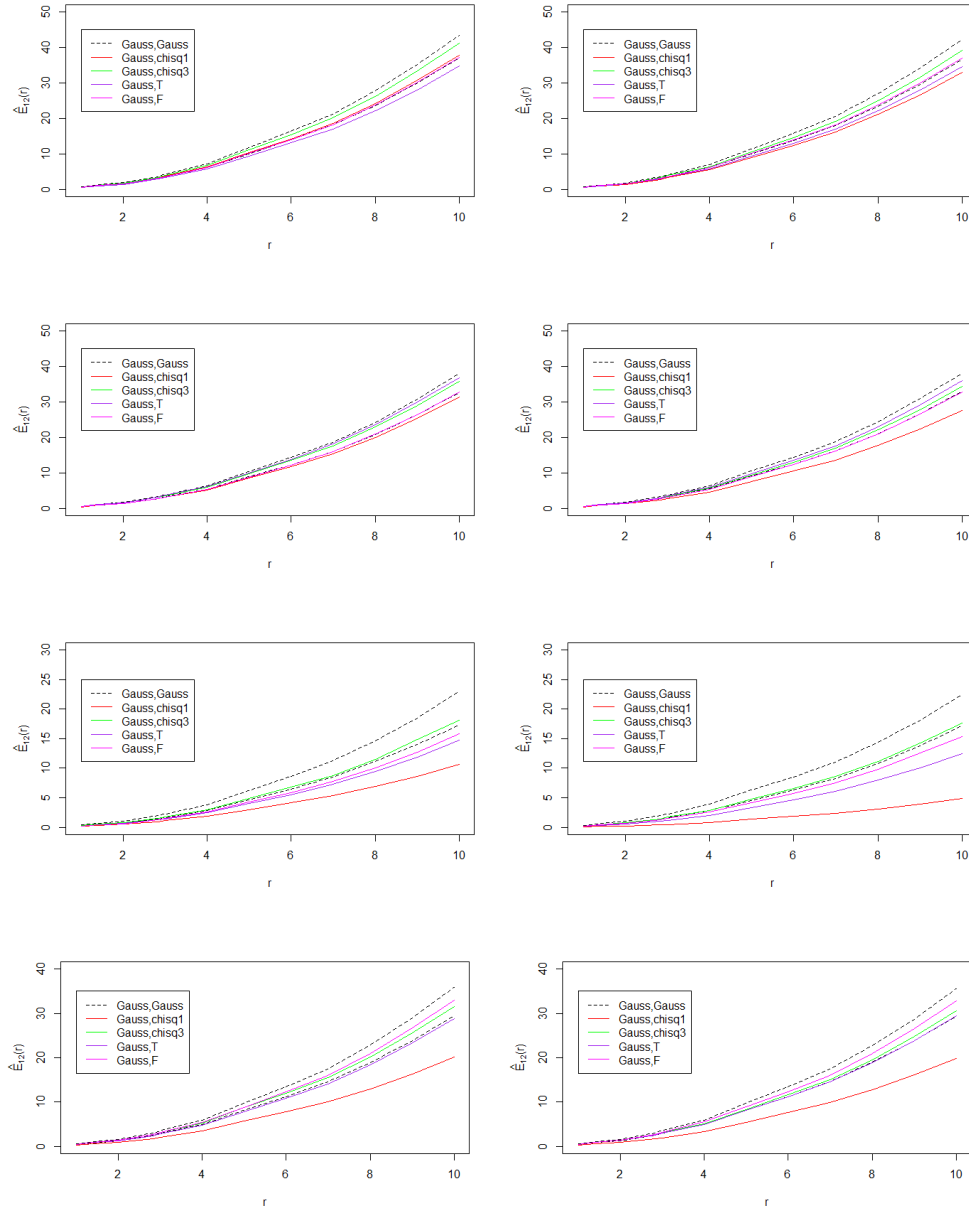


Figure 144: The cross K-function from the point pattern of the components (left) and holes (right) of pairs of each of the distributions (black dashed is Gaussian interval from 100 simulations) on a 65 by 65 lattice. All distributions are marginally normal with similar correlation. Top to bottom: Gaussian Matérn  $(\nu, \eta) = (0.5, 2)$ ,  $(\nu, \eta) = (0.5, 5)$ ,  $(\nu, \eta) = (1, 5)$ ,  $(\nu, \eta) = (0.5, 20)$ .

## 8.7 Application I - Complete Spatial Randomness

In this section, the work of the previous sections will be applied to some two dimensional brain slices. As in previous sections, results for slices  $z = 42$ ,  $z = 60$  and  $z = 70$  will be shown. For a given brain slice we will initially apply the quadrant count analysis to test for complete spatial randomness. Then the K-function and radial distribution function will be calculated, and simulations of random data on the slice will again be used to obtain a confidence envelope. The locations of components and then holes on the brain slice will be used to assess if the components and holes are randomly distributed across the brain slice.

### 8.7.1 Quadrant Count Analysis On Brain Slices

Figure 145 shows the position of the components and holes of the model residual data at  $t = 1$  on brain slice  $z = 42$  (see Figure 101 for image of full brain slice.)

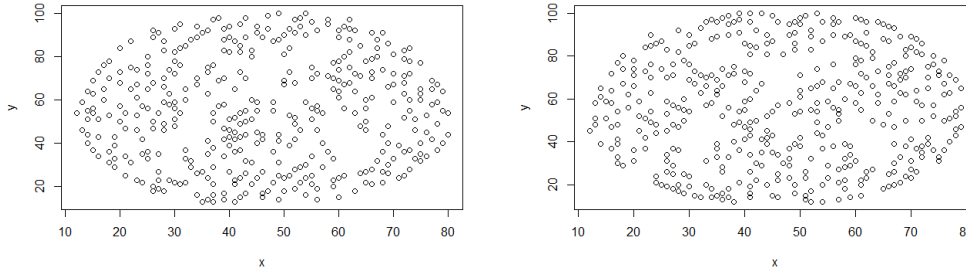


Figure 145: The position of model residuals components (left) and holes (right) on brain slice  $z = 42$  at  $t = 1$ .

From Figure 145, the components and holes of the model residuals do not appear to be clustered in any way. There are a total of 3719 lattice points on slice  $z = 42$ , 382 components and 404 holes at  $t = 1$ , hence here  $\lambda = 0.10$  to two decimal places in both cases.

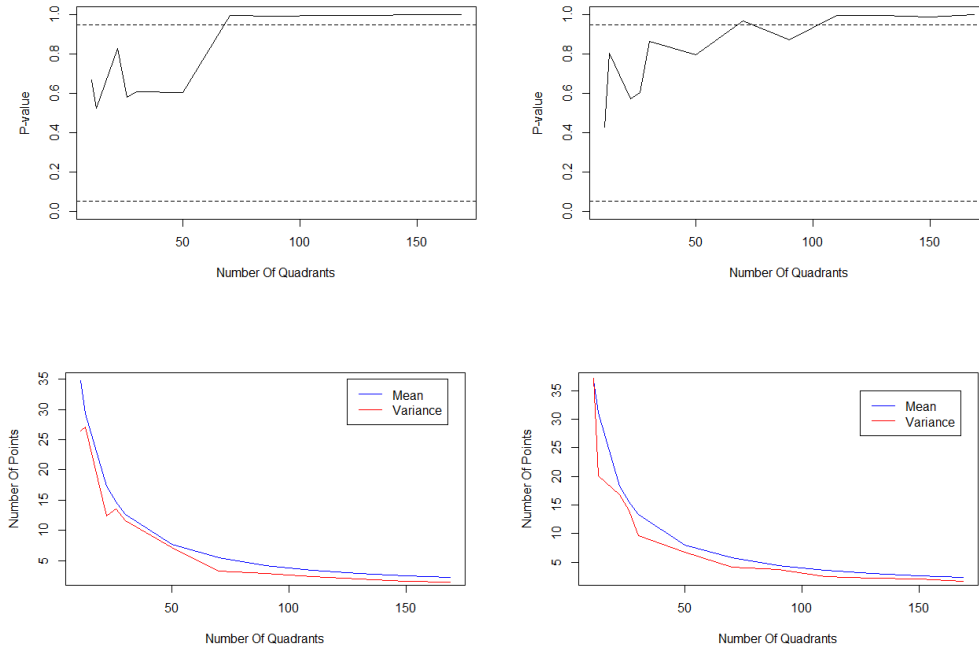


Figure 146: The  $\chi^2$   $p$ -values (top row) and mean and variance (bottom row) for the counts of components (left) and holes (right) of model residuals on brain slice  $z = 42$  at  $t = 1$ .

Figure 146 shows the results of quadrant counts on this slice at the first time point. Here the association test  $p$ -value is shown for several different numbers of quadrants. It can be seen that the counts are consistent with random points. The  $p$ -values are greater than 0.05 for all quadrant sizes considered and the mean and variance plots show that the points are consistent with a homogeneous Poisson point process, in the sense that the mean and variance of the counts are almost equal.

Following are results from two more brain slices. First let us consider brain slice  $z = 60$  (see Figure 101) which has 3007 points, with 290 observed components and 294 holes at  $t = 1$ . Hence here  $\lambda = 0.10$ . Figure 147 shows the position of components and holes on the slice.

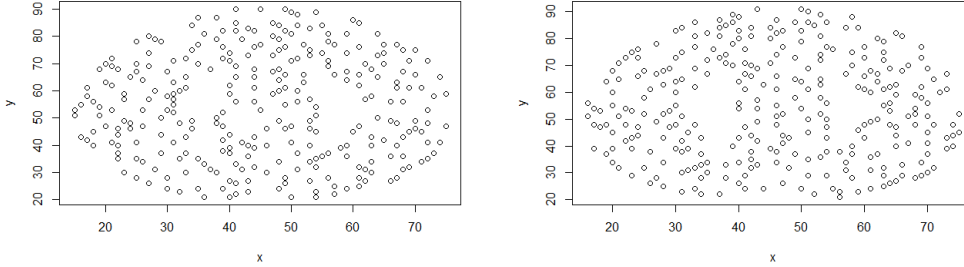


Figure 147: The position of model residuals components (left) and holes (right) on brain slice  $z = 60$  at  $t = 1$ .

Again Figure 147 shows no sign of clustering. Figure 148 shows the quadrant analysis results.

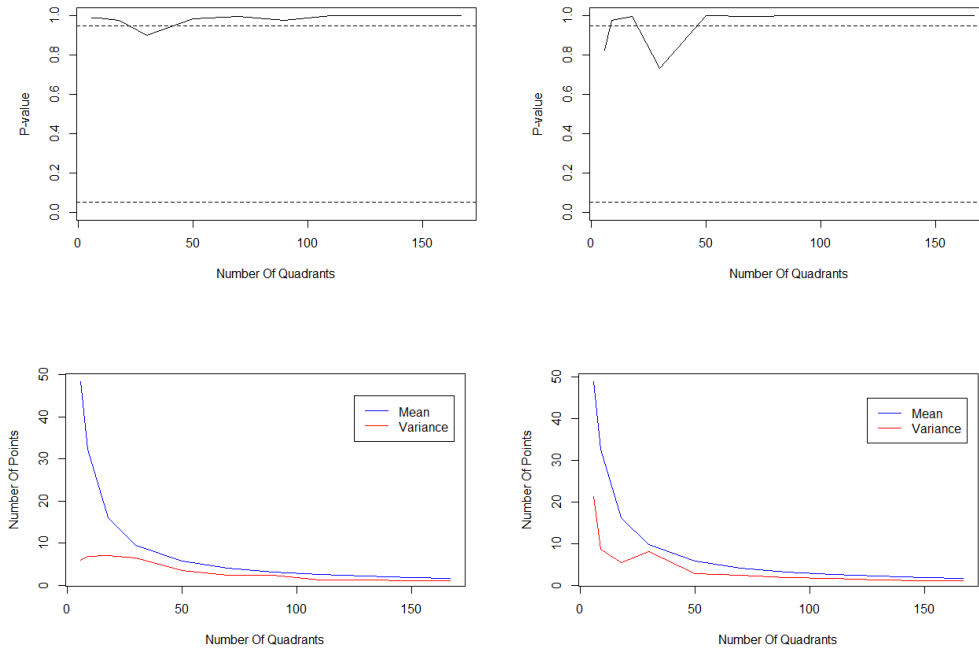


Figure 148: The  $\chi^2$   $p$ -values (top row) and mean and variance (bottom row) for the counts of components (left) and holes (right) of model residuals on brain slice  $z = 60$  at  $t = 1$ .

Figure 148 shows that there is evidence of regularity in the point patterns on slice  $z = 60$  at  $t = 1$ .

Finally let us consider brain slice  $z = 70$  (see Figure 101) which has 1633 points, with 154 observed components and 142 holes at  $t = 1$ . Hence here  $\lambda = 0.09$ . Figure 149 shows the position of components and holes on the slice.

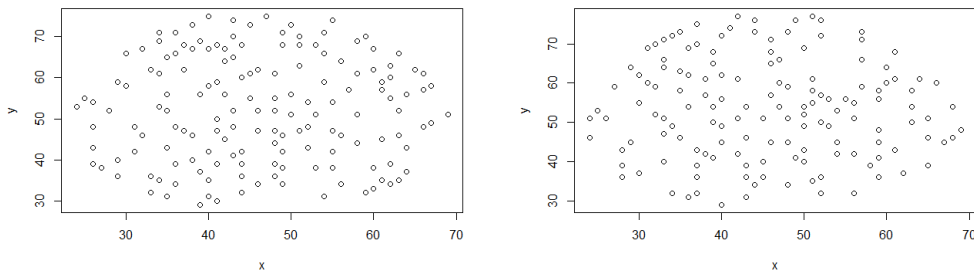


Figure 149: The position of model residuals components (left) and holes (right) on brain slice  $z = 70$  at  $t = 1$ .

Figure 150 shows the quadrant analysis, and shows that there is evidence of regularity in the point patterns on slice  $z = 70$  at  $t = 1$ .

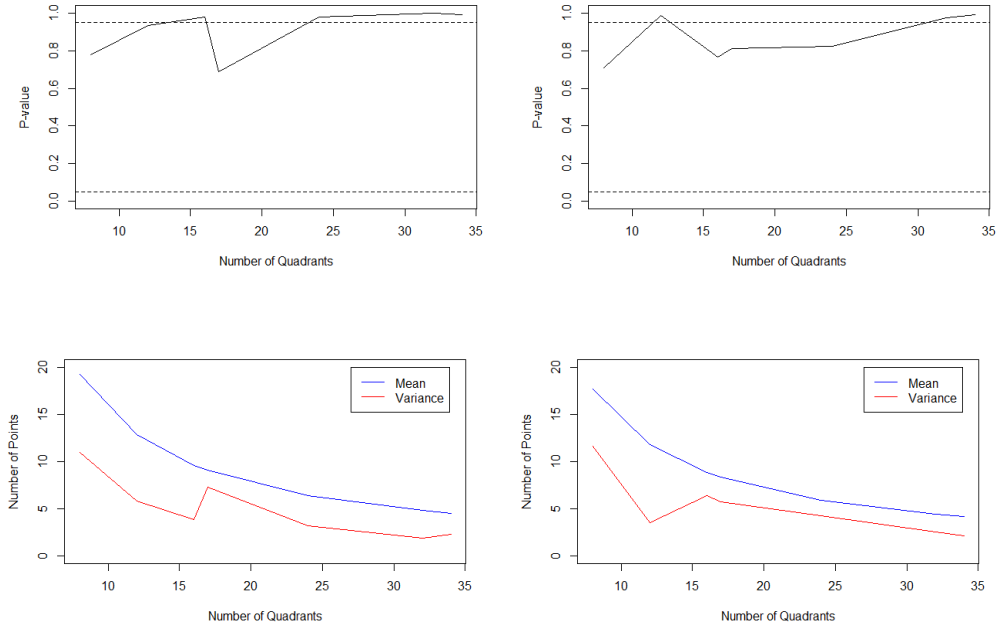


Figure 150: The  $\chi^2$   $p$ -values (top row) and mean and variance (bottom row) for the counts of components (left) and holes (right) of model residuals on brain slice  $z = 60$  at  $t = 1$ .

In this section we have tested for homogeneity on three different brain slices using quadrant count analysis and found that there is no evidence of clustering of the position of components and holes on these brain slices. However the quadrant test showed some evidence for regularity of points for the slices  $z = 60$  and  $z = 70$ . Although only  $t = 1$  results are shown, in all cases similar results were found for other time points.

### 8.7.2 K-function On Brain Slices

Figure 151 shows K-function results for the slice  $z = 42$  at  $t = 1$ .

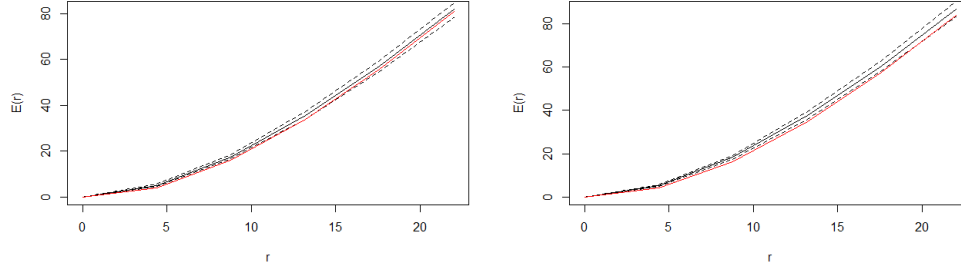


Figure 151:  $E(r)$  for the brain slice  $z = 42$  (black) with confidence envelope from 50 simulations (dashed), and  $\hat{E}(r)$  (red) obtained from the 382 components (left) and 404 holes (right) on the slice at  $t = 1$ .

Figure 151 shows the comparison of  $E(r)$  and  $\hat{E}(r)$ . From this it can be seen that  $\hat{E}(r)$ , obtained from the position of components and holes on the slice, is within the confidence envelope of  $E(r)$ , therefore the components and holes on slice  $z = 42$  at  $t = 1$  seem to be randomly distributed across the slice. Although the result at  $t = 1$  is presented here, very similar plots were found for all time points on the slice. Therefore there is no evidence of clustering of components or holes on the brain slice  $z = 42$ . Figure 152 shows the results of the pairwise correlation function applied to the data.



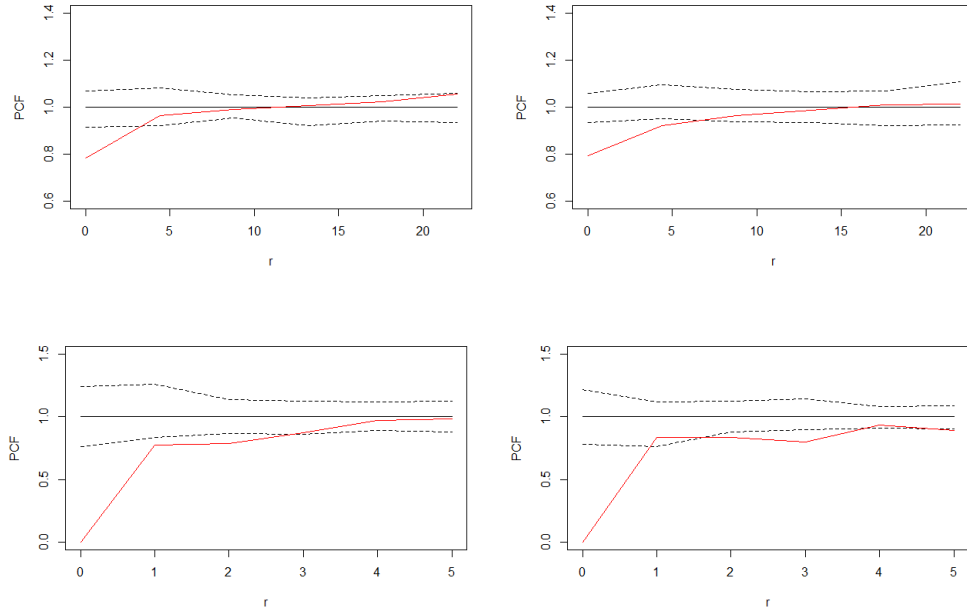


Figure 152: The  $PCF$  for the brain slice  $z = 42$  (black) with confidence envelope from 50 simulations (dashed), and  $\hat{E}_g(r)$  (red) obtained from the 382 components (left) and 404 holes (right) on the slice at  $t = 1$ . Bottom plots show results for smaller distances.

Figure 152 shows that for small distances, there is evidence of regularity in position of components and holes on the slice, ie fewer points are observed than expected when considering intervals of width one up to approximately  $r = 4$ , and thereafter the data in the intervals are consistent with a random point pattern. This is probably due to the lattice structure of the data and that components (or holes) cannot be adjacent to one another on the lattice.

Figures 153 and 155 shows the corresponding results for slices  $z = 60$  and  $z = 70$  respectively.

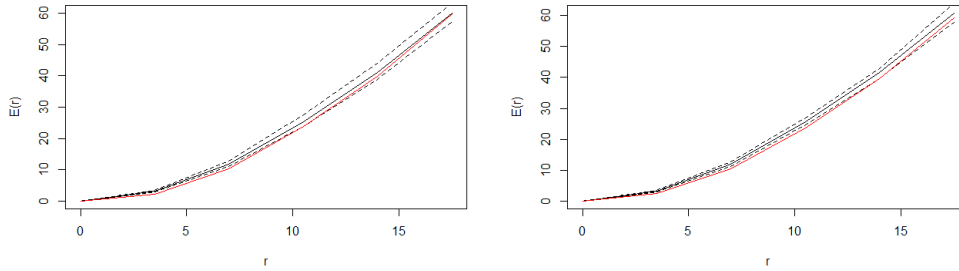


Figure 153:  $E(r)$  for the brain slice  $z = 60$  (black) with confidence envelope from 50 simulations (dashed), and  $\hat{E}(r)$  obtained from the 290 components (left) and 294 holes (right) on the slice at  $t = 1$ .

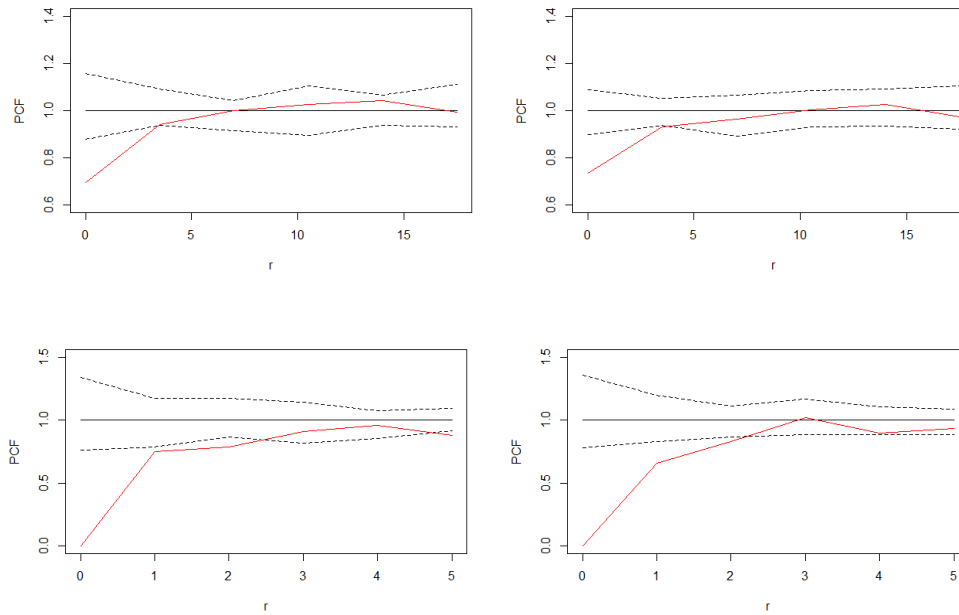


Figure 154: The  $PCF$  for the brain slice  $z = 60$  (black) with confidence envelope from 50 simulations (dashed), and  $\hat{E}_g(r)$  (red) obtained from the 290 components (left) and 294 holes (right) on the slice at  $t = 1$ . Bottom plots show results for smaller distances.

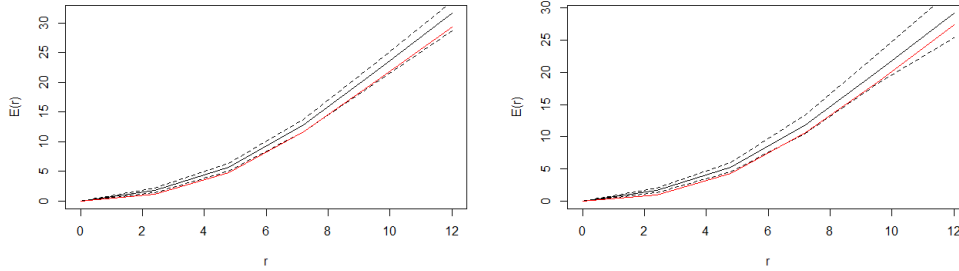


Figure 155:  $E(r)$  for the brain slice  $z = 70$  (black) with confidence envelope from 50 simulations (dashed), and  $\hat{E}(r)$  obtained from the 154 components (left) and 142 holes (right) on the slice at  $t = 1$ .

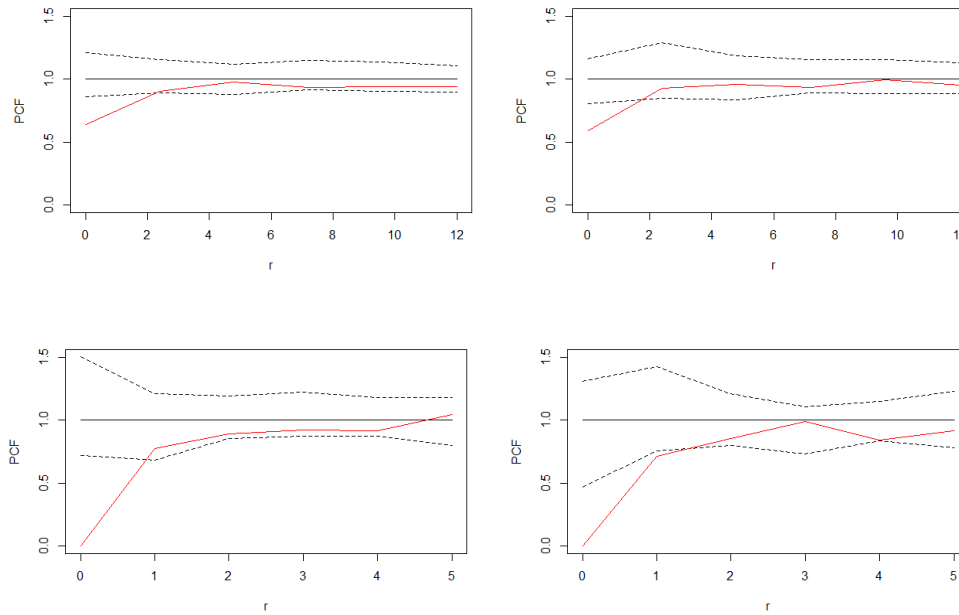


Figure 156: The  $PCF$  for the brain slice  $z = 70$  (black) with confidence envelope from 50 simulations (dashed), and  $\hat{E}_g(r)$  (red) obtained from the 154 components (left) and 142 holes (right) on the slice at  $t = 1$ . Bottom plots show results for smaller distances.

Figures 154 and 156 again show that for small distances, there is evidence of regularity in position of components and holes on the slice, and as noted this is due to the lattice nature of the data.

The results in this section have shown that for the three slices considered, the K-function and radial distribution function show that the distribution of components and holes on these slices could be random except for small values of  $r$ . In these cases the number of observed points fall below expected indicating regularity, which agrees with the results of the quadrant count analysis seen in Section 8.7.1. We note however that this may be due to the lattice nature of the data and the fact that adjacent points cannot both be components, hence for small  $r$ , we may observe fewer components than expected.

### **8.7.3 Testing For Homogeneity On The Full Brain**

In this section we will test for spatial randomness on the full brain. Figure 157 shows the results of quadrant counts on the full brain at the first time point.

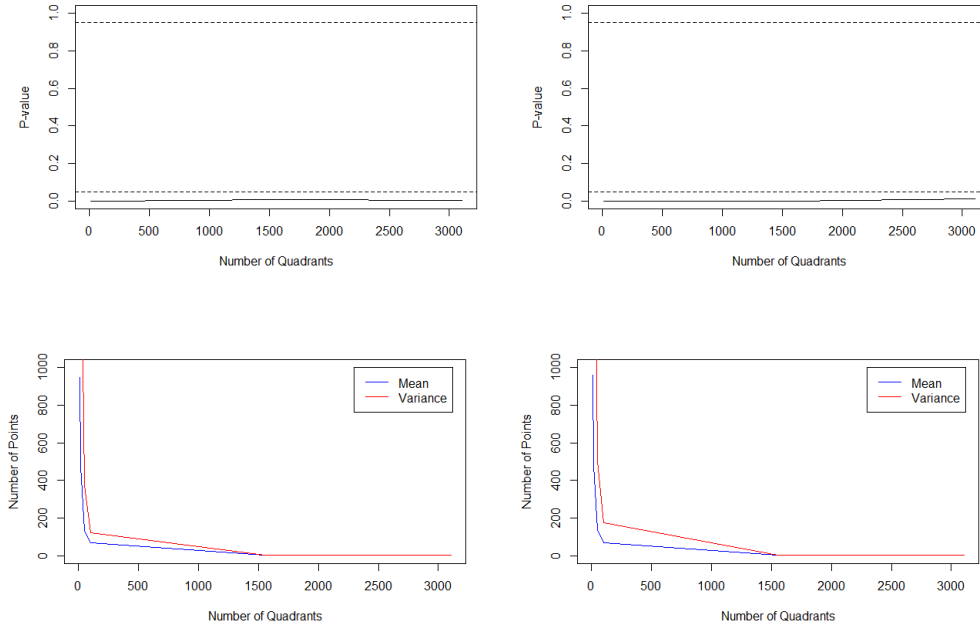


Figure 157: The  $\chi^2$   $p$ -values (top row) and mean and variance (bottom row) for the counts of components (left) and holes (right) of model residuals on the full brain at  $t = 1$ .

From Figure 157 it can be seen that there is evidence of clustering of both the position of components and holes on the full brain. Similar results were found at other time points.

There are a total of 152195 lattice points on the full brain, 6634 components and 6702 holes at  $t = 1$ , hence here  $\lambda = 0.04$  to two decimal places in both cases. Figure 158 shows the comparison of  $E(r)$  and  $\hat{E}(r)$  in this case.

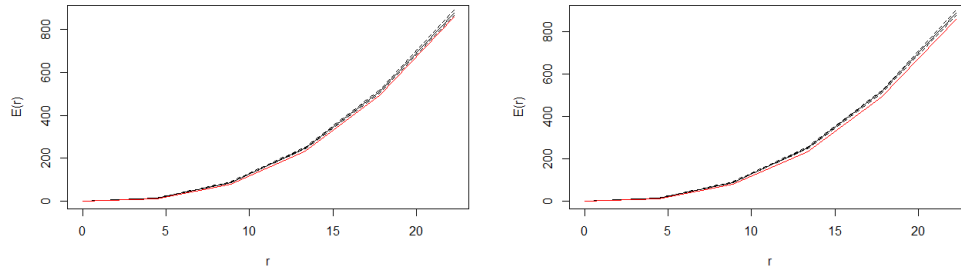


Figure 158:  $E(r)$  for the full brain at  $t = 1$  (black) with confidence envelope from 50 simulations (dashed), and  $\hat{E}(r)$  obtained from the 6634 components (left) and 6702 holes (right) .

Figure 158 shows that  $\hat{E}(r)$ , obtained from the position of components and holes on the full brain, is very slightly lower than the confidence envelope of  $E(r)$  for all  $r$ , therefore there may be some slight evidence that the components and holes on the full brain at  $t = 1$  are more regular than what we would see if the data were randomly distributed across the brain.

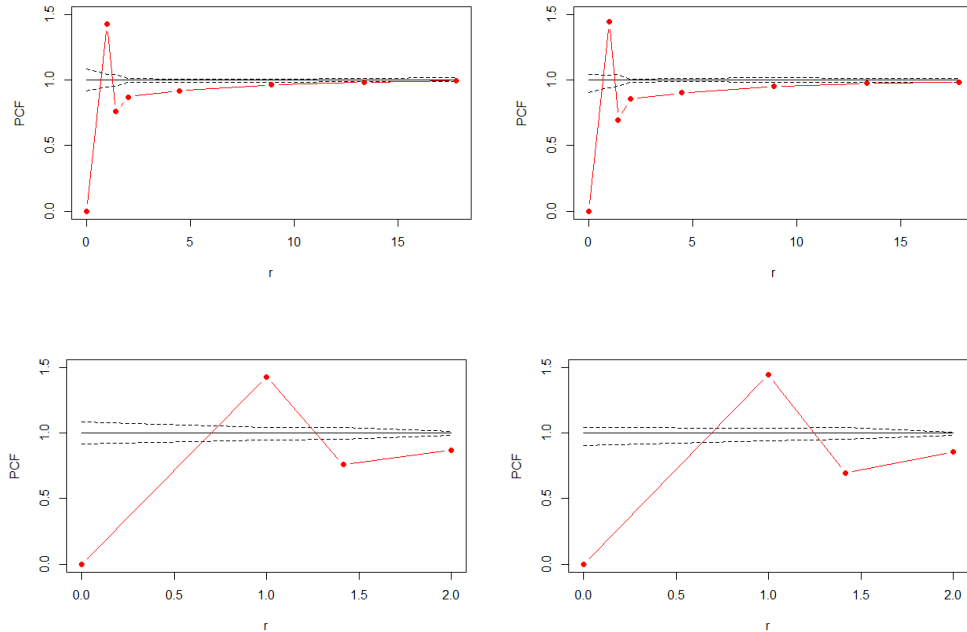


Figure 159: The *PCF* for the full brain at  $t = 1$  (black) with confidence envelope from 50 simulations (dashed), and  $\hat{E}_g(r)$  (red) obtained from the 6634 components (left) and 6702 holes (right). Bottom plots show results for smaller distances.

Figure 159 shows that at smaller distances the distribution of points within intervals are not consistent with a random point pattern. From the bottom plots it can be seen that for  $1 > r > 0$  the pairwise correlation function is zero for both components and holes. As before this is because there cannot be a component within radius one of another component, due to observations being made on the lattice. It can also be seen that for  $\sqrt{2} \geq r > 1$  that there are more observed components/holes than expected under complete randomness. Again, this is due to the lattice structure. To highlight this consider a simple example. Suppose we have a three by three lattice, with nine possible locations of a component. Suppose that two components are observed on the lattice, hence  $\lambda = 2/9 = 0.22$ . That is, the probability of any one of the nine lattice points being a component under randomness is 0.22. Now suppose we observe a component on the central point of the lattice. In this situation, the four immediate neighbours can no longer be

components. However as there is a second component, this increases the probability that the remaining four points (satisfying  $\sqrt{2} \geq r > 1$ ) are components to  $\lambda = 1/4 = 0.25$ .

As we consider intervals at larger distances the point patterns become more consistent with the random point pattern as can be seen in the top plots.

In this section we have seen conflicting results between the quadrant count analysis and the K-function results. The quadrant counts show that there is clustering present, however the K-function and radial distribution function show that the points could be randomly distributed except at short distances as expected for the lattice data. As noted above, the results from the quadrant analysis could be influenced by the increased probability of observing a component (or hole) at distance  $\sqrt{2} \geq r > 1$  from another component when considering lattice data. Therefore, the K-function and radial distribution function (pairwise correlation function) are more useful in this instance as it is clear what is happening at different scales. These results show that as  $r$  increases beyond two, the point patterns become consistent with a random point pattern.

## 8.8 Application II - Comparing Point Patterns

### 8.8.1 Minimum Distance Between Point Patterns

In this section we will compare the spatial point patterns of the components and holes on brain slices by computing the minimum distance between point patterns as described in Section 8.4.1. The aim is to compare components and holes over all time points to investigate if there are any time points where the point patterns change significantly.

Figures 160, 161 and 162 show examples of a mapping between two sets of points which minimises the Euclidean distance between points. Here the mappings are between the components and holes on brain slices  $z = 42$ ,  $z = 60$  and  $z = 70$  respectively at  $t = 1$ , however any two sets of points could be used, for example the components on the slice at two different time points. This was achieved using the *R* package 'clue'. Note that as the pairs of point patterns being considered will be of different sizes, the distance metric (97) given in Section 8.4.1 which adjusts for uneven number of points will be used in all cases.



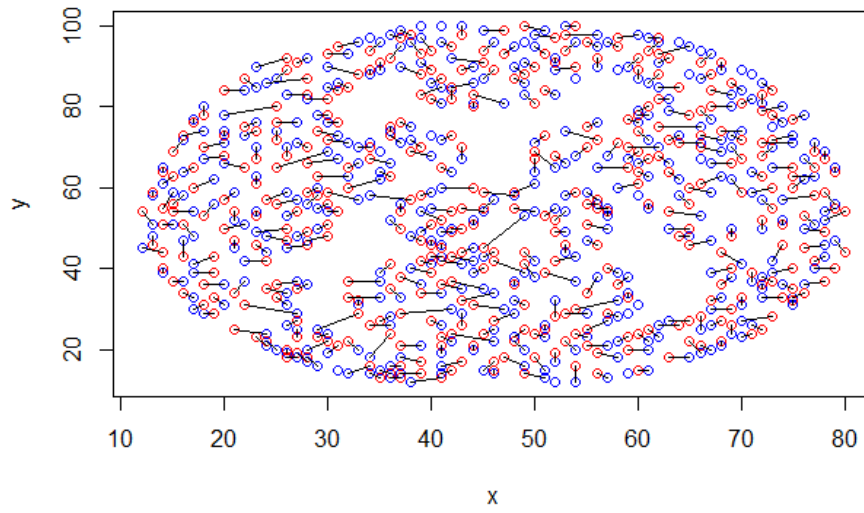


Figure 160: The position of the components and holes (blue and red) on brain slice  $z = 42$  at  $t = 1$ , with the mapping between points with minimal Euclidean distance.

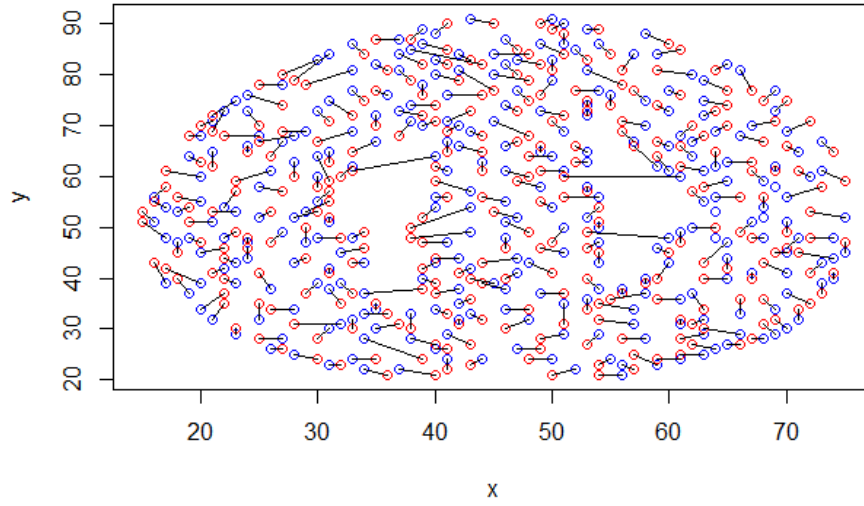


Figure 161: The position of the components and holes (blue and red) on brain slice  $z = 60$  at  $t = 1$ , with the mapping between points with minimal Euclidean distance.

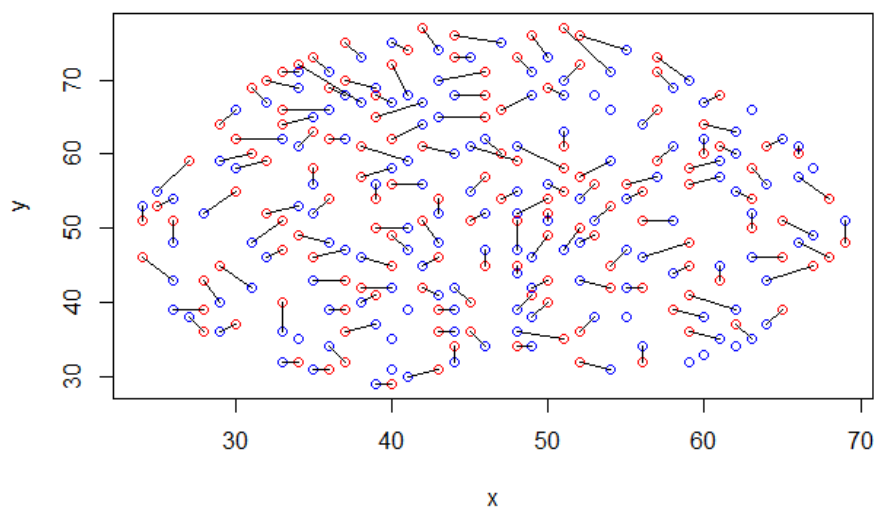


Figure 162: The position of the minima and maxima (blue and red) on brain slice  $z = 70$  at  $t = 1$ , with the mapping between points with minimal Euclidean distance.

Figures 163, 164 and 165 show the minimum distance between components (and holes) of adjacent time points on the brain slices  $z = 42$ ,  $z = 60$  and  $z = 70$  respectively.

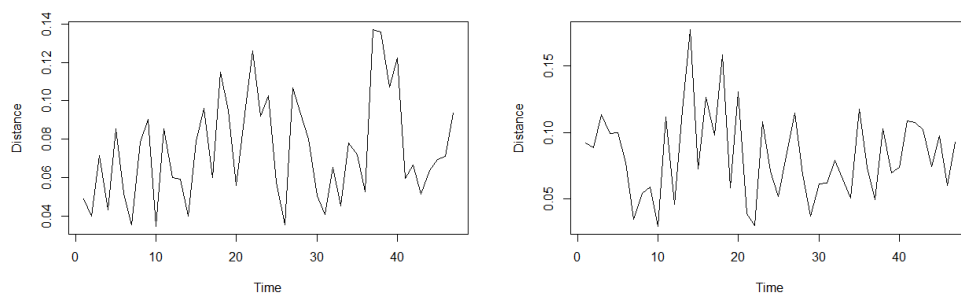


Figure 163: The distance between the components (left) and holes (right) on brain slice  $z = 42$  at adjacent time points, using the mapping between points with minimal Euclidean distance.

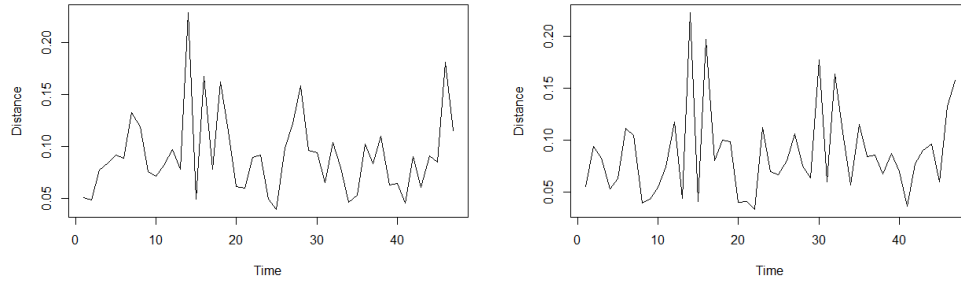


Figure 164: The distance between the components (left) and holes (right) on brain slice  $z = 60$  at adjacent time points, using the mapping between points with minimal Euclidean distance.

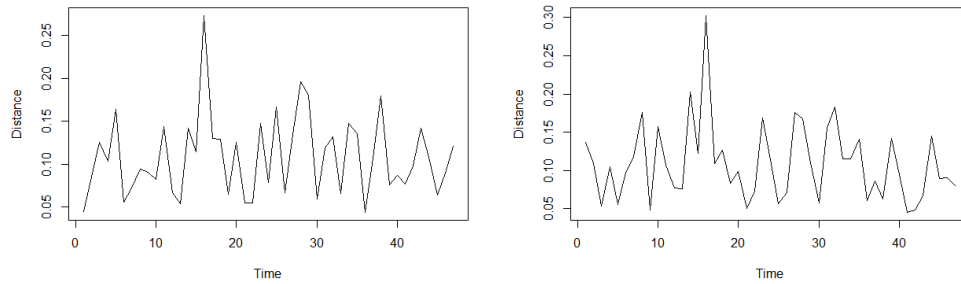


Figure 165: The distance between the components (left) and holes (right) on brain slice  $z = 70$  at adjacent time points, using the mapping between points with minimal Euclidean distance.

Figure 166 shows the corresponding results for the full brain.

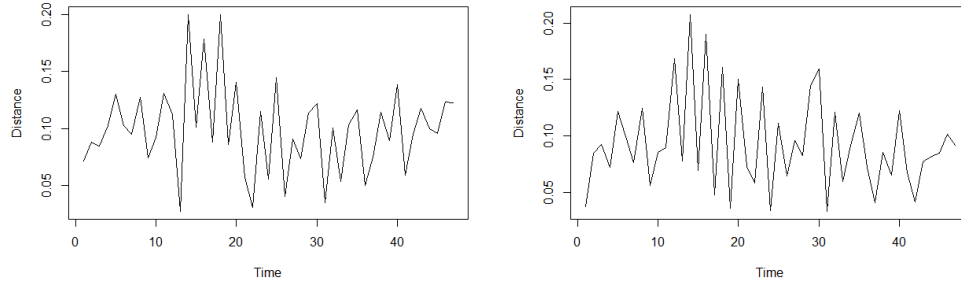


Figure 166: The distance between the components (left) and holes (right) on the full brain at adjacent time points, using the mapping between points with minimal Euclidean distance.

Figures 163-166 show that the distance between adjacent time points for components or holes reveal evidence of an increase at the time of the onset of activity in the experiment. This was not seen in the analysis of the standardised model residuals using traditional topological methods such as persistence diagrams or landscape functions (Section 6.4), nor was this found by analysing the number of topological features in Chapter 7.

### 8.8.2 Comparing Gaussian And Brain Slice Point Patterns

In this section, the aim is to determine if the observed position of components and holes on brain slices are consistent with that of a Gaussian field. In Chapter 7 it was found that the number of components and holes on brain slices is very similar to the number obtained from simulating isotropic Gaussian data on the brain slices. Here we will investigate if the position of the components and holes on the slices are similar to simulated Gaussian data on the slices.

The method used is similar to that of Section 8.6.1. First 50 pairs of Gaussian fields will be simulated on the brain slice, using the estimated correlation on the slice, and for each pair of fields, the distance between the point patterns of the components will be calculated using Equation 96 which corrects for difference in number of points. From this set of 50 distances, the cumulative distribution function will be calculated. This will be repeated 100 times to obtain a mean cdf for pairs of Gaussian fields, as well as an interval for Gaussian cdfs. These will then be plotted against the cumula-

tive distribution function obtained from the distances between 50 simulated Gaussian fields on the slice, and the actual point pattern of the components on the slice. If the point pattern of components on the brain slice is similar to that of a Gaussian field, the cumulative distribution functions should be similar. Figure 167 show two pairs of point patterns on the brain slice with  $z = 42$ . One is the components of two simulated Gaussian fields, and the other the components of one Gaussian field and the actual components on the brain slice. From this we can see that the patterns look similar.

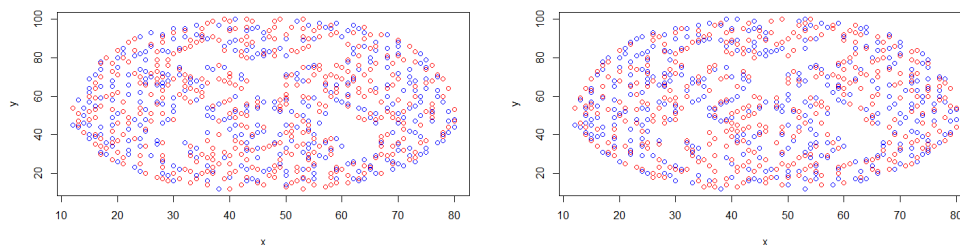


Figure 167: The components of two simulated Gaussian fields on slice  $z = 42$  at  $t = 1$  (left), and the components of a Gaussian simulation (blue) and the components of the brain data (red) (right).

Figures 168, 169 and 170 show results comparing the cumulative distribution functions obtained from pairs of Gaussian fields and pairs of Gaussian field components and actual components on the slices.

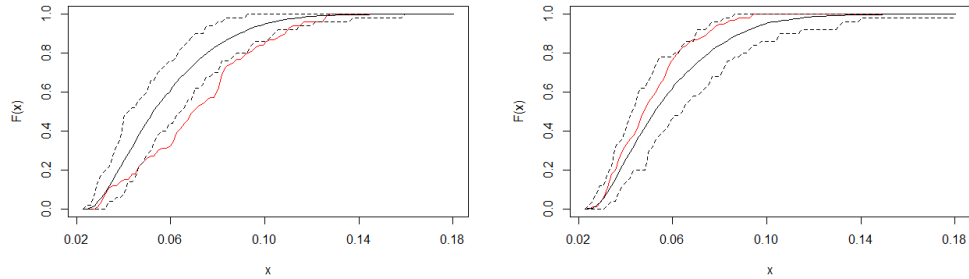


Figure 168: The mean cumulative distribution function (black) and 95% interval (black dashed) of the distances between components (left) and holes (right) of 50 pairs of simulated Gaussian fields on the brain slice  $z = 42$  at  $t = 1$ . Also shown is the cdf of 50 simulated Gaussian and actual point pattern pairs (red).

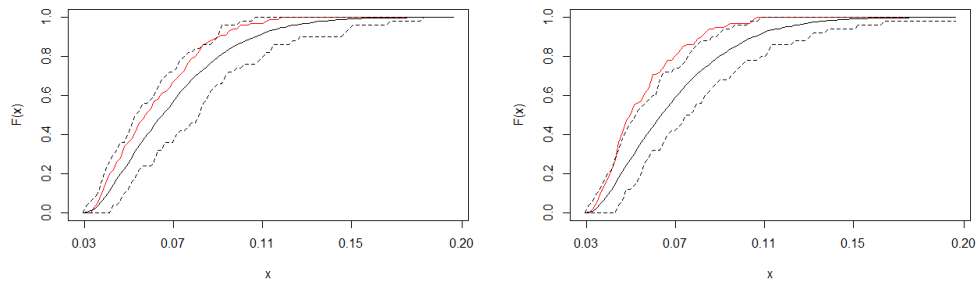


Figure 169: The mean cumulative distribution function (black) and 95% interval (black dashed) of the distances between components (left) and holes (right) of 50 pairs of simulated Gaussian fields on the brain slice  $z = 60$  at  $t = 1$ . Also shown is the cdf of 50 simulated Gaussian and actual point pattern pairs (red).

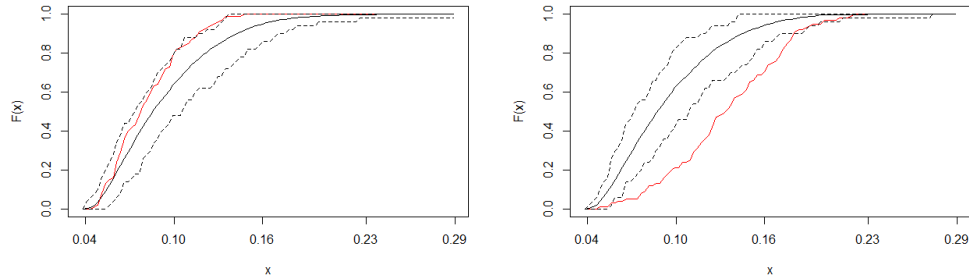


Figure 170: The mean cumulative distribution function (black) and 95% interval (black dashed) of the distances between components (left) and holes (right) of 50 pairs of simulated Gaussian fields on the brain slice  $z = 70$  at  $t = 1$ . Also shown is the cdf of 50 simulated Gaussian and actual point pattern pairs (red).

Although the plots give a good indication of whether the cumulative distribution functions are similar or different, we will use the method as described in Section 8.6.2 to compute the power. As in that section, for the above cases we can compute 1000 distances based on pairs of Gaussian brain slices, and investigate where the distances of 100 simulated Gaussian slice and actual brain slice pairs lie on that distribution. If the power is large, implying that the distances between the Gaussian-brain slice pairings is large with respect to the Gauss-Gauss distances, then we can infer that the brain slice point patterns are non-Gaussian. Table 62 shows the results for the three slices considered.

Brain Slice	Components Power	Holes Power
42	0.05	0.00
60	0.02	0.00
70	0.01	0.23

Table 62: The power obtained from comparing the components/holes of 100 pairs of fields, one a simulated Gaussian field, the other the specified brain slice at  $t = 1$ , to 1000 distances obtained from pairs of Gaussian fields on the brain slice.

Table 62 shows that the power is very small in all cases, hence we cannot



distinguish between the Gauss-Gauss distances and the Gauss-brain slice distances, implying that the point pattern on the brain slices are consistent with a Gaussian point pattern.

### 8.8.3 Cross K-function

Next, the cross K-function will be used to investigate if there are any attraction or repulsion between components and holes at different scales.

Figure 171 shows the cross K-function, which is comparing the position of components and holes on each the brain slices  $z = 42$ ,  $z = 60$  and  $z = 70$ .

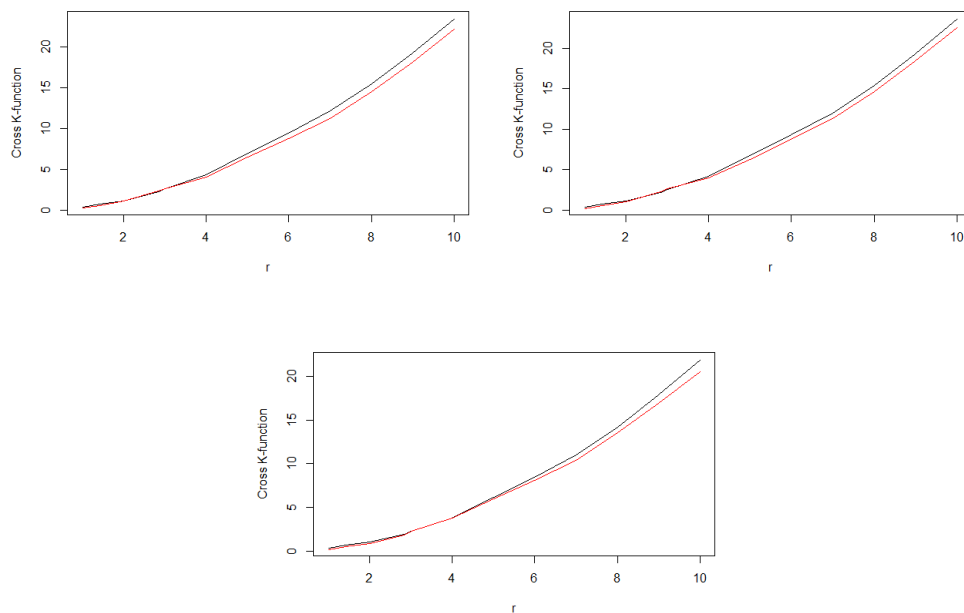


Figure 171: The cross K-function between the components and holes at  $t = 1$  on the brain slices  $z = 42$  (top left),  $z = 60$  (top right) and  $z = 70$  (bottom centre). Red is the observed cross k-function, and black is the expected under randomness.

Figure 171 shows that the observed cross K-function on the slices is very close to the expected. That is the number of components within distance  $r$  of the holes are on average similar to what we would get under the assumption that the components were randomly distributed across the slice. The

figures in Section 8.8.1 which show the position of components and holes on these slices do not show any sign of clustering of components or holes, or attraction or repulsion, hence this result is as expected. Figure 171 does show some difference between the observed and expected cross K-function. As described in Section 8.4, this difference can be measured for significance using a permutation test. Figure 172 shows the results. The plots show the interval obtained by permuting the components and holes on the slice 99 times, recalculating the cross K-function each time, and comparing where the observed lies in relation to the interval.

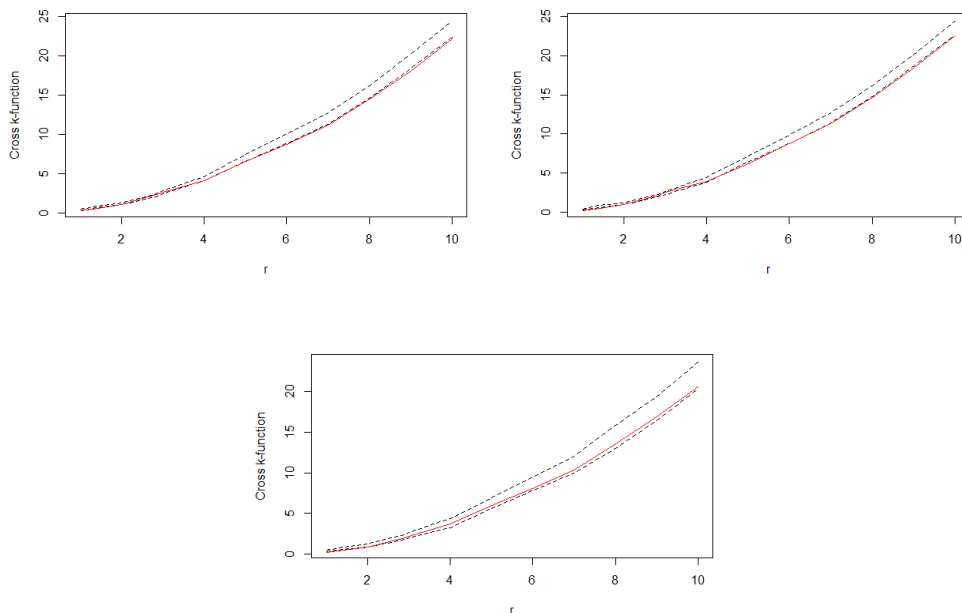


Figure 172: The cross K-function between the components and holes on the brain slices  $z = 42$  (top left),  $z = 60$  (top right) and  $z = 70$  (bottom centre). Red is the observed cross K-function, and black dashed is the confidence envelope obtained from 99 permutations.

Figure 172 shows that in all three slices that the observed cross K-function is consistent with independence.

Similar results were also found when comparing the components and holes on a single slice over time, ie the cross K-function and associated confidence envelopes did not change significantly over time.

Figure 173 shows the cross K-function comparing the position of components and holes on the full brain at  $t = 1$ , along with the confidence envelope for independence arising from 99 random permutations of the components and holes.

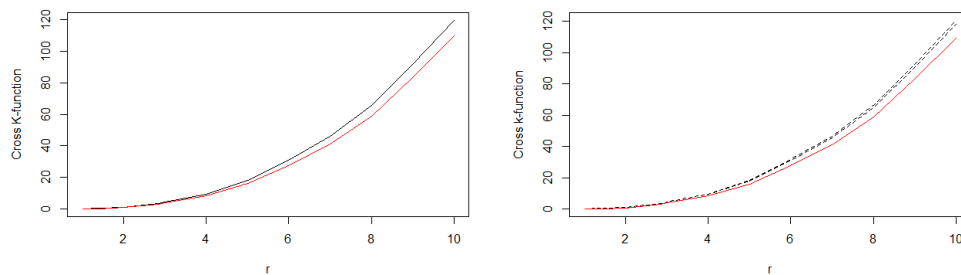


Figure 173: The cross K-function between the components and holes on the full brain at  $t = 1$  (left) and the confidence envelope (right). Red is the observed cross K-function, and black is the expected under randomness.

Figure 173 shows that the observed cross K-function on the full brain is very close to the expected. That is the numbers of components within distance  $r$  of the holes are on average similar to what we would get under the assumption that the components were randomly distributed across the brain. The figure on the right shows the confidence envelope obtained from 99 random permutations of the components and holes on the full brain. Note that at all values of  $r$ , the observed cross K-function falls below the interval, which indicates repulsion between the components and holes which may be statistically significant but small.

Similar results were also found when comparing the components and holes on the full brain over time, ie the cross K-function and associated confidence envelopes did not change significantly over time.

## 8.9 Conclusion

In this chapter we have discussed many methods that can be used to test for complete spatial randomness of a point pattern or to compare two spatial point patterns. Simulations were used to demonstrate how these methods work on lattice data. Quadrant count analysis can be used to obtain a single

statistic to test for spatial randomness, and can distinguish often between random, clustered and regular point patterns. However if one is interested in testing for randomness at different scales, Ripley’s K-function can also be used. Simulations were also used to determine if the point patterns of the components and holes of various fields, all marginally normalised and with the same target correlation, could be used to distinguish between the underlying distributions. It was found, using the K-function of single point patterns, that distinctions could be made between the Gaussian and  $T$  distribution, and also the  $\chi_1^2$  distribution due to its asymmetry in components and holes.

It was found that using the cumulative distribution functions of the distance between point patterns of either the same or different distributions produced plots which gave an idea of the similarity or difference between pairs of such point patterns, however using the probability distribution of such distances gave a more efficient way of testing for differences amongst components and holes of various pairs of distributions. It was found that little distinction could be made between pairs of Gaussian fields and pairs of the other fields, however it was found that two Gaussian point patterns were almost always statistically distinct from a Gaussian-other distribution pair. This is interesting as this provides a test for determining if brain data are Gaussian using the spatial point patterns of the topological features.

In the application section, quadrant count analysis showed that the point patterns of components and holes on some slices were random, and some were regularly distributed across the slice. However the K-function analysis did not show this except for at small scales, and this may be attributed to the lattice nature of the data. Also note that when applying these methods to the full brain, quadrant count analysis revealed clustering. The radial distribution function showed that this is only true at for  $1 < r \leq \sqrt{2}$ , which is again a consequence of the lattice data as explained previously. Therefore it is recommended to use the  $K$ -function and radial distribution function (pairwise correlation function) to investigate spatial randomness at different scales.

Comparing the point patterns of the brain data with Gaussian data showed that the observed components and holes were consistent with that of Gaussian data. Finally it was found by comparing the point patterns of the components or holes on brain slices at adjacent time points that evidence of a change could be seen at the onset of activity. This was not seen in the analysis of the standardised model residuals using traditional topological methods such as persistence diagrams or landscape functions (Section 6.4),

nor was this found by analysing the number of topological features in Chapter 7. Therefore this method could be used to detect the onset of activity in noisy fMRI data, where other explored methods fail.

## 9 Conclusion

### 9.1 Summary Of Findings And Contribution

The aim of this thesis was to fit a statistical model to the fMRI data and develop a method to detect changes in neural activity in response to activity. As discussed in Section 3.1, previously proposed models had the disadvantage of being unable to model the spatial dependence between the large number of voxels due to computational limitations. This often resulted in models which make simplifying assumptions, such as independence between observations at different voxels or regions. It has been shown that such assumptions significantly increase the rate of falsely identifying activity in voxels. In Chapter 3 we proposed a model which captures the temporal and spatial dependence between sets of voxels using a sequential approach to parameter estimation, which significantly reduces computation time. Model checks in Section 3.9 show that the model residuals were approximately noise indicating an adequate fit.

Analysis of the model residuals in Chapter 5 revealed interesting findings. In particular, clustering of the residual regional means into two groups very closely matched the partitioning of the brain into two groups of two lobes (one consisting of the frontal or parietal lobes, and the other of the temporal or occipital lobes) (see Section 5.3). It is well known (2.2) that the frontal and parietal lobes are associated with movement, and the temporal and occipital with audio and visual processing. Therefore, the clustering of the model residuals can distinguish between regions of the brain which are involved with movement from those which are not.

In Chapter 4, we used the estimated model parameters to test which regions of the brain were associated with activity. However the main contribution of this work is the new methods developed in Chapter 8 which analyse the spatial distribution of topological features to detect neural activity. In Chapter 6 we introduced the main concepts of TDA, and found a promising result using the changes in persistence homology over time. It appeared that changes could be seen at points of activity within the experiment. However, this could also be seen in the variance of the residuals over time. When the residuals were standardised over time, no changes could be seen in the mean or variance of the data or in the persistence diagrams.

In Chapter 7 we extended the work of [51] to investigate the number of topological features on various fields, all normal standardised with the

same target correlation. The main difference being the extension to three dimensional, and irregular lattices. It was found through simulations that distinctions between the fields may not be possible for smaller lattices using components and holes alone. Also of interest is that the expected number of components and holes on the Gaussian field is extremely sensitive to local correlation, with only small changes in correlation causing large differences in number of components and holes on the field. In addition, through simulations we found that a very popular *R* package used for simulating random fields (*RandomFields*) fails in three dimensions to produce fields with the theoretical number of topological features. Although the package appears to produce fields which are Gaussian, the number of components and voids expected are far from those observed on simulated fields (see Section 7.2.5). This package has over 400,000 total downloads, and over 90,000 within the last year, as found using the *R* package *cranlogs*. Therefore this is an interesting finding, as any analysis of minima or maxima on three dimensional fields produced in this package may be incorrect.

The main finding is in Chapter 8. We worked with residuals standardised over time, where as stated no changes could be seen in the mean or variance of the data, or in the persistence diagrams of the data. However, applying a new idea of analysis of the spatial distribution of the topological features and how these change over time, we were able to detect a significant change due to the onset of activity. This is a new method that could be used in future work to detect changes in a signal where standard approaches cannot.

## 9.2 Suggestions For Future Work

In Chapter 3 we proposed a space-time model for the data. One limitation of our model is the assumption of stationarity. Although checks were made to ensure that this was a sensible assumption for this data set, a suggestion for future work could be to incorporate a non-stationary spatial dependence between observations. The model we proposed also assumed the spatial dependence is constant though time, an assumption we found to be adequate for this data set. In future work allowing for spatial dependence between time points could be an interesting addition to the model, however computational limitations will need to be considered. Finally, although we had access to only a single data set, incorporating additional data sets from future fMRI studies into the modelling could strengthen the inference.

The main contribution of this work is the new method of detecting acti-

vation using the model residuals and spatial topological data analysis. Note that in the application of the topological methods, the uncertainty in the model residuals has been ignored. Uncertainty in the model residuals comes from two sources, the variance of the data, and the uncertainty arising from the fitting of the model. It is the second source of uncertainty that is ignored in our application. An interesting extension to this work would be to measure this uncertainty. Figures 174 and 175 show examples of a starting point of this. Here data is simulated from the spatial model in Section 3.7, with mean and autoregressive parameters used in Table 10 such that the voxels are active, and the Matérn spatial covariance function with parameters  $(\nu, \eta) = (0.5, 2)$ . Here data is simulated on a 65 by 65 lattice where each time series is of length 48. Figure 174 shows the number of components found using the residuals used to simulate the data, and the number found when using the estimated residuals after the model is fit, thus showing the impact of the uncertainty arising from the model fitting. Similarly Figure 175 shows the distance (using metric (97)) between the point patterns of the components at adjacent time points using the two sets of residuals.



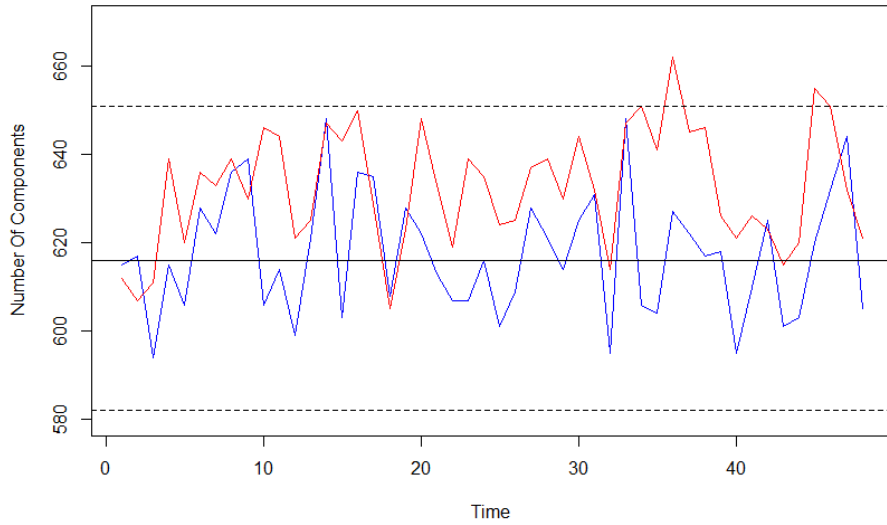


Figure 174: The number of components at each time point on a 65 by 65 lattice using simulated residuals (blue) and estimated residuals (red) after model fitting. Black solid horizontal line shows the expected number for the Gaussian field, with the mean plus/minus two standard deviations (black dashed).

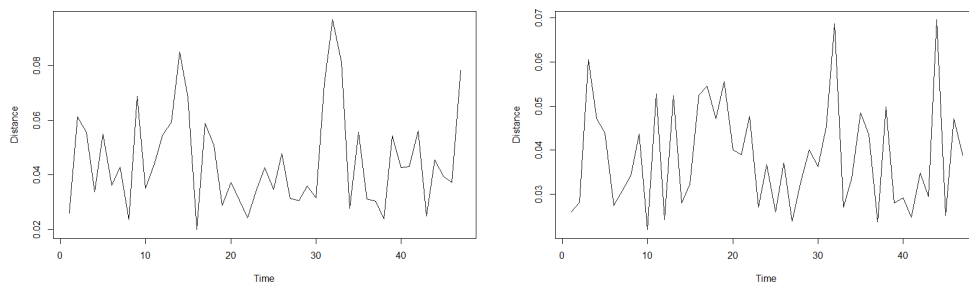


Figure 175: The distance between the components of adjacent time points on a 65 by 65 lattice using simulated residuals (left) and estimated residuals after model fitting (right).

Figure 174 shows that there are some differences in the number of com-

ponents of the two sets of residuals, however the difference is not substantial, as the estimated residuals still give results which lie within two standard deviations of the expected as calculated in Table 38. However Figure 175 shows differences between the two sets of results. In particular, using the simulated residuals we can see the two peaks of activation at  $t = 12$  and  $t = 36$ , however after the model fitting, the results are much more noisy, with much less evidence of activation. An option for future work would be a more in depth analysis of how the uncertainty from the model affects this and other topological methods discussed in the thesis.

Finally, an interesting possibility of extension to this work could be to investigate how the topological features evolve on a field as the level set changes. This could reveal further differences between different fields, even when they are normalised and with the same target correlation, or if used in combination with the temporal evolution of the field, could reveal more information regarding neural activation.

## References

- [1] Anatomy of the brain. <https://mayfieldclinic.com/pe-anatbrain.htm>. Accessed: 2018-11-06.
- [2] Brain structures and their functions. <https://serendipstudio.org/bb/kinser/Structure1.html>. Accessed: 2018-11-06.
- [3] Comparative analyses of point patterns. [https://www.seas.upenn.edu/~ese502/NOTEBOOK/Part\\_I/5\\_Comparative\\_Analyses.pdf](https://www.seas.upenn.edu/~ese502/NOTEBOOK/Part_I/5_Comparative_Analyses.pdf). Accessed: 2019-05-16.
- [4] Convex hull. <https://www.ti.inf.ethz.ch/ew/courses/CG13/lecture/Chapter%203.pdf>. Accessed: 2020-07-29.
- [5] A few words about reml. <http://users.stat.umn.edu/~gary/classes/5303/handouts/REML.pdf>. Accessed: 2018-08-31.
- [6] How to calculate the pair correlation function. <http://www.physics.emory.edu/faculty/weeks//idl/gofr2.html>. Accessed: 2019-05-02.
- [7] Lobes of the brain. <https://qbi.uq.edu.au/brain/brain-anatomy/lobes-brain>. Accessed: 2018-11-06.
- [8] Norms and condition numbers. [http://www.cse.iitd.ernet.in/~dheerajb/CS210\\_lect07.pdf](http://www.cse.iitd.ernet.in/~dheerajb/CS210_lect07.pdf). Accessed: 2018-08-03.
- [9] Origins of the bold signal. <http://www.sbirc.ed.ac.uk/cyril/fMRI4.html>. Accessed: 2018-06-07.
- [10] The pair correlation function. [http://www.soft-matter.uni-tuebingen.de/teaching/vorlesung\\_condensed\\_matter\\_ws2017/Summary-g-of-r-2017-11-20-short.pdf](http://www.soft-matter.uni-tuebingen.de/teaching/vorlesung_condensed_matter_ws2017/Summary-g-of-r-2017-11-20-short.pdf). Accessed: 2019-05-02.
- [11] Stationarity conditions for an ar(2) process. <http://www.sfu.ca/~baa7/Teaching/econ818/StationarityAR2.pdf>. Accessed: 2018-12-11.

- [12] Stochastic modelling of fractures in rock masses. [http://www.leeds.ac.uk/StochasticRockFractures/Download/ReportsSeminars/Report3\\_Pattern\\_Analysis.pdf](http://www.leeds.ac.uk/StochasticRockFractures/Download/ReportsSeminars/Report3_Pattern_Analysis.pdf). Accessed: 2019-06-03.
- [13] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.
- [14] Alex M Andrew. Another efficient algorithm for convex hulls in two dimensions. *Information Processing Letters*, 9(5):216–219, 1979.
- [15] Barry C Arnold. Flexible univariate and multivariate models based on hidden truncation. *Journal of Statistical Planning and Inference*, 139(11):3741–3749, 2009.
- [16] Adelchi Azzalini. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32(2):159–188, 2005.
- [17] Andrew David Barbour and Timothy Carlisle Brown. Stein’s method and point process approximation. *Stochastic Processes and their Applications*, 43(1):9–31, 1992.
- [18] Jessica Barrett, Peter Diggle, Robin Henderson, and David Taylor-Robinson. Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):131–148, 2015.
- [19] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.
- [20] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188, 2001.
- [21] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.

- [22] Douglas W Bigwood and David W Inouye. Spatial pattern analysis of seed banks: an improved method and optimized sampling. *Ecology*, 69(2):497–507, 1988.
- [23] J Martin Bland and Douglas G Altman. Multiple significance tests: the Bonferroni method. *Bmj*, 310(6973):170, 1995.
- [24] F Dubois Bowman. Spatio-temporal modeling of localized brain activity. *Biostatistics*, 6(4):558–575, 2005.
- [25] F Dubois Bowman. Spatiotemporal models for region of interest analyses of functional neuroimaging data. *Journal of the American Statistical Association*, 102(478):442–453, 2007.
- [26] F DuBois Bowman, Brian Caffo, Susan Spear Bassett, and Clinton Kilts. A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage*, 39(1):146–156, 2008.
- [27] Peter Bubenik. Statistical topological data analysis using persistence landscapes. *The Journal of Machine Learning Research*, 16(1):77–102, 2015.
- [28] Ed Bullmore, Chris Long, John Suckling, Jalal Fadili, Gemma Calvert, Fernando Zelaya, T Adrian Carpenter, and Mick Brammer. Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains. *Human brain mapping*, 12(2):61–78, 2001.
- [29] Petrutza Caragea and Richard L Smith. Approximate likelihoods for spatial processes. *Preprint*, 2006.
- [30] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.
- [31] Stefano Castruccio, Hernando Ombao, and Marc G Genton. A multi-resolution spatio-temporal model for brain activation and connectivity in fMRI data. *arXiv preprint arXiv:1602.02435*, 2016.
- [32] Timothy M Chan. Optimal output-sensitive convex hull algorithms in two and three dimensions. *Discrete & Computational Geometry*, 16(4):361–368, 1996.

- [33] Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*, 2017.
- [34] Mark S Cohen. Parametric analysis of fMRI data using linear systems methods. *Neuroimage*, 6(2):93–103, 1997.
- [35] David R Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society: Series B (Methodological)*, 17(2):129–157, 1955.
- [36] Mark RT Dale, Philip Dixon, Marie-Josée Fortin, Pierre Legendre, Donald E Myers, and Michael S Rosenberg. Conceptual and mathematical relationships among methods for spatial analysis. *Ecography*, 25(5):558–577, 2002.
- [37] Peter Diggle, Peter J Diggle, Patrick Heagerty, Patrick J Heagerty, Kung-Yee Liang, Scott Zeger, et al. *Analysis of longitudinal data*. Oxford University Press, 2002.
- [38] Peter J Diggle, Julian Besag, and J Timothy Gleaves. Statistical analysis of spatial point patterns by means of distance methods. *Biometrics*, pages 659–667, 1976.
- [39] Philip M Dixon. Ripley’s k function. *Wiley StatsRef: Statistics Reference Online*, 2014.
- [40] Robin A Dubin. Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *The Review of Economics and Statistics*, pages 466–474, 1988.
- [41] Luc Duchateau, Paul Janssen, and John Rowlands. *Linear mixed models. An introduction with applications in veterinary research*. ILRI (aka ILCA and ILRAD), 1998.
- [42] Idris A Eckley, Paul Fearnhead, and Rebecca Killick. Analysis of change-point models. *Bayesian Time Series Models*, pages 205–224, 2011.
- [43] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 454–463. IEEE, 2000.

- [44] Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, Aarti Singh, et al. Confidence sets for persistence diagrams. *The Annals of Statistics*, 42(6):2301–2339, 2014.
- [45] Montserrat Fuentes. Spectral methods to approximate the likelihood for irregularly spaced spatial data. Technical report, North Carolina State University. Dept. of Statistics, 2004.
- [46] Ramón Giraldo, Pedro Delicado, and Jorge Mateu. Hierarchical clustering of spatially correlated functional data. *Statistica Neerlandica*, 66(4):403–421, 2012.
- [47] François Goreaud and Raphaël Pélissier. Avoiding misinterpretation of biotic interactions with the intertype k12-function: population independence vs. random labelling hypotheses. *Journal of vegetation science*, 14(5):681–692, 2003.
- [48] Ronald L. Graham. An efficient algorithm for determining the convex hull of a finite planar set. *Info. Pro. Lett.*, 1:132–133, 1972.
- [49] Arjun K Gupta, Graciela González-Farías, and J Armando Dominguez-Molina. A multivariate skew normal distribution. *Journal of multivariate analysis*, 89(1):181–190, 2004.
- [50] RA Haggarty, CA Miller, and EM Scott. Spatially weighted functional clustering of river network data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(3):491–506, 2015.
- [51] Robin Henderson, Irina Makarenko, Paul Bushby, Andrew Fletcher, and Anvar Shukurov. Statistical topology and the random interstellar medium. *arXiv preprint arXiv:1703.07256*, 2017.
- [52] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.
- [53] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.
- [54] Jung Won Hyun, Yimei Li, John H Gilmore, Zhaohua Lu, Martin Styner, and Hongtu Zhu. Sgpp: spatial gaussian predictive process models for neuroimaging data. *NeuroImage*, 89:70–80, 2014.

- [55] Janine Illian, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan. *Statistical analysis and modelling of spatial point patterns*, volume 70. John Wiley & Sons, 2008.
- [56] Hakmook Kang, Hernando Ombao, Crystal Linkletter, Nicole Long, and David Badre. Spatio-spectral mixed-effects model for functional magnetic resonance imaging data. *Journal of the American Statistical Association*, 107(498):568–577, 2012.
- [57] Cari G Kaufman, Mark J Schervish, and Douglas W Nychka. Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555, 2008.
- [58] Yuriy V Kozachenko, Oleksandr O Pogorilyak, Iryna V Rozora, and Antonina M Tegza. *Simulation of stochastic processes with given accuracy and reliability*. Elsevier, 2016.
- [59] Yanchi Liu, Zhongmou Li, Hui Xiong, Xuedong Gao, and Junjie Wu. Understanding of internal clustering validation measures. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 911–916. IEEE, 2010.
- [60] Greta M Ljung and George EP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.
- [61] Joseph J Locascio, Peggy J Jennings, Christopher I Moore, and Suzanne Corkin. Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging. *Human brain mapping*, 5(3):168–193, 1997.
- [62] David G Luenberger, Yinyu Ye, et al. *Linear and nonlinear programming*, volume 2. Springer, 1984.
- [63] Sebastian Mika, Gunnar Ratsch, Jason Weston, Bernhard Scholkopf, and Klaus-Robert Mullers. Fisher discriminant analysis with kernels. In *Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop.*, pages 41–48. Ieee, 1999.
- [64] Rajul Parikh, Annie Mathai, Shefali Parikh, G Chandra Sekhar, and Ravi Thomas. Understanding and using sensitivity, specificity and predictive values. *Indian journal of ophthalmology*, 56(1):45, 2008.



- [65] Brian D Ripley. The second-order analysis of stationary point processes. *Journal of applied probability*, 13(2):255–266, 1976.
- [66] Christian Rohrbeck. Detection of changes in variance using binary segmentation and optimal partitioning, 2013.
- [67] Dominic Schuhmacher and Aihua Xia. A new metric between distributions of point processes. *Advances in applied probability*, 40(3):651–672, 2008.
- [68] R John Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.
- [69] Douglas Steinley, Michael J Brusco, and Lawrence Hubert. The variance of the adjusted Rand index. *Psychological methods*, 21(2):261, 2016.
- [70] Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001.
- [71] Ruey S Tsay. *Analysis of financial time series*, volume 543. John Wiley & Sons, 2005.
- [72] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni MRI single-subject brain. *Neuroimage*, 15(1):273–289, 2002.
- [73] Peter CM van Zijl, Jun Hua, and Hanzhang Lu. The bold post-stimulus undershoot, one of the most debated issues in fMRI. *Neuroimage*, 62(2):1092–1102, 2012.
- [74] Mark W Woolrich, Christian F Beckmann, Thomas E Nichols, and Stephen M Smith. Statistical analysis of fMRI data. In *fMRI techniques and protocols*, pages 179–236. Springer, 2009.
- [75] Keith J Worsley, Alan C Evans, S Marrett, and P Neelin. A three-dimensional statistical analysis for cbf activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, 12(6):900–918, 1992.

- [76] Keith J Worsley and Karl J Friston. Analysis of fMRI time-series revisited again. *Neuroimage*, 2(3):173–181, 1995.
- [77] Keith J Worsley, CH Liao, J Aston, V Petre, GH Duncan, F Morales, and AC Evans. A general statistical analysis for fMRI data. *Neuroimage*, 15(1):1–15, 2002.
- [78] KJ Worsley. Detecting activation in fMRI data. *Statistical methods in medical research*, 12(5):401–418, 2003.
- [79] Rui Xu and Don Wunsch. *Hierarchical Clustering*, page 400. Wiley-IEEE Press, 2009.
- [80] Ka Yee Yeung and Walter L Ruzzo. Details of the adjusted Rand index and clustering algorithms, supplement to the paper an empirical study on principal component analysis for clustering gene expression data. *Bioinformatics*, 17(9):763–774, 2001.
- [81] Tingting Zhang, Fan Li, Lane Beckes, and James A Coan. A semi-parametric model of the hemodynamic response for multi-subject fMRI data. *NeuroImage*, 75:136–145, 2013.
- [82] Tingting Zhang, Fan Li, Marlen Z Gonzalez, Erin L Maresh, and James A Coan. A semi-parametric nonlinear model for event-related fMRI. *NeuroImage*, 97:178–187, 2014.
- [83] Tingting Zhanga, Fan Lib, Lane Beckesa, and James A Coana. Non-parametric inference of hemodynamic response for multi-subject fMRI data under multi-stimulus design. *NeuroImage*.
- [84] Hongtu Zhu, Jianqing Fan, and Linglong Kong. Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association*, 109(507):1084–1098, 2014.
- [85] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.



## 10 Appendix

### 10.1 The Regions Of Interest

Region Number	Region Name
1,2	Precentral Gyrus (Left,Right)
3,4	Superior Frontal Gyrus, Dorsolateral (Left,Right)
5,6	Superior Frontal Gyrus, Orbital Part (Left,Right)
7,8	Middle Frontal Gyrus (Left,Right)
9,10	Middle Frontal Gyrus, Orbital Part (Left,Right)
11,12	Inferior Frontal Gyrus, Opercular Part (Left,Right)
13,14	Inferior Frontal Gyrus, Triangular Part (Left,Right)
15,16	Inferior Frontal Gyrus, Orbital Part (Left,Right)
17,18	Rolandic Operculum (Left,Right)
19,20	Supplementary Motor Area (Left,Right)
21,22	Olfactory Cortex (Left,Right)
23,24	Superior Frontal Gyrus, Medial (Left,Right)
25,26	Superior Frontal Gyrus, Medial Orbital (Left,Right)
27,28	Gyrus Rectus (Left,Right)
29,30	Insula (Left,Right)
31,32	Anterior Cingulate and Paracingulate Gyri (Left,Right)
33,34	Median Cingulate and Paracingulate Gyri (Left,Right)
35,36	Posterior Cingulate Gyri (Left,Right)
37,38	Hippocampus (Left,Right)
39,40	Parahippocampal Gyrus (Left,Right)
41,42	Amygdala (Left,Right)
43,44	Calcarine Fissure and Surrounding Cortex (Left,Right)
45,46	Cuneus (Left,Right)
47,48	Lingual Gyrus (Left,Right)
49,50	Superior Occipital Gyrus (Left,Right)
51,52	Middle Occipital Gyrus (Left,Right)
53,54	Inferior Occipital Gyrus (Left,Right)
55,56	Fusiform Gyrus (Left,Right)
57,58	Postcentral Gyrus (Left,Right)
59,60	Superior Parietal Gyrus (Left,Right)
61,62	Inferior Parietal Gyrus (Left,Right)
63,64	Supramarginal Gyrus (Left,Right)
65,66	Angular Gyrus (Left,Right)
67,68	Precuneus (Left,Right)
69,70	Paracentral Lobule (Left,Right)
71,72	Caudate Nucleus (Left,Right)
73,74	Lenticular Nucleus, Putamen (Left,Right)
75,76	Lenticular Nucleus, Pallidum (Left,Right)
77,78	Thalamus (Left,Right)
79,80	Heschl's Gyrus (Left,Right)
81,82	Superior Temporal Gyrus (Left,Right)
83,84	Temporal Pole: Superior Temporal Gyrus (Left,Right)
85,86	Middle Temporal Gyrus (Left,Right)
87,88	Temporal Pole: Middle Temporal Gyrus (Left,Right)
89,90	Inferior Temporal Gyrus (Left,Right)

Table 63: The 90 Regions of Interest

## 10.2 Model Residual Topology Summary Statistics

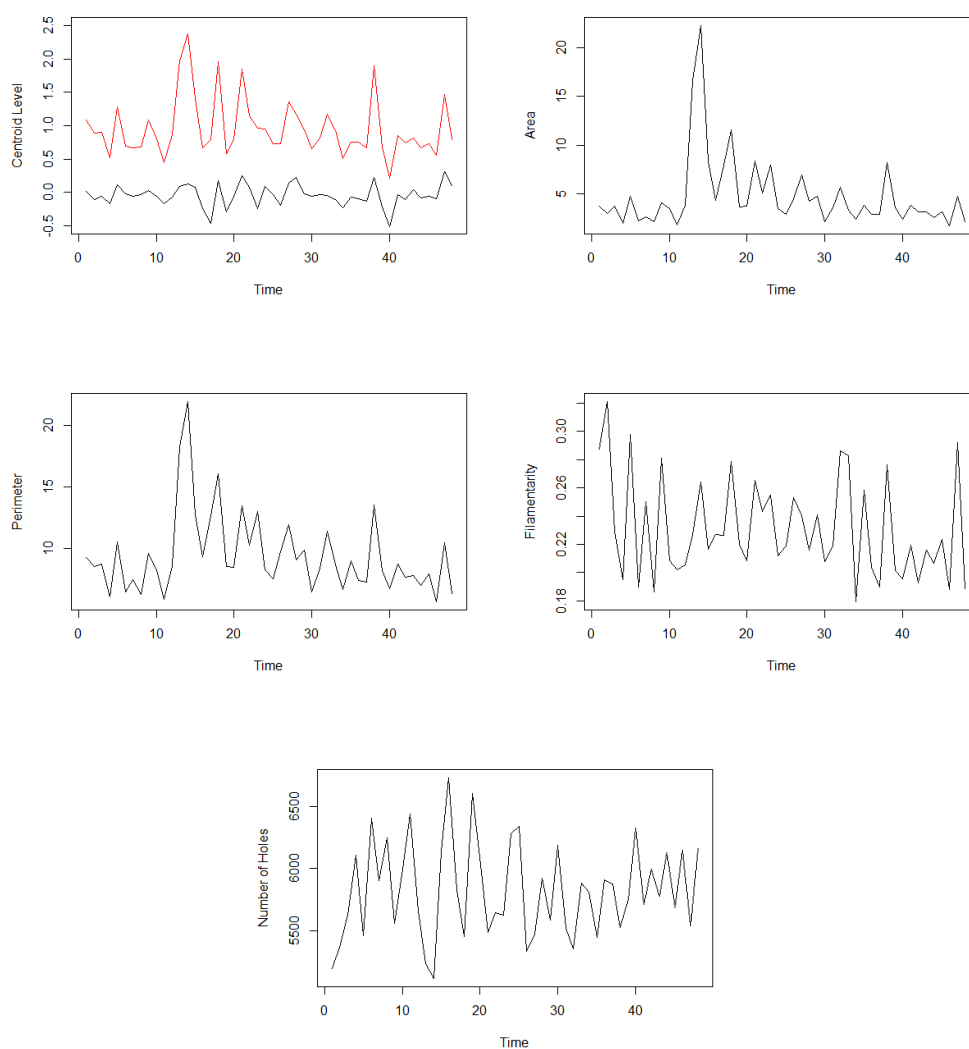


Figure 176: Summary statistics derived using 95% convex peels of persistence diagrams of the holes of the full brain model residuals. Top left shows the birth (black) and death (red) centroid levels over time, and the remaining plots the other summary statistics.

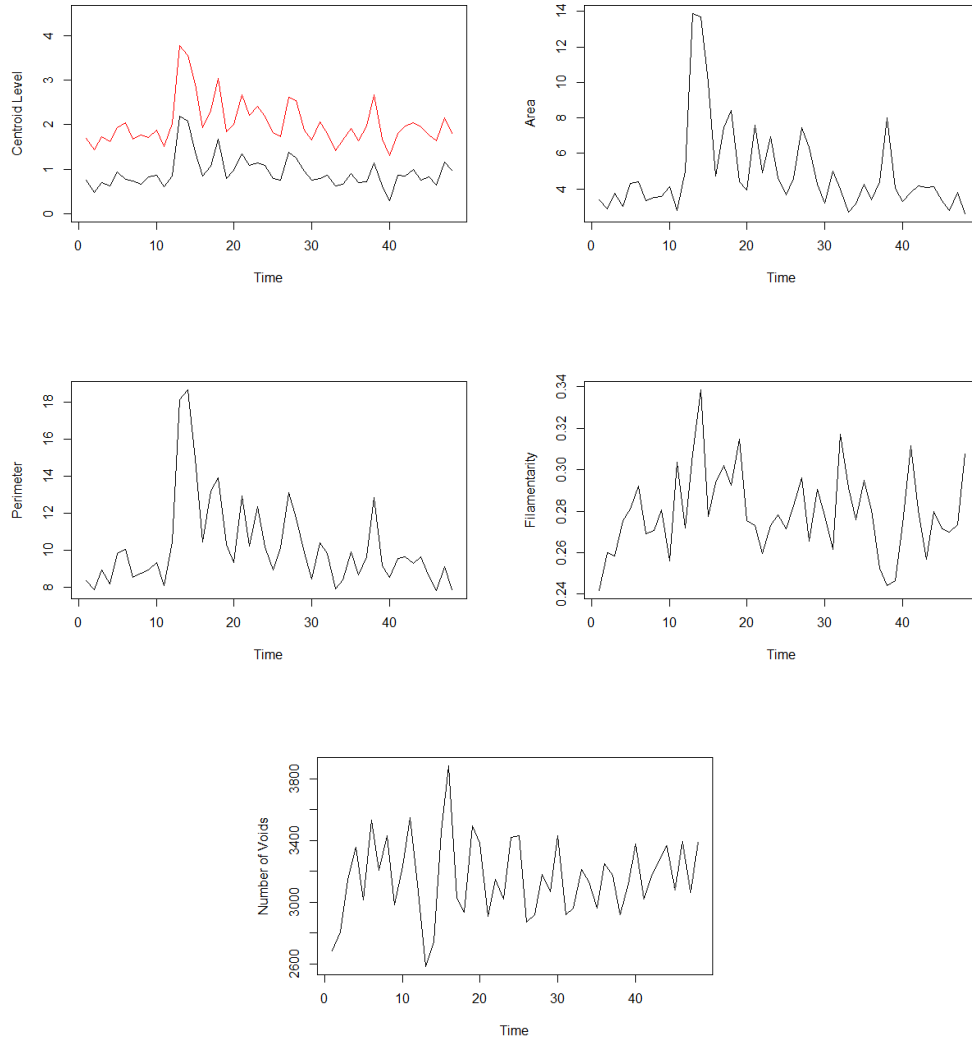


Figure 177: Summary statistics derived using 95% convex peels of persistence diagrams of the voids of the full brain model residuals. Top left shows the birth (black) and death (red) centroid levels over time, and the remaining plots the other summary statistics.

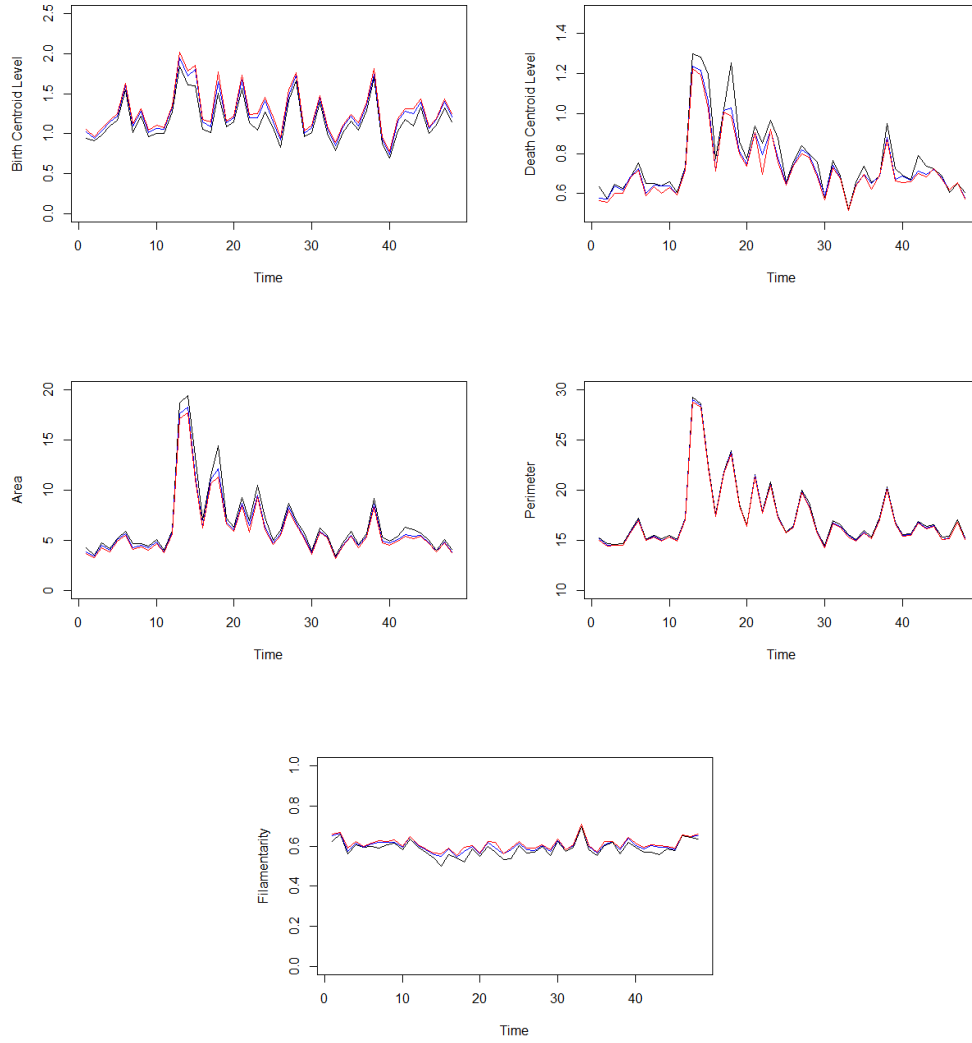


Figure 178: Summary statistics derived using first three landscape functions (black, blue, red) of persistence diagrams of the holes of the full brain model residuals.

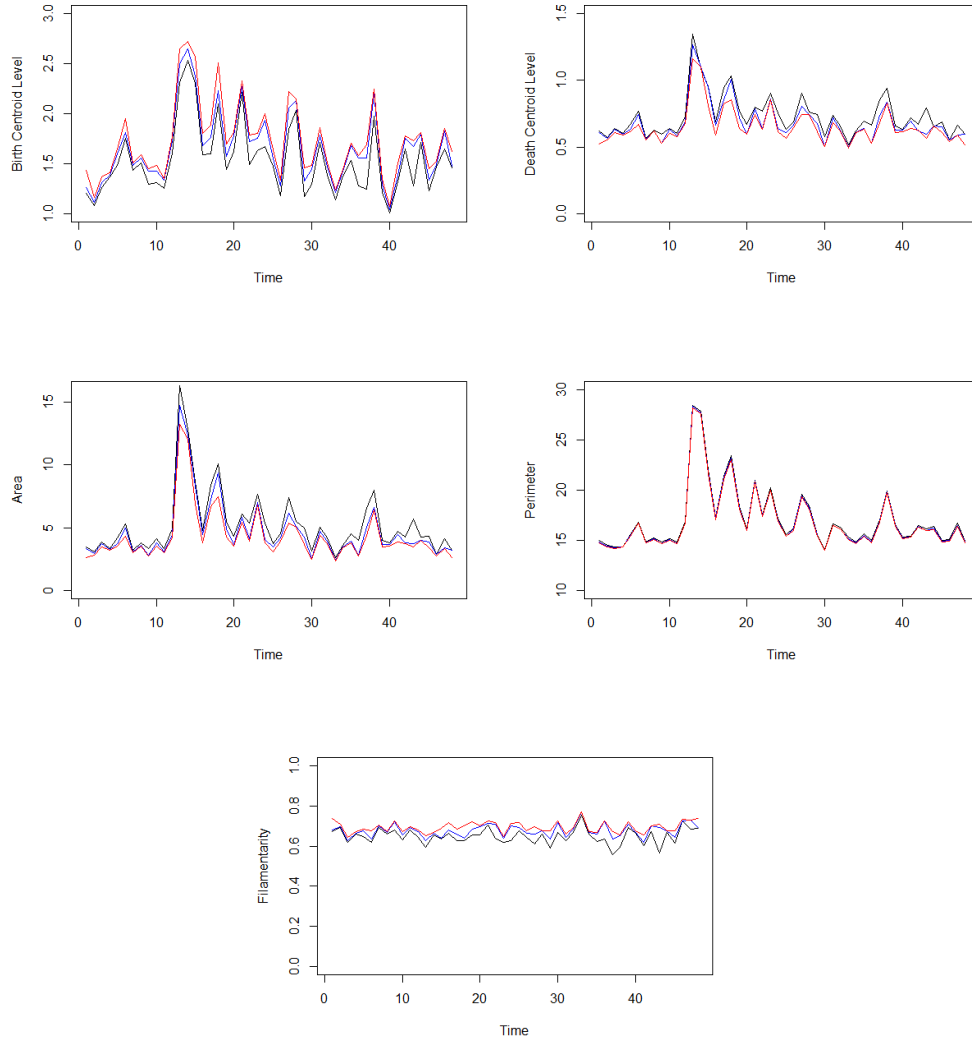


Figure 179: Summary statistics derived using first three landscape functions (black, blue, red) of persistence diagrams of the voids of the full brain model residuals.



### 10.3 Standardised Model Residual Topology Summary Statistics

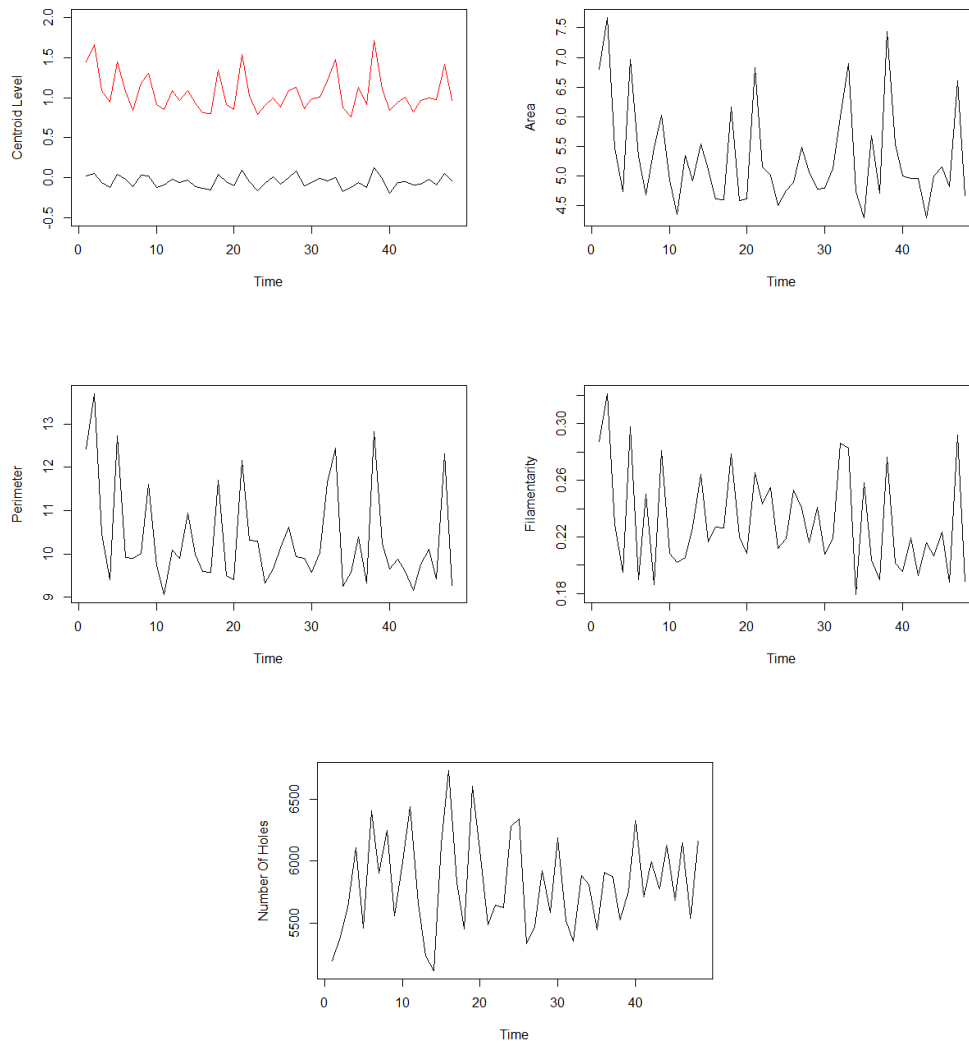


Figure 180: Summary statistics derived using 95% convex peels of persistence diagrams of the holes of the full brain model residuals (standardised). Top left shows the birth (black) and death (red) centroid levels over time, and the remaining plots the other summary statistics.

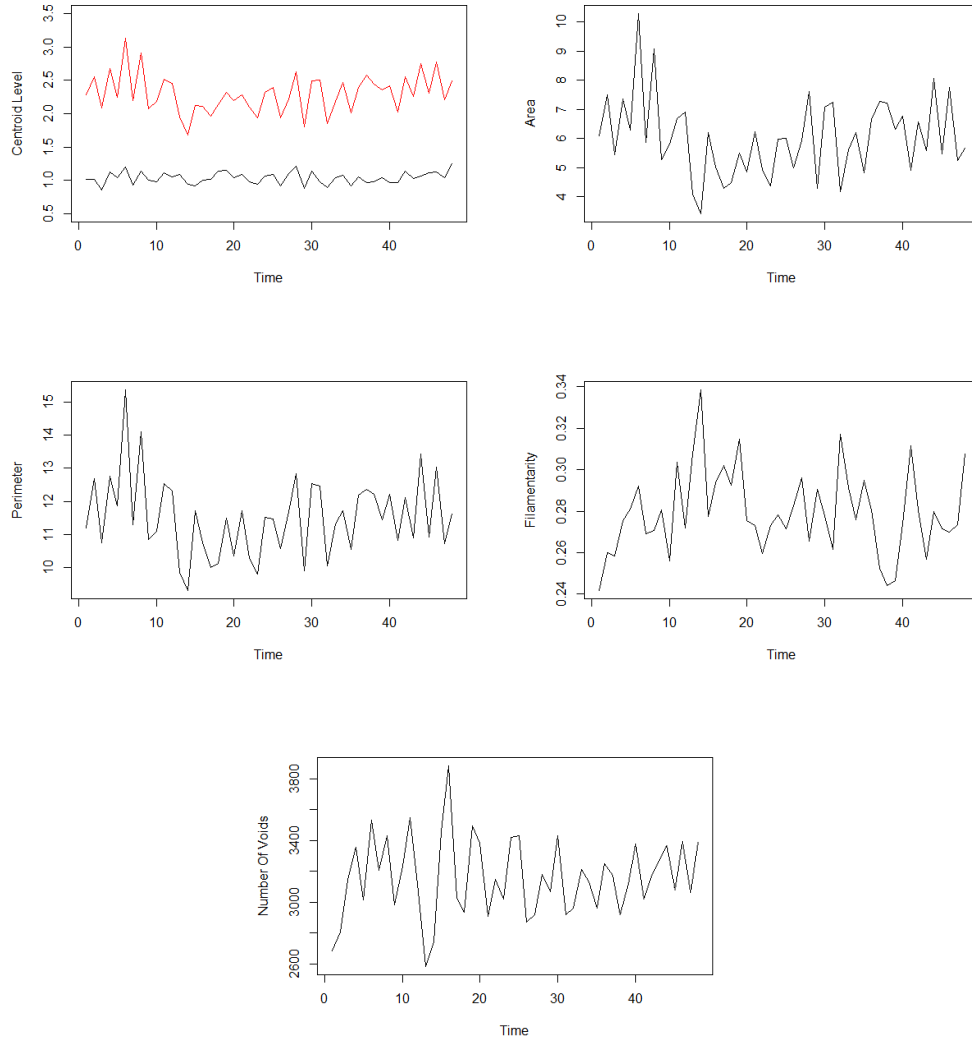


Figure 181: Summary statistics derived using 95% convex peels of persistence diagrams of the voids of the full brain model residuals (standardised). Top left shows the birth (black) and death (red) centroid levels over time, and the remaining plots the other summary statistics.

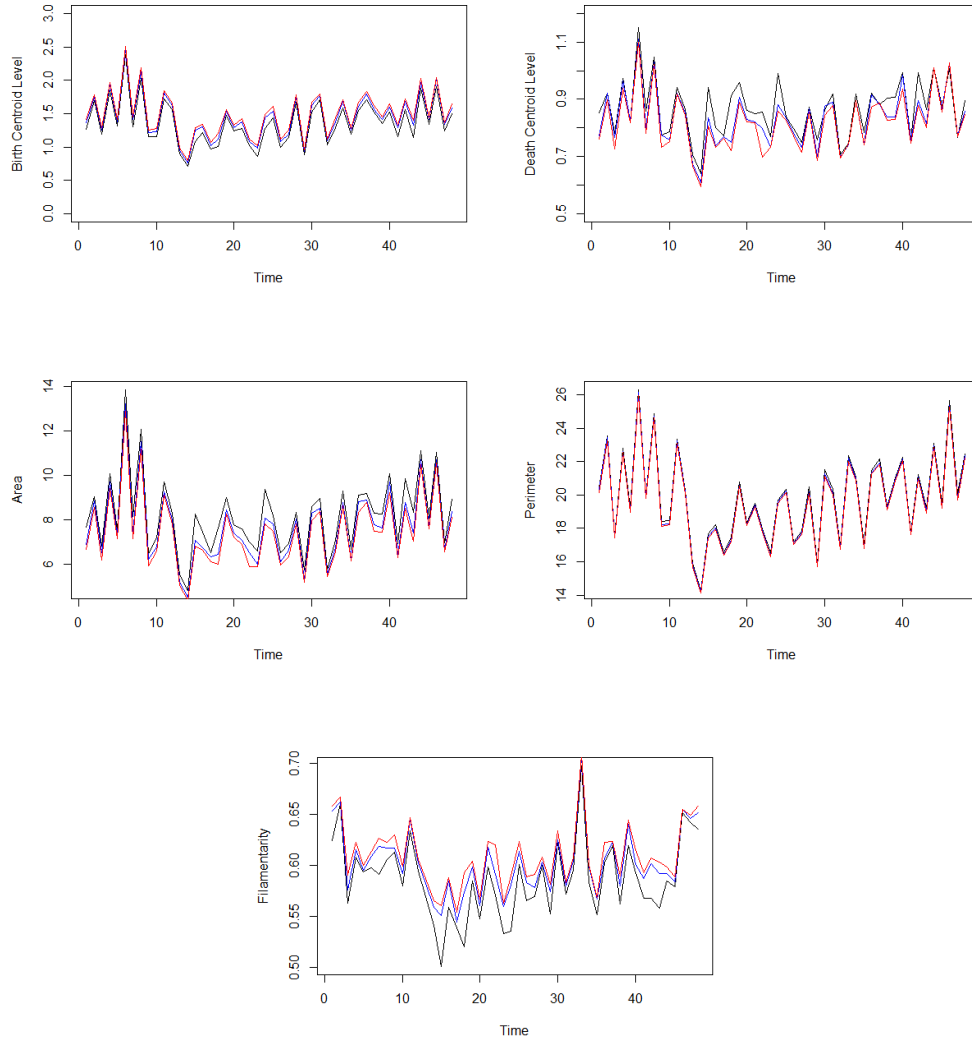


Figure 182: Summary statistics derived using first three landscape functions (black, blue, red) of persistence diagrams of the holes of the full brain model residuals (standardised).

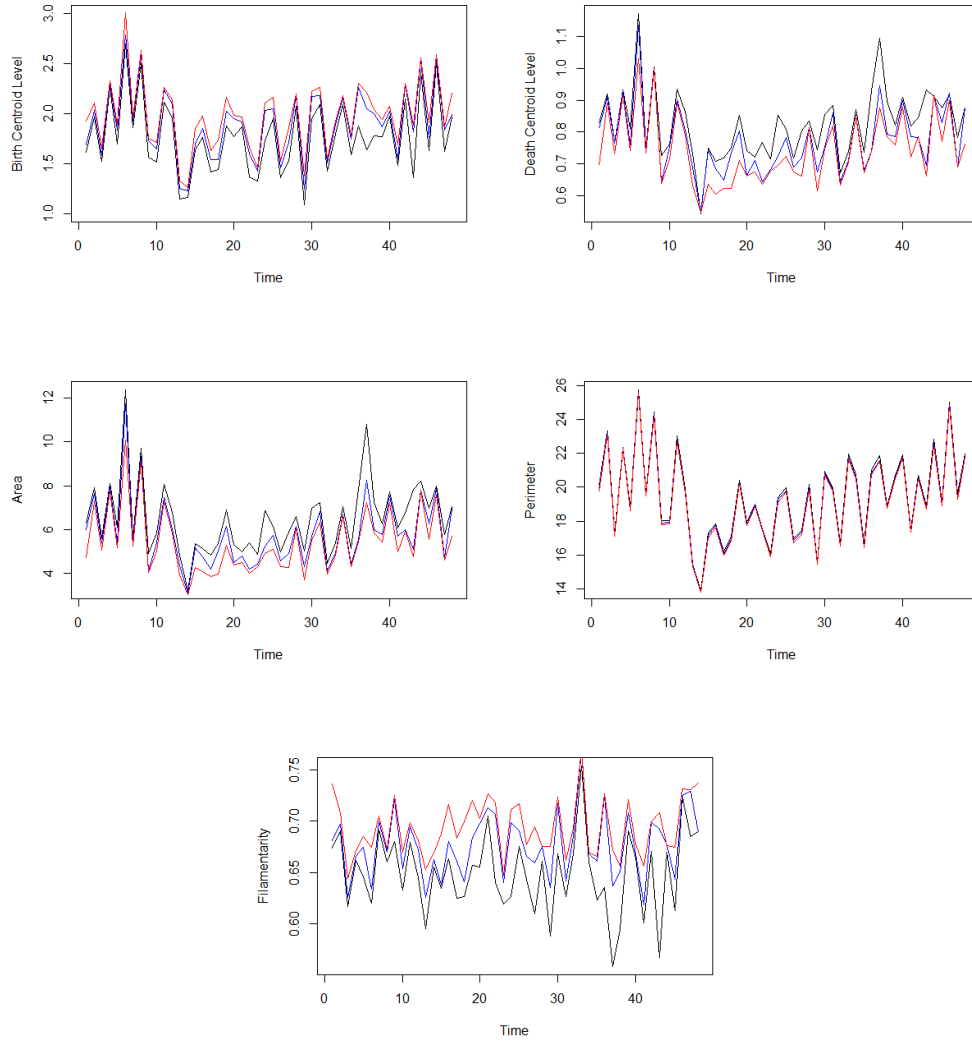


Figure 183: Summary statistics derived using first three landscape functions (black, blue, red) of persistence diagrams of the voids of the full brain model residuals (standardised).

## 10.4 Images Of The Fields On Two Dimensional Lattices

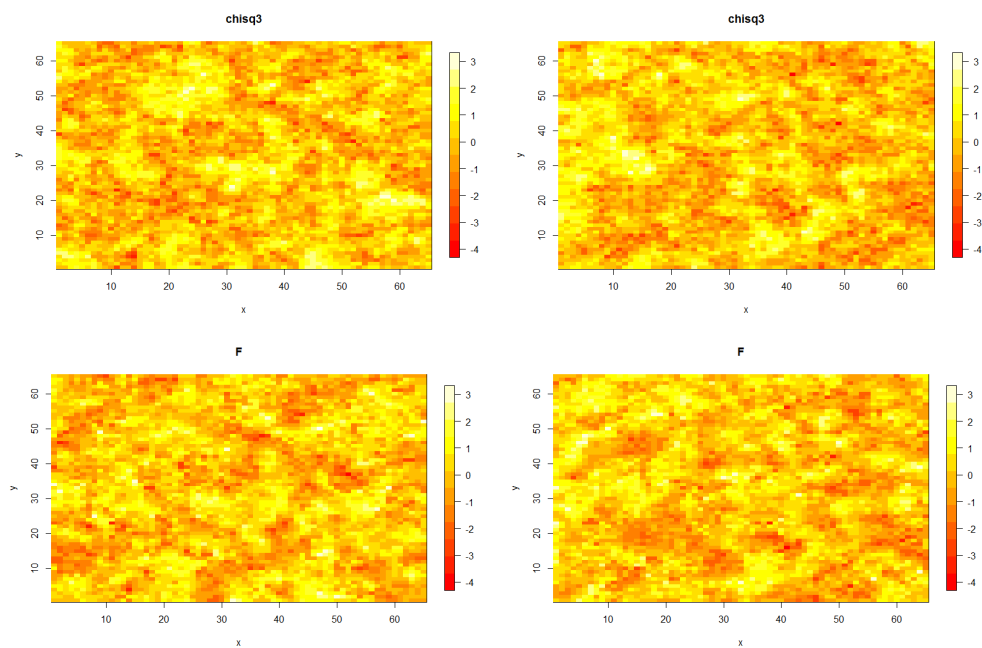


Figure 184: Heatmap images of simulated data from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 2)$ .

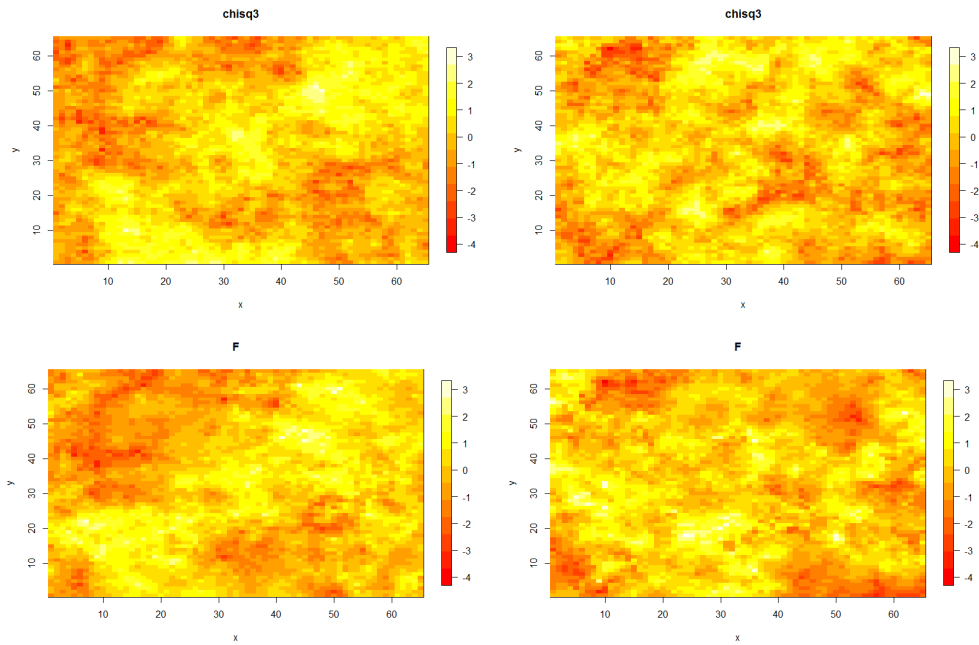


Figure 185: Heatmap images of simulated data from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 5)$ .

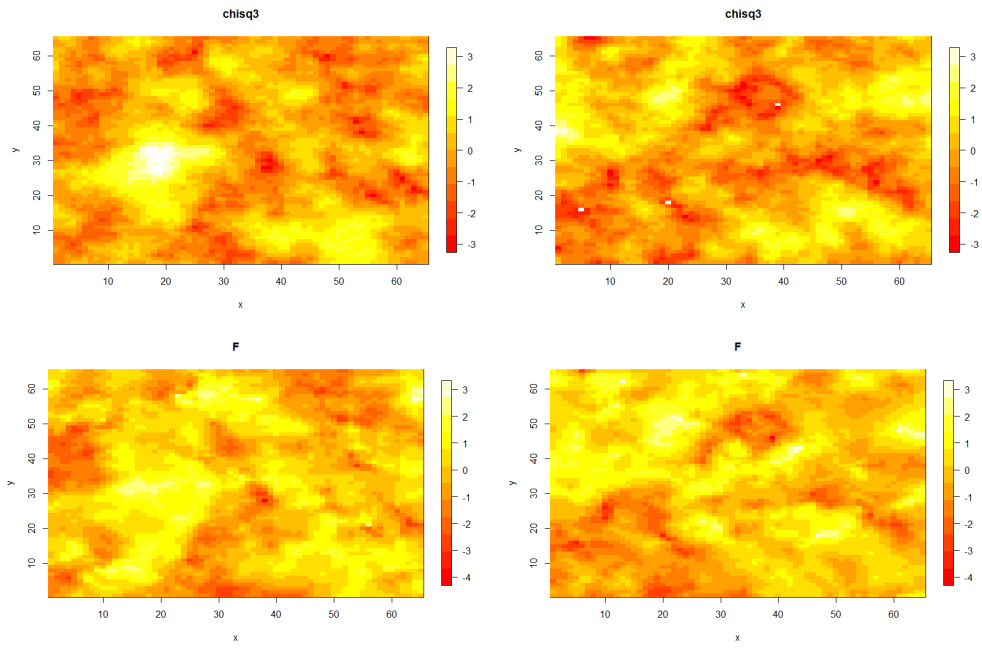


Figure 186: Heatmap images of simulated data from the various distributions with target correlation parameters  $(\nu, \eta) = (1, 5)$ .

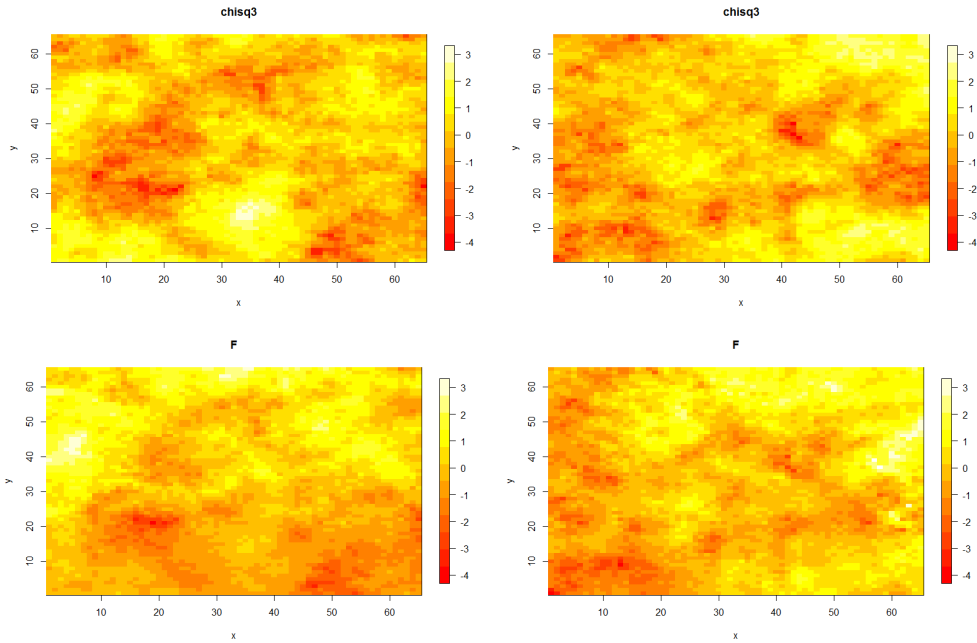


Figure 187: Heatmap images of simulated data from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 20)$ .

## 10.5 Locations Of Components And Holes Of The Fields On Two Dimensional Lattices

### 10.5.1 $(\nu, \eta) = (0.5, 2)$

Figures 188 and 190 show the position of the components and holes of the random realisations seen in Figure 78. Figures 189 and 191 correspond to the heatmaps in Figure 184 .



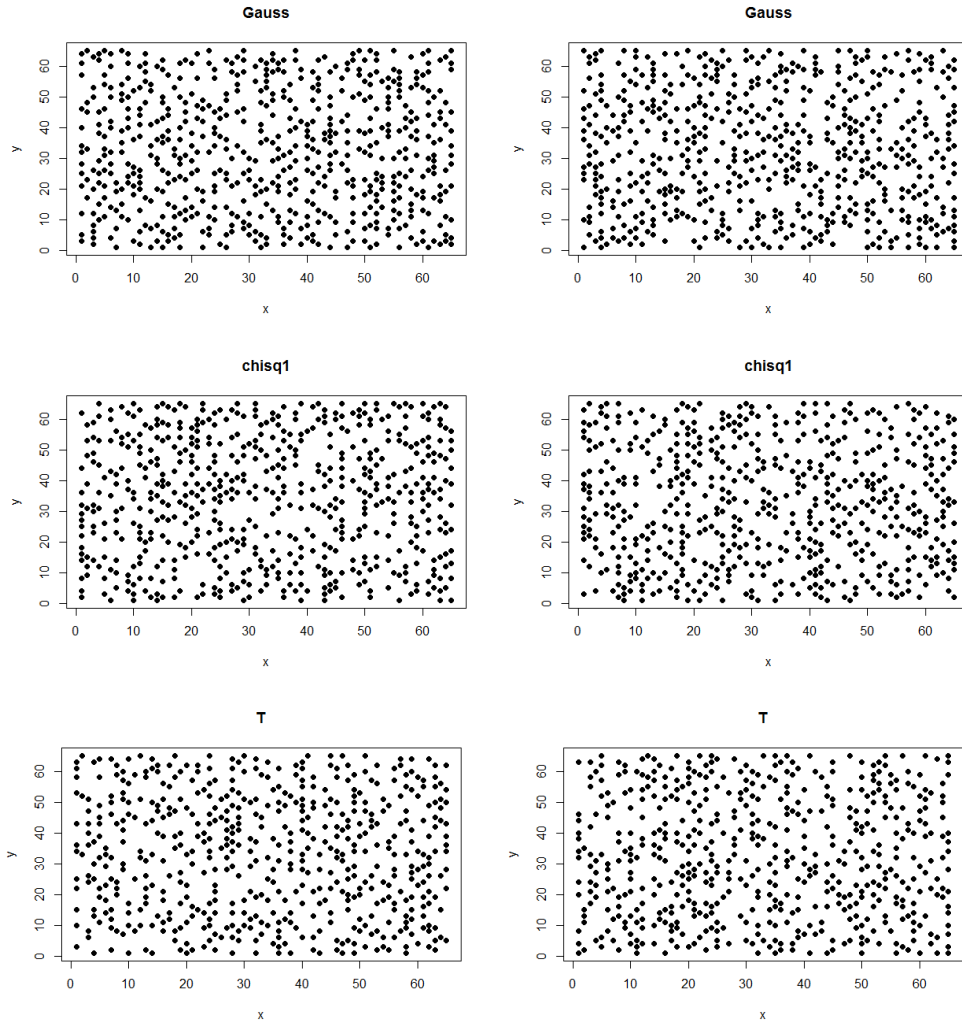


Figure 188: The position of the components from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 2)$ .

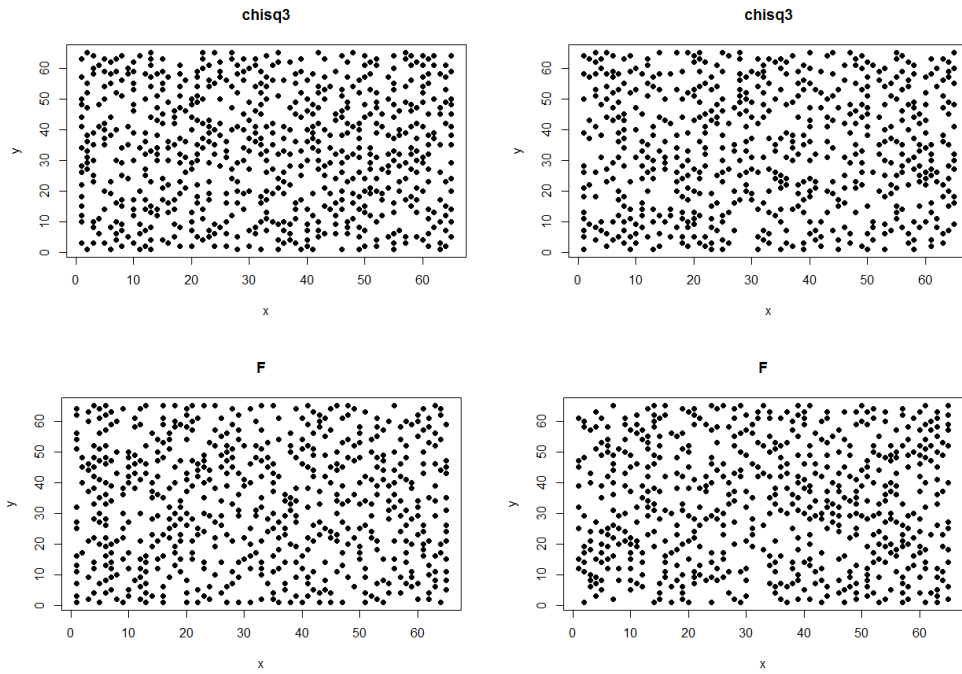


Figure 189: The position of the components from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 2)$ .

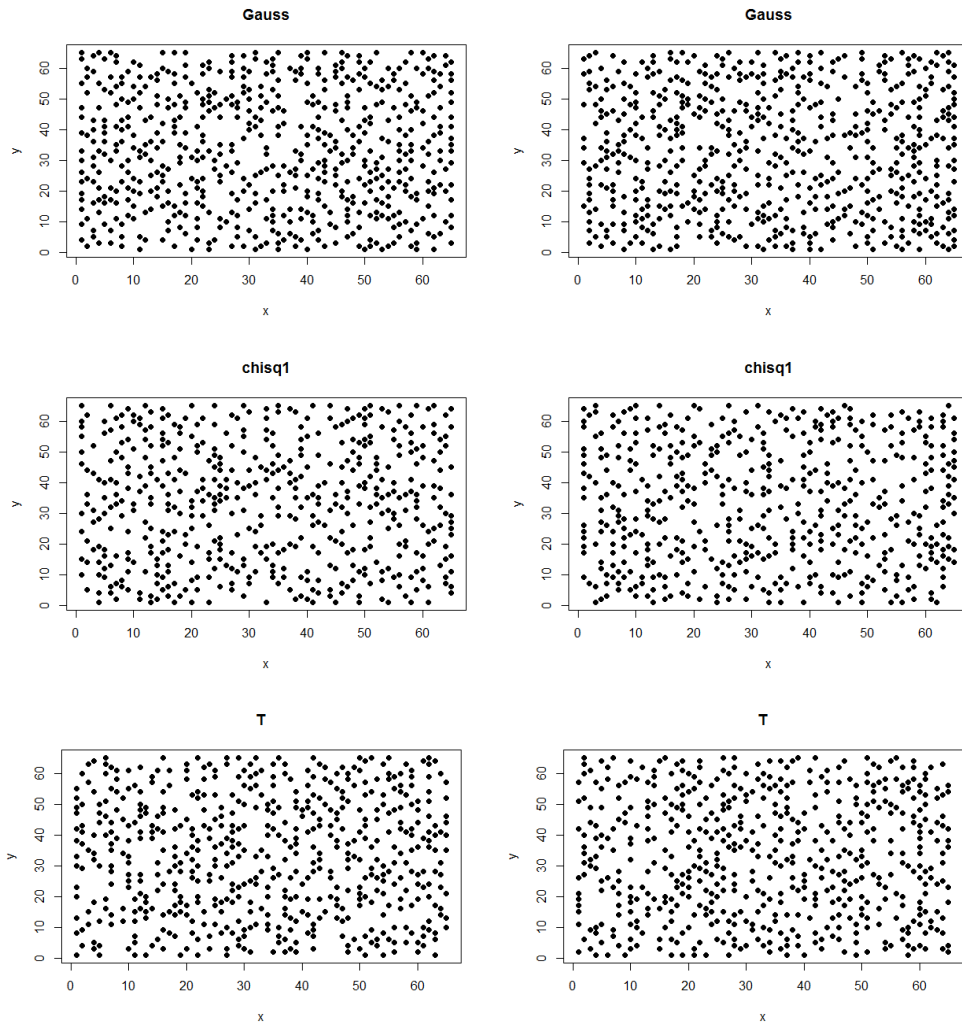


Figure 190: The position of the holes from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 2)$ .

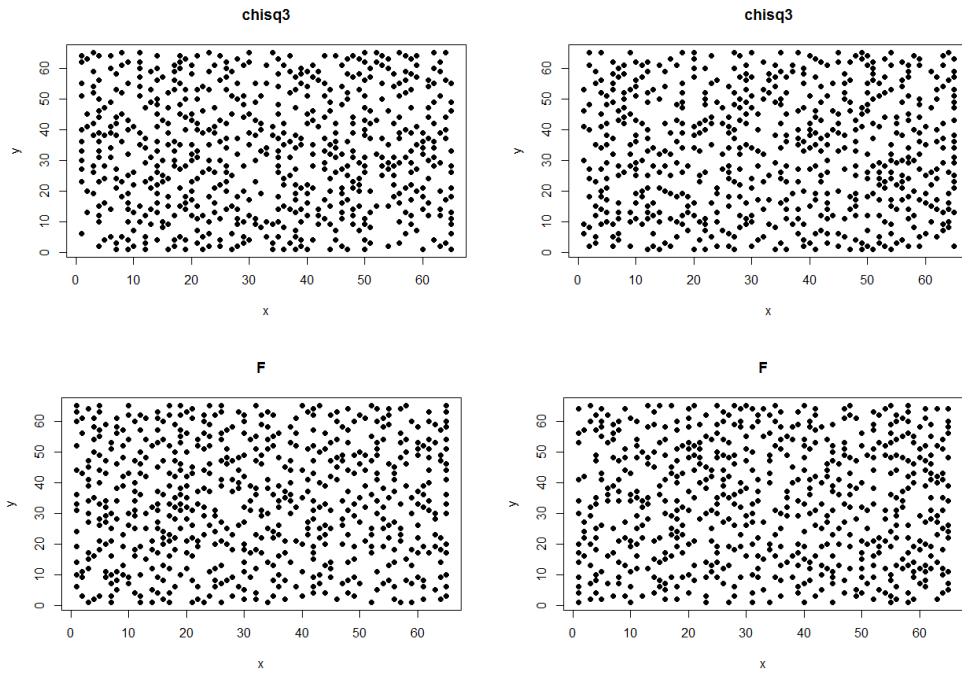


Figure 191: The position of the holes from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 2)$ .

### 10.5.2 $(\nu, \eta) = (0.5, 5)$

Figures 192 and 194 show the position of the components and holes of the random realisations seen in Figure 79. Figures 193 and 195 correspond to the heatmaps in Figure 185 .

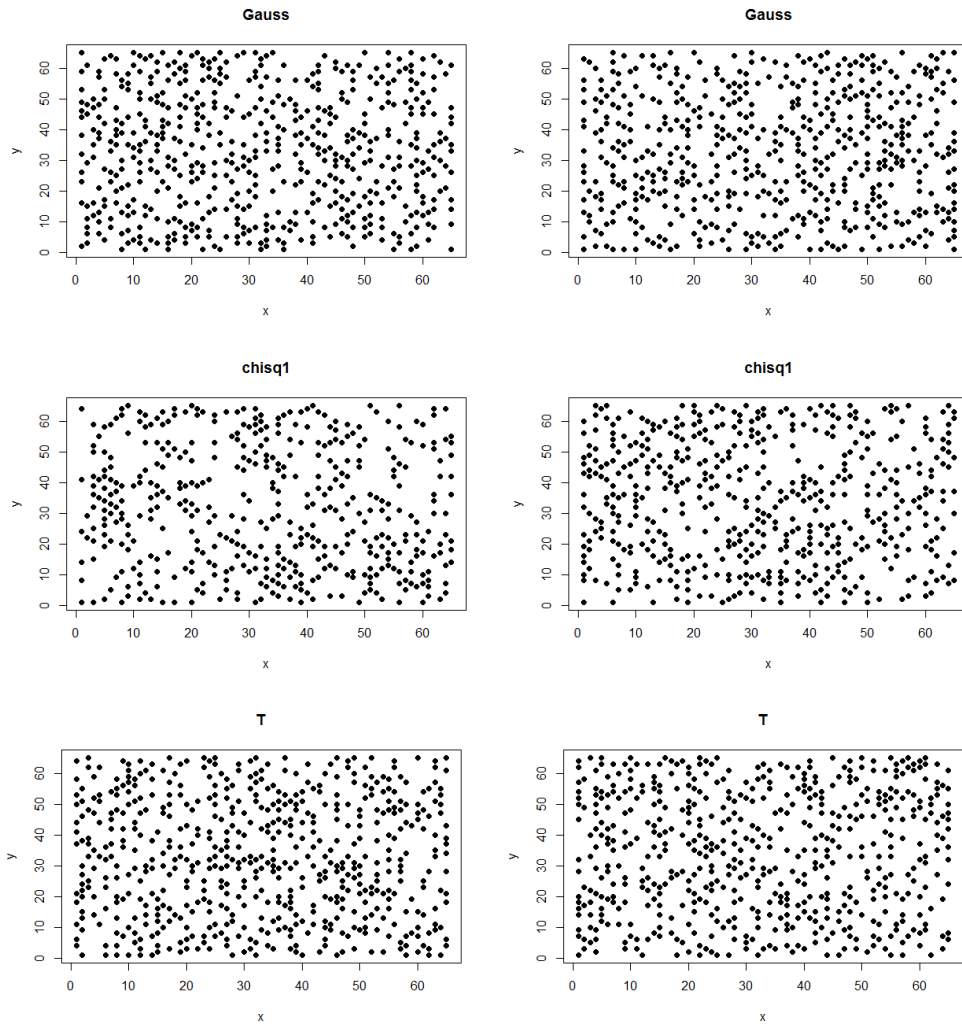


Figure 192: The position of the components from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 5)$ .

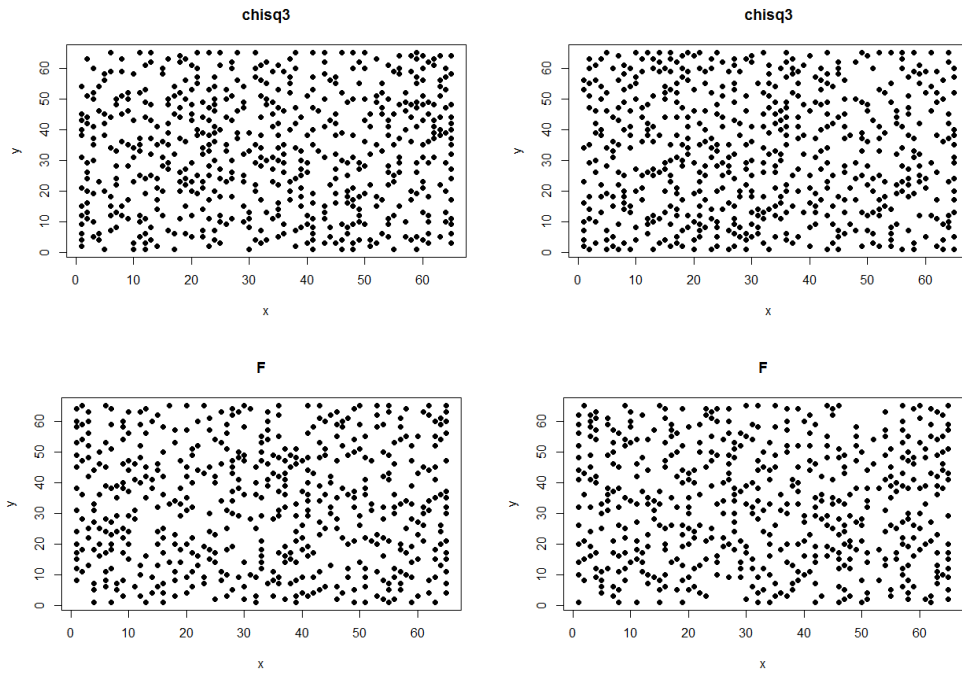


Figure 193: The position of the components from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 5)$ .

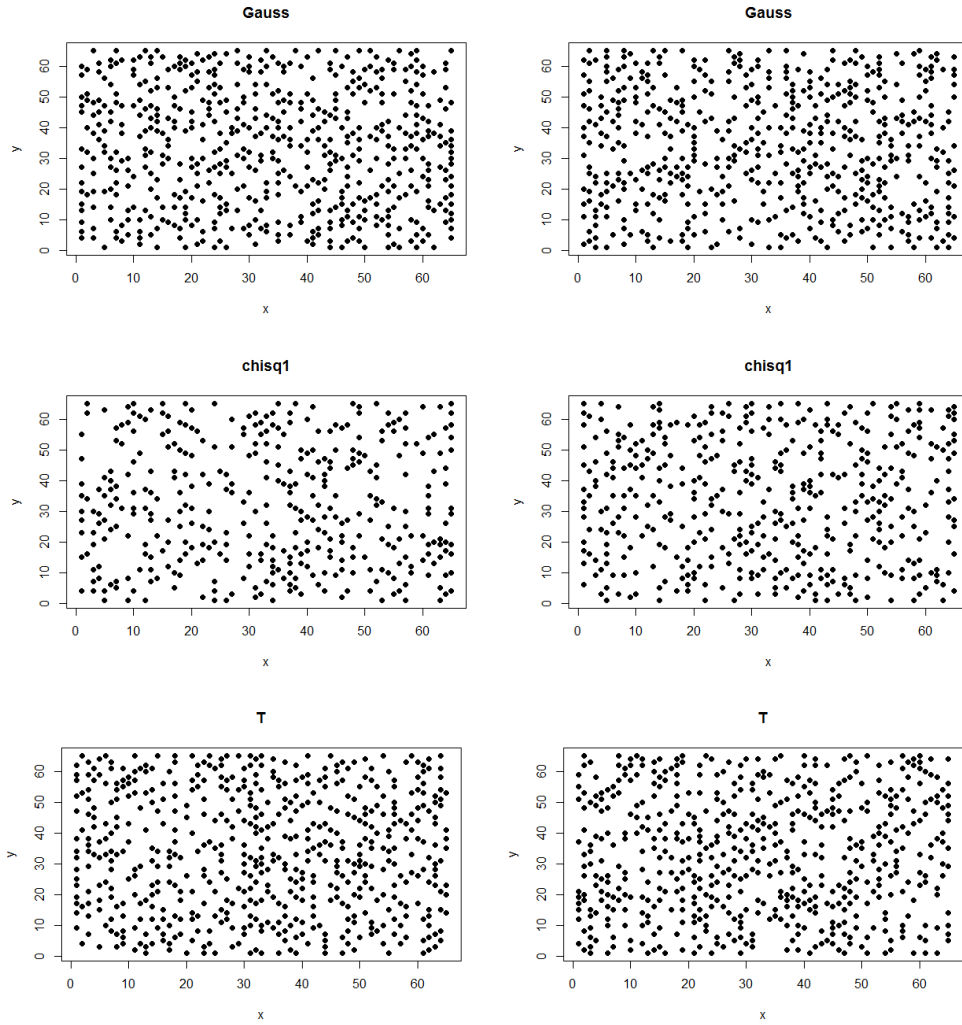


Figure 194: The position of the holes from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 5)$ .

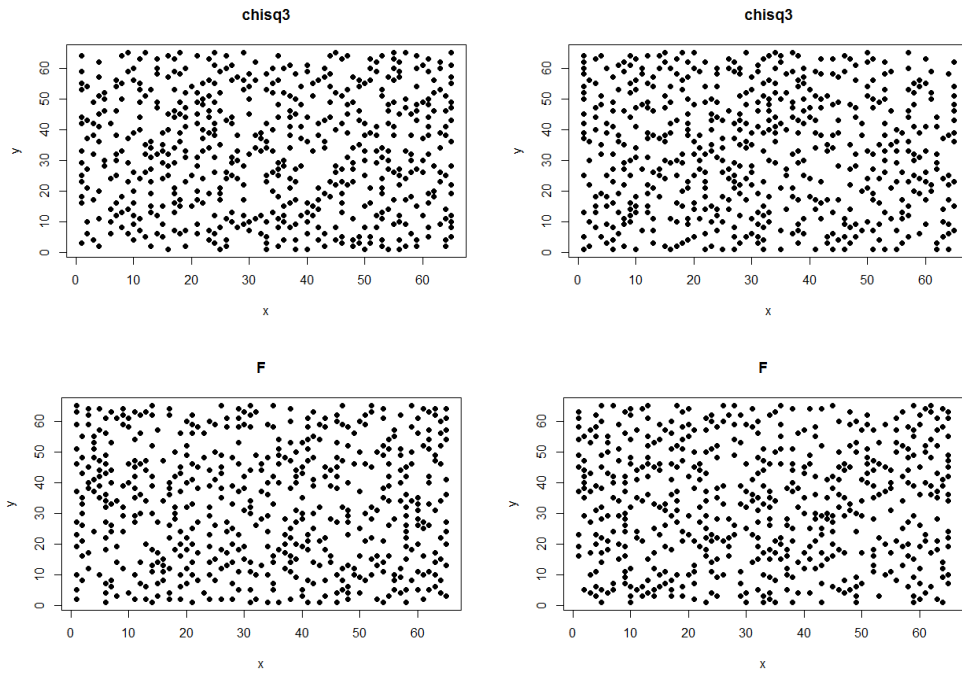


Figure 195: The position of the holes from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 5)$ .

### 10.5.3 $(\nu, \eta) = (1, 5)$

Figures 196 and 197 show the position of the components and holes of the random realisations seen in Figure 186.



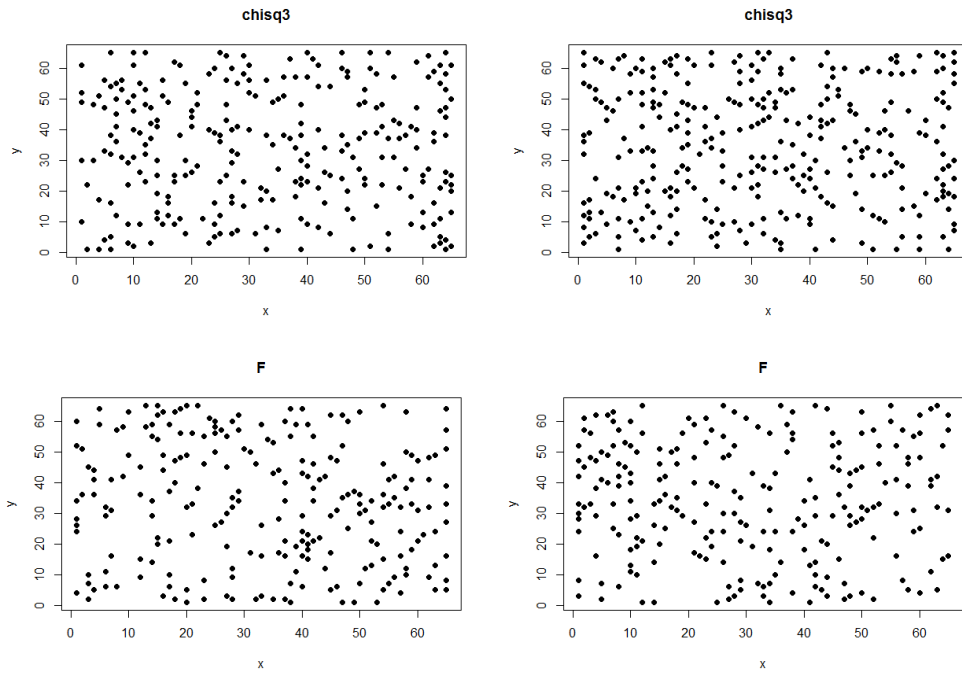


Figure 196: The position of the components from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (1, 5)$ .

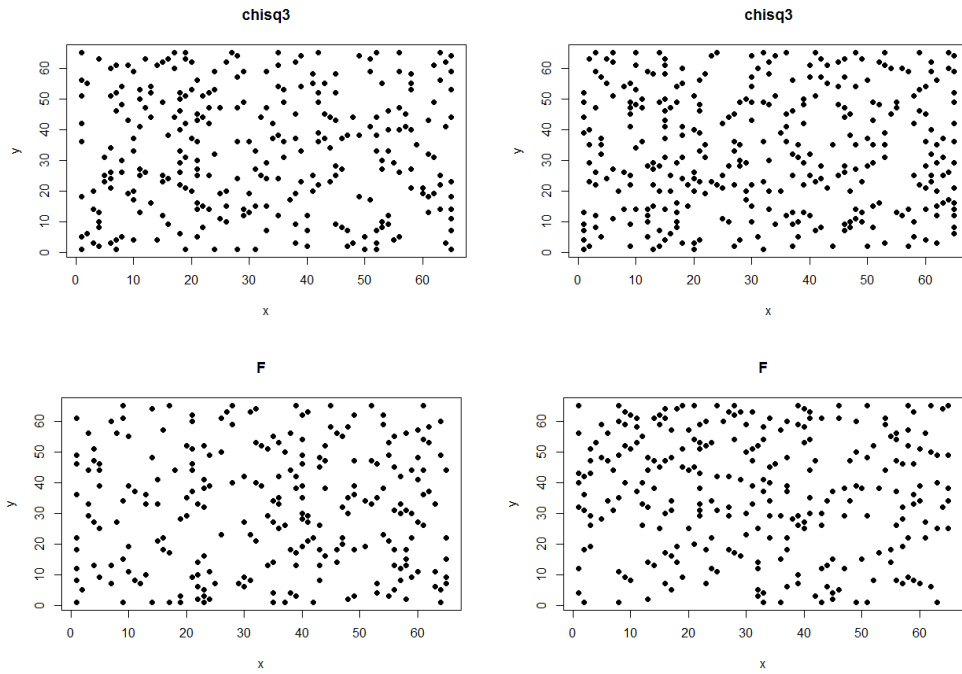


Figure 197: The position of the holes from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (1, 5)$ .

#### 10.5.4 $(\nu, \eta) = (0.5, 20)$

Figures 198 and 200 show the position of the components and holes of the random realisations seen in Figure 81. Figures 199 and 201 correspond to the heatmaps in Figure 187 .

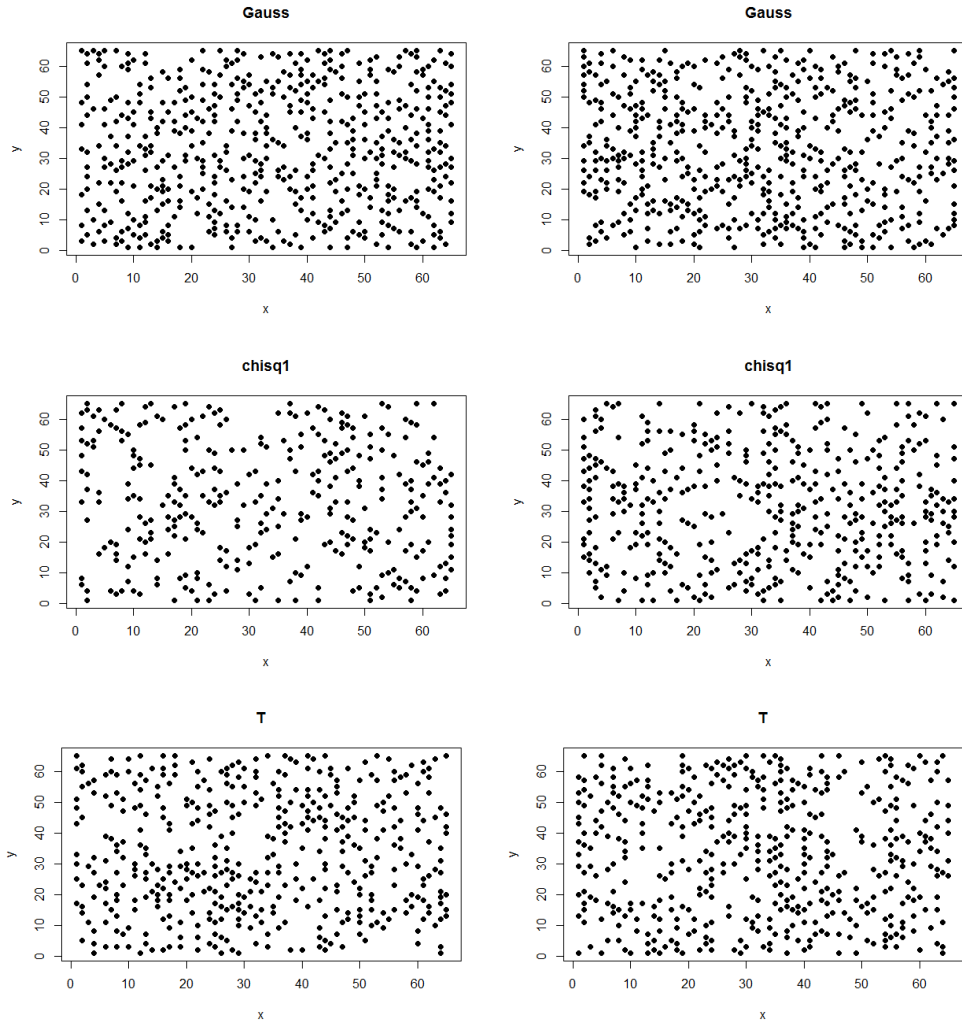


Figure 198: The position of the components from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 20)$ .

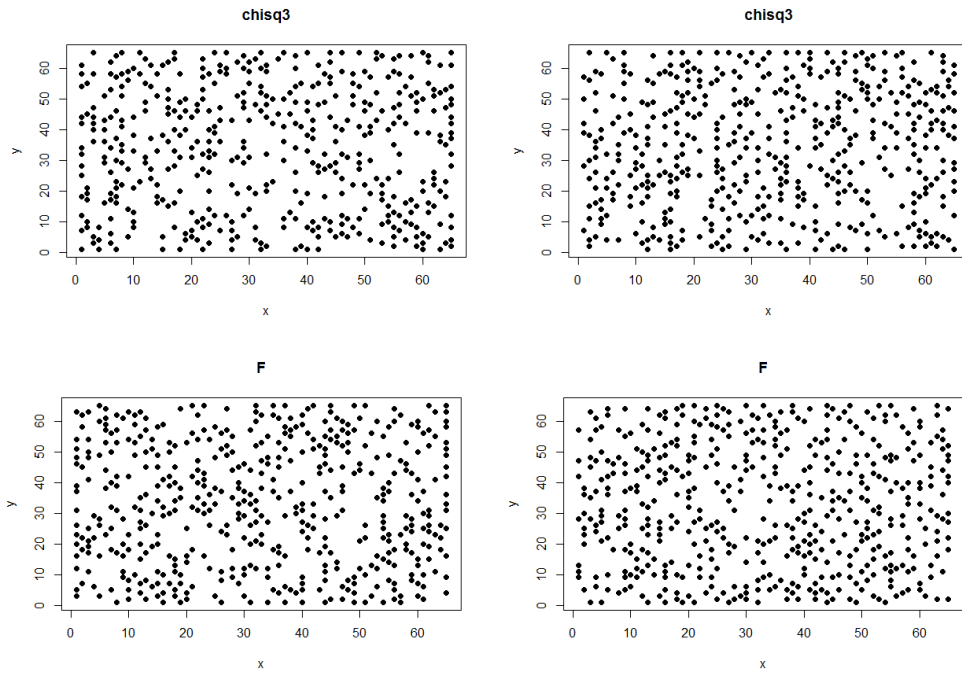


Figure 199: The position of the components from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 20)$ .

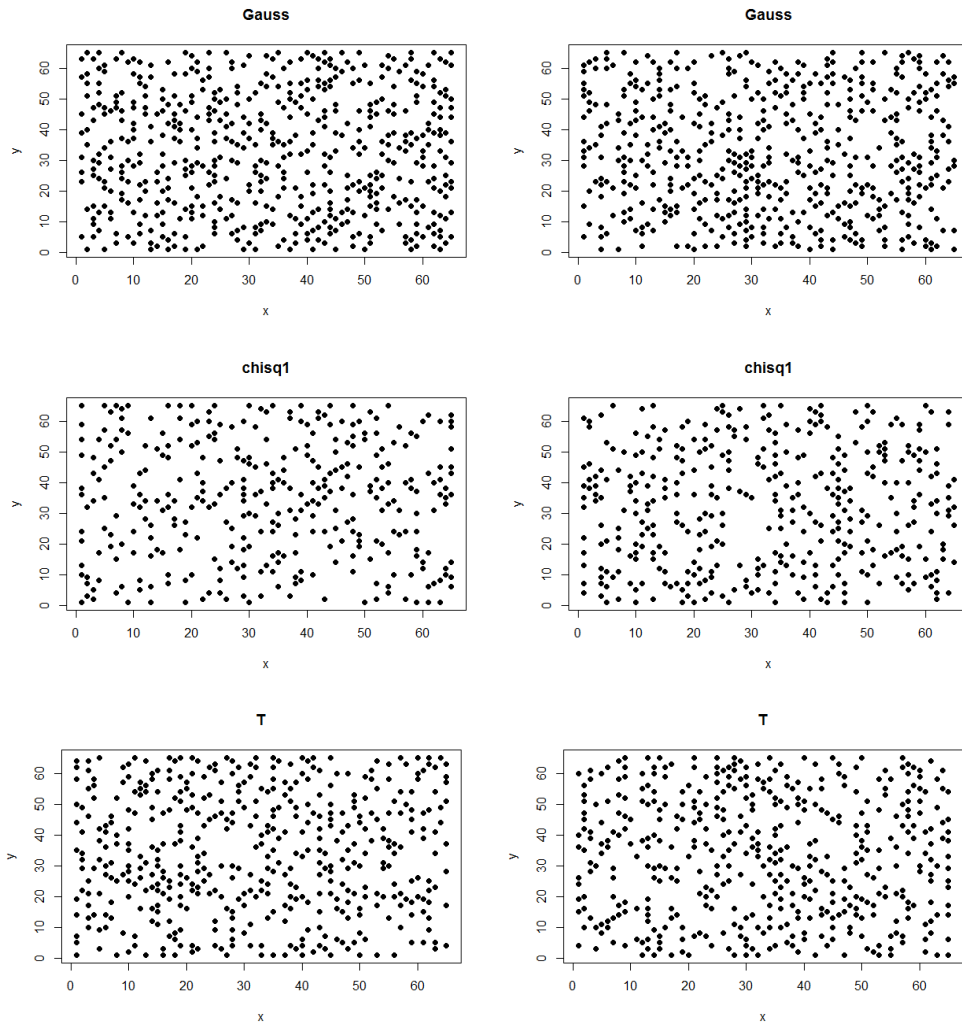


Figure 200: The position of the holes from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 20)$ .

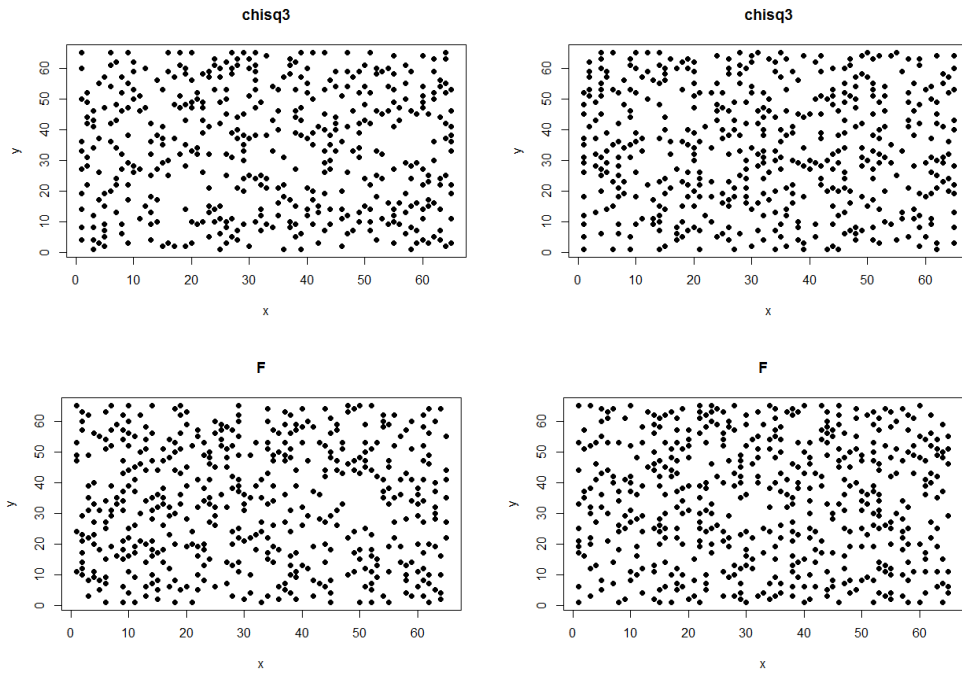


Figure 201: The position of the holes from two realisations from the various distributions with target correlation parameters  $(\nu, \eta) = (0.5, 20)$ .